Title
Sortase Enzyme Mediated Surface Protein Display in Gram-positive Bacteria: Mechanistic Studies and Inhibitor Development

Author
Chan, Albert

Publication Date
2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Sortase Enzyme Mediated Surface Protein Display in Gram-positive Bacteria: Mechanistic Studies and Inhibitor Development

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Biochemistry and Molecular Biology

by

Albert Hay Wah Chan

2014

ABSTRACT OF THE DISSERTATION

Sortase Enzyme Mediated Surface Protein Display in Gram-positive Bacteria:

Mechanistic Studies and Inhibitor Development

by

Albert Hay Wah Chan

Doctor of Philosophy in Biochemistry and Molecular Biology

University of California, Los Angeles, 2014

Professor Robert T. Clubb, Chair

Staphylococcus aureus is a leading cause of hospital-acquired infections in the United States. The rapid emergence of multidrug-resistant strains has created an urgent need for new antibiotics. *S. aureus* and other Gram-positive pathogens use sortase enzymes to display surface virulence factors to promote infections. Compounds that inhibit sortase might therefore function as potent anti-infective agents. This dissertation focuses on the discovery of sortase A (SrtA) inhibitors, as well as more fundamental studies of the mechanism of catalysis. Both structural (NMR) and *in silico* approaches were used for inhibitor development, which resulted in the discovery of a very potent pyridazinone class compound that effectively inhibits *S. aureus* SrtA at nanomolar concentrations. This compound and its related derivatives also inhibit *Bacillus anthracis* SrtA, but they are not toxic to human cells. Therefore, they are promising leads for further development into a therapeutic. To gain insights into the sortase catalytic mechanism, I

also determined the 3D structure of the *B. anthracis* SrtA bound to a substrate analog using NMR methods. The structure reveals a novel N-terminal extension that regulates substrate entry, and a substrate-induced disorder to order transition of one of the active site loops. Ultimately, this new structural information may be useful in guiding the design of pan-sortase inhibitors that can be used to treat a broad spectrum of bacterial infections.

The dissertation of Albert Hay Wah Chan is approved.

Kenneth A. Bradley

Michael E. Jung

Robert T. Clubb, Committee Chair

University of California, Los Angeles

2014

To my parents

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank my advisor, Professor Robert T. Clubb, for his guidance, friendship, and devotion in helping me with this dissertation. I thank him for his dedication in teaching me how to use NMR in research, as well as being a role model for the type of scientist I wish to become one day. Additionally, I would like to thank my collaborators, Professors Michael E. Jung and J. Andrew McCammon for kindly providing me with the resources and direction in all projects. I thank members of my dissertation committee, Professors Kenneth A. Bradley, Genhong Cheng, David S. Eisenberg, and Michael E. Jung for providing valuable advice and reviewing this manuscript. Special thanks go to Drs. Robert Peterson, Robert Damoiseaux, and Gregory Khitrov for their kind technical assistance, Dr. Sung Wook Yi for synthesizing compounds used in all of my research projects, and Dr. Jeff Wereszczynski for teaching me how to use various computational chemistry tools. I thank all of my colleagues in the Robert Clubb laboratory and friends at UCLA, including Brendan Amer, Dr. Timothy Anderson, Dr. Pik "Becky" Chan, Andrea Hadjikyriacou, Grace Huang, Alex Jacobitz, Michele Kattke, Ramsay Macdonald, Dr. Reza Malmirchegini, Melanie Marohn, Dr. Scott Robson, Megan Sjodt, Dr. Thomas Spirig, Dr. Nuttee "Tom" Suree, Dr. Valerie Villareal, and Dr. Ethan Weiner, for their great support. I thank my family in Hong Kong for their nurturing all these years. Finally, I would like to thank the Chemistry-Biology Interface Training Grant (NIH T32GM008496) and Dissertation Year Fellowship (Graduate Division, UCLA) for their financial support.

Chapter 1 contains figures from "The structure of the *Staphylococcus aureus* sortase-substrate complex reveals how the universally conserved LPXTG sorting signal is recognized" (Suree, N., Liew, C. K., Villareal, V. A., Thieu, W., Fadeev, E. A., Clemens, J. J., Jung, M. E.,

and Clubb, R. T. (2009) *J Biol Chem* **284**, 24465-24477) and "Targeting virulence: a new paradigm for antimicrobial therapy" (Clatworthy, A. E., Pierson, E., and Hung, D. T. (2007) *Nat Chem Biol* **3**, 541-548). These figures were reproduced or modified with permission from the Journal of Biological Chemistry and Nature Chemical Biology.

Chapter 2 contains a modified version of "Discovery of *Staphylococcus aureus* Sortase A Inhibitors Using Virtual Screening and the Relaxed Complex Scheme" (Chan, A.H., Wereszczynski, J., Amer, B.R., Yi, S.W., Jung, M.E., McCammon, J.A., and Clubb, R.T. (2013) *Chem Biol Drug Des* **82**, 418-428). AHC, MEJ, JAM, and RTC designed the experiments. AHC performed the molecular docking and FRET based enzymatic assay experiments. JW performed the molecular dynamics experiment. BRA performed the HPLC experiment. SWY synthesized the compounds. AHC, JW, and RTC wrote the manuscript.

Chapter 3 is a manuscript of "NMR assisted docking of a pyridazinone *Staphylococcus aureus* Sortase A inhibitor for structure-based drug design" (Chan, A.H., Weiner, E.M., Yi, S.W., Wereszczynski, J., McCammon, J.A., Jung, M.E., and Clubb, R.T. (2014) To be submitted). AHC, JAM, MEJ, and RTC designed the experiments. AHC performed the LC-MS, molecular docking, enzymatic assays and cytotoxicity test. EMW solved the NMR structure. SWY synthesized the compounds. JW performed the molecular dynamics experiment. AHC, EMW, and RTC wrote the manuscript.

Chapter 4 is a manuscript of "The Structure and Dynamics of the *Bacillus anthracis* Sortase-Substrate Complex Reveals a Regulatory Role of the N-terminal Extension in Substrate Entry and Catalysis" (Chan, A.H., Yi, S.W., Jung, M.E., and Clubb, R.T. (2014) To be submitted). AHC, JME, and RTC designed the experiments. AHC performed all the experiments. SWY synthesized the substrate analog. AHC and RTC wrote the manuscript.

VITA

Albert Hay Wah Chan received primary and secondary education in Hong Kong. In 2004, Albert came to the United States and attended Foothill College in Los Altos Hills, California, majoring in Chemistry. In 2006, he transferred to University of California, Berkeley, where he received a Bachelor of Science degree in Chemical Biology with High Honors in May 2008. In September 2008, he entered the Biochemistry and Molecular Biology graduate program at University of California, Los Angeles. While in graduate school, Albert received three years of support from the Chemistry-Biology Interface training program from July 2009 to June 2012, and Dissertation Year Fellowship from October 2013 to June 2014. Albert is an author on three papers and a poster in an international conference:

Suree, N., Yi, S.W., Thieu, W., Marohn, M., Damoiseaux, R., **Chan, A.**, Jung, M., and Clubb, R.T. (2009). Discovery and structure-activity relationship analysis of *Staphylococcus aureus* sortase A inhibitors. Bioorg. Med. Chem. *17*, 7174-85.

Chan, A.H., Wereszczynski, J., Amer, B.R., Yi, S.W., Jung, M.E., McCammon, J.A., and Clubb, R.T. (2013). Discovery of *Staphylococcus aureus* Sortase A Inhibitors Using Virtual Screening and the Relaxed Complex Scheme. Chem. Biol. Drug Des. *82*, 418-28.

Sjodt, M., Macdonald, R., Sprig, T., **Chan, A.H.**, Fabian, M., Olson, J.S., Gell, D.A., Clubb, R.T. (2014). NMR Model of the 38.8kDa hemoglobin receptor from *Staphylococcus aureus*. Submitted.

Chan, A.H., Weiner, E.M., Wereszczynski, J., McCammon, J.A., and Clubb, R.T. "Computational Design of *Staphylococcus aureus* SrtA inhibitors." Poster presentation at the 52nd Interscience Conference on Antimicrobial Agents and Chemotherapy, San Francisco, CA, September 11, 2012.

Chapter 1

Introduction to the Gram-positive Bacterial Protein Export and Surface Display Systems

1.1 Overview

The increasing occurrence of multi-antibiotic resistant pathogenic Gram-positive and Gram-negative bacteria has caused alarm in the medical community. In particular, new antibiotics are urgently needed to combat the ESKAPE pathogens (*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumonia, Acinetobacter baumanii, Pseudomonas aeruginosa,* and *Enterobacter* species) (1,2). We are especially interested in finding new antibiotics to treat *S. aureus* infections, since this pathogen is widespread and is a leading cause of both hospital- and community-acquired infections in the United States. *S. aureus* is also capable of causing a wide range of diseases, and is estimated to kill 11,000 people in the United States each year, which is more than the number of deaths caused by HIV (3).

To put our antibiotic work into context, in this chapter I first review the general structure of the Gram-positive bacterial cell wall, the protein export systems, and the different ways that proteins are displayed on the cell surface, followed by a more focused discussion of our drug target, sortase, the enzyme that is responsible for covalently anchoring many surface proteins to the cell wall peptidoglycan of Gram-positive bacteria. Finally, the current classes of antibiotics and their targets are summarized, and past and current work in sortase inhibitor development is discussed.

1.2 Gram-positive bacterial cell wall structure

The cell wall of bacteria serves as an exoskeleton to maintain cell shape and to allow bacteria to withstand mechanical stresses. It has been shown that isolated cell walls are capable of retaining the shape of the intact bacterial cell (4), and that cells become round when the cell wall is removed (5). The cell wall also serves as a barrier against hydrolases or membrane toxic

compounds from the surrounding environment. In addition, it functions as an anchor point for a variety of proteins to be presented on the cell surface. Those surface proteins often play important roles in the infection process, such as by promoting cell adhesion, nutrient acquisition, and immune evasion. At the minimum, the cell wall of Gram-positive bacteria is composed of a single cytoplasmic membrane surrounded by a thick layer of peptidoglycan (~55 nm and 35 nm for Gram-positive bacteria *Bacillus subtilis* and *Stapylococcus aureus*, respectively, compared to ~6 nm for the Gram-negative bacterium *Escherichia coli* (6-8)). A general depiction of the cell wall is shown in Fig. 1.1. As reviewed by Vollmer and colleagues, peptidoglycan is a polymer of muropeptide that consists of a disaccharide of *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) linked by a β(1→4) bond, and a peptide stem of which the exact composition depends on the bacterium (9). In *S. aureus*, the peptide stem has the sequence L-Ala-D-iGln-L-Lys-D-Ala-D-Ala, with L-Ala attached to MurNAc, and a $(L\text{-}Gly)_5$ "bridge" attached to the side chain amino group of L-Lys (Fig. 1.2). In *Bacillus anthracis*, the peptide stem is identical to that of *S. aureus*, except for the 3rd residue which is a *meso*-diaminopimelic acid instead of L-Lys. The $(L\text{-}Gly)_5$ bridge is also absent in the *B. anthracis* peptidoglycan. The muropeptide subunits are linked together via β(1→4) linkages between the disaccharides to form long glycan strands, and individual glycan strands are cross-linked via the peptide stems, commonly between the side chain amino group of the 3rd residue of one peptide stem and the carboxyl group of the 4th residue of another peptide stem. The linkage can either be direct (such as in *B. anthracis*) or through a cross-bridge (such as in *S. aureus*). The degree of cross-linkage varies from ~20% in *E. coli* to over 93% in *S. aureus*.

A major component of the cell wall that is present in Gram-positive bacteria, but not in Gram-negative bacteria is teichoic acid (10). There are two types of teichoic acid, one known as

lipoteichoic acid (LTA) which is attached to the cell membrane, and another one known as wall teichoic acid (WTA) which is covalently attached to the peptidoglycan. It has been estimated that the total mass of WTAs constitutes up to 60% of the cell wall (11). A teichoic acid polymer can be divided into two parts: a main chain polymer composed of phosphodiester-linked polyol repeat units such as ribitol 5-phosphate and glycerol 3-phosphate, and a disaccharide linkage unit that is used to covalently link the polymer to the peptidoglycan MurNAc or lipid. Several roles have been proposed for teichoic acids, including host cell attachment as demonstrated by the loss of epithelial and endothelial cell binding by *S. aureus* mutants that lack WTA (12,13). Due to the negative charges along the polymer, teichoic acids have a high affinity for bivalent cations, and might provide an ion-storage mechanism for magnesium (14). LTA is also found to be indispensable for cell division (15), possibly because teichoic acids can recruit the positively charged autolysins, enzymes that cleave peptidoglycan strands to separate daughter cells upon cell division (11).

Some bacteria contain additional components in their cell walls. *B. anthracis* cell wall, for example, contains a layer of secondary cell wall polysaccharide (SCWP), an S-layer, and a capsule (16). *B. anthracis* SCWP is a polymer with the repeating unit [→6)-α-GlcNAc-(1→4)-β-ManNAc-(1→4)-β-GlcNAc-(1→], where α-GlcNAc is substituted with α-Gal and β-Gal at O3 and O4, respectively, and the β-GlcNAc is substituted with α-Gal at O3. SCWP is linked to the peptidoglycan MurNAc through a β-ManNAc-(1→4)-β-GlcNAc disaccharide linkage unit similar to how WTA is linked to the peptidoglycan (16). SCWP serves as a docking spot for S-layer proteins. S-layer proteins generally consist of at least two domains: a crystallization domain that mediates protein-protein interaction required to form a two-dimensional crystalline array on the surface of the cell, and an anchoring domain such as a surface layer homology (SLH) domain

or a *Clostridium difficile* CW_binding_2 (CWB2) domain that allows the S-layer protein to be anchored to the SCWP (17). *B. anthracis* has two major S-layer proteins, Sap and EA1, which are incorporated into the S-layer at different stages of growth (18). Although lacking known functional domains that specify peptidoglycan hydrolase activity, both Sap and EA1 were found to have such activity, and elongated chains of bacteria are seen in a *sap* mutant (19), highlighting the role of the S-layer in cell division. Finally, some bacteria contain capsules composed of polysaccharides (e.g. *S. aureus*), hyaluronic acid (e.g. *Streptococcus pyogenes*) or poly-D-γ-glutamic acid (e.g. *B. anthracis*). Capsules allow bacteria to resist phagocytosis, and may block bactericidal activities of neutrophil extracts and some cationic peptides (20,21). It has been shown that capsules promote staphylococcal virulence, but loss of capsule expression, interestingly, may play a role in *S. aureus* persistence in the chronically infected host (22,23).

1.3 Gram-positive protein export system

A major protein export system into or across the bacterial cell membrane is the general secretion pathway (Sec) (24). This pathway has been studied extensively in Gram-negative bacteria using *E. coli* as the model system, and to a lesser extent, in Gram-positive bacteria using *B. subtilis*. Most Sec substrates are initially synthesized as precursors with an N-terminal signal peptide to signal their export. The Sec signal peptide consists of three parts: a positively charged N-terminal region, a hydrophobic core, and a polar C-terminal region. Lipoprotein precursors contain an extra element in their signal peptide called a lipobox motif, which includes a highly conserved cysteine residue for lipid addition. During translocation, non-lipoprotein Sec precursor proteins are processed by a type I signal peptidase to cleave off the signal peptide, whereas

lipoprotein precursors are processed by a type II signal peptidase. The released signal peptides are further degraded by signal peptidases such as RasP in *B. subtilis*.

Sec precursor proteins may be translocated in a co- or posttranslational manner. In both cases, precursor proteins are threaded into a translocation channel formed by SecY, SecE, and SecG (hereafter referred as SecYEG). Co-translationally translocated proteins are typically lipoproteins destined to be inserted into the cell membrane. When their highly hydrophobic signal peptides or transmembrane segments emerge from the ribosome, they are recognized and bound by a signal recognition particle (SRP) consisting of an Ffh protein and a small cytoplasmic RNA, and in the case of *B. subtilus*, an HbsU protein as well. This SRP-ribosome nascent chain complex then docks to the SRP receptor FtsY, which is further transferred to the SecYEG translocon. The energy for the translocation is thought to be provided by the further elongation of the precursor protein at the ribosome (25). On the other hand, post-translationally trancslocated Sec precursor proteins are fully synthesized first but kept in an unfolded, translocation-competent state by chaperones. The chaperone then delivers the precursor protein to the SecA ATPase associated with SecYEG. The precursor protein is then pushed through the SecYEG translocon by SecA in an ATP-dependent manner (26). In Gram-negative bacteria, SecB is the general secretion-specific chaperone protein that binds to many Sec precursor proteins. However, no homologs of SecB can be found in Gram-positive bacteria. In *B. subtilis*, it has been proposed that CsaA might fulfill the role of SecB since it was shown to possess an affinity for SecA and can bind to a subset of Sec precursor proteins (27). In addition, the general chaperones GroEL-GroES/DnaK-DnaJ-GrpE/trigger factor and the soluble form of SecA might also be involved in the SecA-dependent protein export pathway (28-30).

Interestingly, some Gram-positive bacteria possess a second SecA (SecA2), or even a second SecY (SecY2) protein (31). In contrast to the housekeeping SecA and SecY proteins, these accessory Sec2 components are mostly dispensable and are required for the export of much fewer precursor proteins, many of which are virulence factors. Depending on whether SecY2 is present, the accessory Sec2 system can be divided into two groups: the SecA2-only and the SecA2/SecY2 systems. In the SecA2-only system, it is believed that SecA2 mediates membrane translocation of its substrate through the same SecYEG translocon, either alone, or in cooperation with the housekeeping SecA (32). At this stage, it is unclear how SecA2-only substrates are specifically recognized by SecA2, especially because some SecA2-dependent substrates lack a signal peptide (33). The SecA2/SecY2 system is specifically involved in the export of large, serine-rich glycoproteins, which cannot be translocated via the housekeeping Sec system due to their glycosylation. Substrates of this pathway possess an unusually long N-terminal region in their signal peptide, and two or more glycines within the hydrophobic core region act as SecA2/SecY2 pathway specificity determinants (34). In *Streptococcus gordonii* where the SecA2/SecY2 system was first discovered, the gene encoding the SecA2/SecY2 substrate is part of an accessory *sec2* locus. The *sec2* locus also encodes five accessory secretory proteins (Asp1-5), of which Asp4 and Asp5 show some sequence similarities to SecE and SecG, respectively, suggesting that SecA2/SecY2-Asp4-Asp5 may constitute a second translocon distinct from the housekeeping SecA/SecYEG translocon (35). However, homologs of Asp4 and Asp5 are not always present in bacteria with SecA2/SecY2 systems.

Besides the Sec pathway, other protein export pathways exist in Gram-positive bacteria but they have more limited sets of substrates. While most unfolded precursor proteins are typically exported through the Sec pathway, folded proteins are exported by the twin-arginine

translocation (Tat) pathway (36). This pathway is named as such because proteins destined for the Tat pathway possess a conserved motif (S/T-R-R-*X*-F-L-K, where *X* stands for any amino acid) located between the N-terminal domain and the hydrophobic domain of the signal peptide. This pathway is important in exporting cofactor-binding proteins, as correct folding is required for cofactor binding in the cytosol. The number of Tat-dependent substrates varies a lot in Gram-positive bacteria, ranging from one iron-dependent peroxidase FepB in *S. aureus* (37), to more than 100 in *Streptomyces coelicolor* (38). Gram-positive bacteria also possess a novel secretion system that is not found in Gram-negative bacteria. This system, called the type VII or WXG100 secretion system, is responsible for exporting WXG100 proteins, a family of small (~100 residues) α-helical proteins that are often virulence factors (39,40). The molecular mechanisms of substrate recognition and secretion of this pathway are still under investigation.

1.4 Non-covalent surface protein display

After protein export, proteins can be integrated into the membrane, secreted into the environment, or attached to various components of the cell wall either covalently or non-covalently. In both Gram-positive and -negative bacteria, most proteins that attach to the cell wall non-covalently make use of the lysin motif (LysM) domain. LysM domains can be found in the N-terminal, C-terminal, or even the central part of proteins (41). A typical LysM motif is about 40 residues long, and adopts a β-α-α-β structure (42). Often, multiple copies (1-6) of the LysM motif are present in a LysM domain. While the molecular mechanism of binding is unknown, it is likely that the GlcNAc component of the peptidoglycan is involved in the binding. Many of the LysM bearing proteins are bacterial lysins, hence the name of the motif. These

lysins may be found in bacteriophages to cleave bacteria cell wall, or in bacteria as carefully controlled autolysins which are required for cell growth and division.

Not all the proteins that bind non-covalently to the cell surface bind to peptidoglycan. Some of them bind to lipoteichoic acids (LTA), and some bind to secondary cell wall polysaccharides (SCWP) (43). Two types of modules allow proteins to bind LTA: a choline binding domain, and a GW module. A choline binding domain is usually found at the C-terminus of the LTA-binding protein, and contains 5-10 imperfect repeats of ~20 amino acids which include several conserved aromatic residues (44). Proteins harboring a choline binding domain specifically bind to choline-modified LTA. For bacteria that do not modify their LTA with choline, a GW module is used to bind LTA instead. A GW motif is about 80 residues long and begins with a glycine and a tryptophan. Similar to choline binding domains, several repeats of the GW motif is often found at the C-terminal of the LTA-binding protein to ensure binding (45). Similar to proteins that contain a LysM domain, most of the LTA-binding proteins are amidases that hydrolyze the cell wall of the bacteria that synthesize them and are important for cell growth and division. As mentioned in section **1.2**, proteins that bind to SCWP typically contain a SLH module or a CWB2 module. The two major S-layer proteins of *B. anthracis*, Sap and EA1, each contains three SLH domains at the N-terminus. The three SLH domains fold into a pseudo-trimer that resembles a three-pronged spindle, where a conserved cationic residue within each "prong" is necessary for tethering the S-layer protein to the pyruvylated SCWP (46). Most other S-layer proteins also contain three copies of SLH or CWB2, and are likely to interact with SCWP in a similar manner.

1.5 Covalent attachment of surface proteins

In Gram-positive bacteria, surface proteins are covalently attached to the cell wall by the sortase enzymes (47), cysteine transpeptidases that catalyze the formation of a peptide bond between the surface protein and the peptide stem of the cell wall precursor molecule lipid II. The lipid II linked protein complex is then incorporated into the peptidoglycan by the transglycosylation and transpeptidation reactions that synthesize the cell wall. Proteins destined to be anchored covalently are first produced as precursor proteins with a N-terminal Sec-dependent signal peptide, and a C-terminal cell wall sorting signal (CWSS) that consists of a short sortase recognition motif, followed by a segment of hydrophobic amino acids, and a tail that contains mostly positively charged amino acids. The anchoring/sorting mechanism is best studied in the *S. aureus* sortase A (Sa-SrtA) system, which was first reported in 1999 by Schneewind and colleagues (48). In this system, a full length precursor protein bearing a N-terminal signal peptide is first directed to the cell membrane and translocated through the SecYEG channel after synthesis in the cytoplasm (Fig. 1.3). The signal peptide is cut off by a signal peptidase and the protein is partially embedded in the cell membrane by its C-terminal hydrophobic segment, while the positively charged tail presumably remains in the cytoplasm to retard secretion. The precursor protein is recruited to the membrane associated Sa-SrtA, of which the extracellular catalytic domain binds to the LP*X*TG motif (where *X* denotes any amino acid) within the CWSS of the precursor protein. A highly conserved active site cysteine residue in Sa-SrtA then nucleophilically attacks the backbone carbonyl carbon of the threonine residue in the LP*X*TG motif to cleave the scissile threonine-glycine peptide bond and to create a relatively stable sortase-protein complex via a thioacyl bond. Finally, lipid II [C_{55}-$(PO_3)_2$-MurNAc(L-Ala-D-iGlu-L-Lys(NH_2-Gly_5)-D-Ala-D-Ala] is recruited to the sortase-protein complex, and the

terminal amino group of the pentaglycine bridge that branches off from the peptide stem L-Lys nucleophically attacks the thioacyl linkage to create an isopeptide linked protein-lipid II product. Regular transglycosylation and transpeptidation reactions carried out by penicillin binding proteins (PBP) then incorporate this product to the peptidoglycan so that the protein is covalently attached to the cell wall. In other bacteria such as *B. anthracis* which contains *meso*-diaminopimelic acid instead of L-Lys(NH_2-Gly_5) in lipid II, the sortase-protein complex is resolved by the side chain amino group of *meso*-diaminopimelic acid instead. Also, different classes of sortase recognize different motifs in the CWSS. For example a class B sortase recognizes a NP(Q/K)(T/S)(N/G/S)(D/A) sorting signal motif.

Some Gram-positive bacteria display hair-like fibres known as pili on the cell surface to promote bacterial adhesion (47). Pili are composed of pilin subunits that are joined together through isopeptide bonds before being anchored to the peptidoglycan by sortases. The transpeptidation step that links pilin subunits together is also carried out by sortases, specifically class C sortases. In this reaction, the sorting signal motif in the pilin subunit is recognized and nucleophilically attacked by the class C sortase to form a thioacyl-linked pilin-sortase complex. A conserved lysine residue, usually found within the pilin motif (W*XXX*V*X*VYPK), on a different pilin protein then attacks the acyl bond to form an isopeptide bond between the sorting signal motif and the side chain of the lysine. This reaction happens many times to polymerize pilins into a pilus, until a minor pilin subunit is attached to the base of the pilus and subsequently to the peptidoglycan by a class A sortase.

1.6 Sortases

The sortase enzymes were first discovered by Schneewind and colleagues in *S. aureus* in a genetic screen for *S. aureus* mutants that failed to cleave the protein A CWSS (48). Since then, many more sortases and their substrates have been identified in other Gram-positive bacteria. Since many virulence factors are anchored to the cell wall by sortases, sortases have attracted great interest as potential drug targets (49,50). Moreover, they have been developed into powerful molecular tools to site-specifically attach proteins to a myriad of biomolecules (51). Sortase genes have been identified based on the presence of a conserved motif TL*X*TC, where *X* is any amino acid (52). With the exception of Mollicutes, it was discovered that sortase is present in all Gram-positive bacteria. Based on their primary sequences, sortases can be grouped into six classes A-F (47). Most bacteria contain more than one class of sortase, but the housekeeping class A or class E sortase is always present. For example, *S. aureus* has a class A sortase Sa-SrtA and a class B sortase Sa-SrtB, and *B. anthracis* has a class A sortase Ba-SrtA, a class B sortase Ba-SrtB, and a class D sortase Ba-SrtC.

Class A sortases are often regarded as the housekeeping sortases since they anchor the majority of surface proteins. These sortases are special in that each enzyme is capable of anchoring many functionally distinct proteins to the cell wall, whereas other classes of sortases have only a limited number of substrates. Also, unlike other classes of sortases, the genes encoding class A sortases are not genomically clustered with genes encoding their protein substrates. Class A sortases recognize a sorting signal motif containing the sequence LP*X*TG. Based on this, 21 surface proteins are predicted to be anchored by *S. aureus* SrtA, and in an extreme case, as many as 43 proteins are anchored by the class A sortase in *Listeria monocytogenes* (52,53). Since many surface proteins attached by class A sortases play key roles

in the infection process by promoting nutrient acquisition from the host, bacterial adhesion and immune evasion, class A sortases have attracted significant interest as potential drug targets (54).

The gene for the class B sortase is located within the iron-regulated surface determinant (*isd*) gene cluster of staphylococci, bacilli, listeria and clostridia (55). In *S. aureus*, the class B enzyme Sa-SrtB attaches only one protein, IsdC, to the cell wall. Unlike Sa-SrtA, Sa-SrtB and IsdC are only expressed under iron deplete conditions (56), since IsdC is important in iron acquisition. The sorting signal motif, NP(Q/K)(T/S)(N/G/S)(D/A), that class B sortases recognize is very different than the LP*X*TG sorting signal motif that class A sortases recognize. Moreover, it seems that at least in *S. aureus*, Sa-SrtB may attach its substrate at a different site within the cell wall that is not heavily cross-linked, so it is possible that Sa-SrtB may attach IsdC to preassembled peptidoglycan instead of lipid II (55,57).

As mentioned in the previous section, class C sortases are responsible for pili assembly. They are unique in that they accept the lysine side chain amino group on the pilin subunit as their secondary substrate, though in some cases such as in *Corynebacterium diphtheriae* they can also accept lipid II as their secondary substrate to anchor pili to the cell wall, albeit at a reduced rate (58). Class D, E and F sortases are not as extensively studied and little is known about them. So far only the *B. anthracis* class D sortase, Ba-SrtC, has been characterized (59,60). It is known to anchor two proteins, BasH and BasI, to the cell wall for spore formation under oxygen limiting conditions. Both BasH and BasI contain a LPNTA sorting signal, which is very similar to the class A sortase sorting signal LP*X*TG. Nevertheless, experiments have shown that Ba-SrtC cannot process a LPATG peptide and Ba-SrtA cannot process a LPNTA peptide, suggesting that they may have evolved a high degree of specificity for their respective sorting signals (59). Class E sortases are only found in high GC content bacteria. Since bacteria harboring class E sortases

do not contain class A sortases, and the genes encoding class E enzymes are not genomically clustered near their substrates, it has been postulated that class E sortases function as housekeeping sortases in high GC content bacteria (61). Unlike the class A enzymes, class E sortases recognize an LA*X*TG sorting signal. Class E sortases have been shown to anchor pili to the cell wall in *C. diphtheriae*, and display chaplin on the surface of *Streptomyces coelicolor,* which presumably mediate aerial hyphae formation. Lastly, actinobacteria such as *S. coelicolor* are predicted to contain class F sortases, but none of these enzymes have been characterized yet.

The catalytic mechanism of sortases has been studied in detail in Sa-SrtA. Structural, bioinformatics, and mutagenesis studies have shown that all sortases adopt an eight-strand beta barrel fold, where the active site contains three proximally positioned residues His120, Cys184, and Arg197 (Sa-SrtA numbering) (62-64). In Sa-SrtA, catalysis occurs through a ping-pong mechanism that is initiated when the thiol group of Cys184 nucleophilically attacks the carbonyl carbon of the threonine residue within the LP*X*TG motif of the precursor surface protein (65,66) (Fig. 1.4). This forms the first transient tetrahedral intermediate, which rearranges into a more stable thioacyl intermediate upon breakage of the threonine-glycine peptide bond and protonation of the leaving group glycine-peptide, presumably by His120. The terminal amine group of the pentaglycine bridge of lipid II, presumably deprotonated by His120, then nucleophilically attacks the carbonyl carbon of the threonine, forming a second transient tetrahedral intermediate that is resolved into a lipid II-linked protein product. Although mutagenesis studies have shown that Arg197 is critically important in catalysis, its role is not as clear as Cys184. Recently, substrate analog-bound structures of Sa-SrtB and Ba-SrtA (see Chapter 4) have been solved. Using molecular dynamics to model the thioacyl intermediate, it appears that Arg197 may be held in place by the side chain of the threonine within the sorting

signal such that its guanidino group is close to the carbonyl oxygen of threonine. This would facilitate catalysis by stabilizing the negative charge that appears on the high energy tetrahedral intermediates that form before and after the thioacyl intermediate (67). Although the presence of cysteine and histidine in the active site may imply that Sa-SrtA uses a thiolate-imidazolium ion-pair mechanism for catalysis (62), pKa measurements revealed that those two residues are mostly uncharged in solution, and that Sa-SrtA actually employs a reverse protonation mechanism in which only ~0.06% of the enzyme has the correct ionization states ready for catalysis (66,68), which might explain the very low enzymatic activity observed *in vitro*.

1.7 Antibiotic development

Since the discovery of penicillin in 1929 and its major deployment in the 1940s, many antibiotics have been discovered and used to treat bacterial infections (69). These antibiotics generally function in one of five ways. They inhibit cell wall synthesis, protein synthesis, DNA or RNA synthesis, folate synthesis, or disrupt the membrane (70). Compounds that inhibit cell wall synthesis include beta-lactam class drugs such as penicillin, methicillin, cephalosporins, and carbapenems, as well as glycopeptide class drugs such as vancomycin. Beta-lactams function by inhibiting penicillin-binding proteins (PBPs) so that the transpeptidation reaction required for cross-linking peptidoglycan strands cannot occur (71). Similarly, glycopeptides prevent cross-linking from happening, but they do so by binding to the terminal D-Ala-D-Ala dipeptide of the peptide stem of peptidoglycan (72). Many classes of compounds inhibit protein synthesis by binding either to the 50S or 30S subunit of bacterial ribosomes. 50S ribosome inhibitors include macrolides (e.g. erythromycin), streptogramins (e.g. dalfopristin), phenicols (e.g. chloramphenicol), and oxazolidinones (e.g. linezolid) (73,74). They work by physically blocking

either the initiation of protein translation or the translocation of peptidyl tRNAs, eventually triggering dissociation of the peptidyl tRNA. Examples of 30S ribosome inhibitors include tetracyclines and aminoglycosides (e.g. streptomycin). Tetracyclines work by blocking the access of aminoacyl tRNAs to the ribosome (75), while aminoglycosides promote mRNA-tRNA mismatching (76). Compounds that inhibit DNA or RNA synthesis include fluoroquinolones (e.g. ciprofloxacin) and rifamycins. Fluoroquinolones work by binding to topoisomerase II and IV, trapping these enzymes at the DNA cleavage stage and preventing strands from rejoining (77), leading to inhibition of DNA synthesis. Rifamycins bind stably to the beta-subunit of RNA polymerase at the initiation stage, thereby prevent RNA synthesis and downstream protein synthesis (78). Folate biosynthesis is a pathway unique in bacteria, and it is inhibited by trimethoprim and a sulfonamide class compound sulfamethoxazole, and they inhibit different stages in the folate synthesis pathway (79). Since folate is essential in nucleic acid synthesis, inhibiting folate synthesis in bacteria leads to bacteriostasis and eventual cell death. Lastly, lipopeptides such as daptomycin can be inserted into the bacterial cell membrane, forming pores that leak ions therefore leading to rapid depolarization of the membrane and cell death (80).

Many novel classes of compounds were discovered between 1940 and 1960, a period that Walsh and Wencewicz termed the "Golden Age of discovery" (69). However, no major classes of compound were introduced since 1962 until 2000. Although new generations of existing classes of antibiotics are constantly being developed, it is unlikely that modifications of these antibiotic classes can go on indefinitely. Fig. 1.5 shows the timeline of antibiotic deployment and the evolution of antibiotic resistance. As multidrug resistant bacteria are on the rise, new classes of antibiotic are needed urgently. One of the most problematic multidrug resistant bacteria is *Staphylococcus aureu*s. *S. aureus* is a leading cause of hospital- and community-acquired

infections in the United States. A wide spectrum of diseases can be caused by this pathogen, ranging from skin and soft tissue infections to potentially life-threatening diseases such as pneumonia, endocarditis, meningitis, and toxic shock syndrome (81,82). The multidrug resistant strain, methicillin-resistant *S. aureus* (MRSA), has made a significant impact on mortality, hospital stay, and associated costs (83). A surveillance conducted by Centers for Disease Control and Prevention (CDC) indicates that in 2011, there were 80,000 cases of invasive MRSA infection in the United States, which led to more than 11,000 deaths (3). Unfortunately and inevitably, vancomycin, which was once regarded as the drug of "last resort," has become less effective in treating MRSA infections due to the emergence of vancomycin-intermediate and vancomycin-resistant *S. aureus* (VISA and VRSA) (84). Recently, a few clinical isolates of *S. aureus* were also found to be resistant to even newer classes of antibiotics such as linezolid and daptomycin (85,86). New therapeutic agents are therefore desperately needed to treat MRSA infections, preferably ones that do not lead to rapid emergence of drug-resistant strains.

The *S. aureus* sortase A (Sa-SrtA) protein may be an excellent choice as a new target for drug development. As mentioned previously, sortases are transpeptidases that anchor surface proteins to the cell wall covalently. In *S. aureus*, many of the surface proteins are virulence factors such as protein A, which helps the bacteria to evade our immune system by sequestering antibodies, fibronectin binding proteins A and B (FnBPA and FnBPB), which allow *S. aureus* to adhere to different host tissues, and various iron-regulated surface determinant (Isd) proteins, important for the bacteria to acquire iron from within the host (54). By designing an inhibitor that targets Sa-SrtA, we may prevent many virulence factors from being displayed on the bacterial surface and thus promote bacterial clearance by the host. To test this idea, *srtA*$^{-}$ strains of *S. aureus* have been generated and their infectivity tested in mice, which shows that SrtA-

deficient mutants are less capable of killing mice in a lethal-infection experiment (100% of mice survived an intravenous injection of 6 x 10^6 cfu sortase mutant, whereas 80% mice died from an injection of 6 x 10^6 cfu isogenic Newman strain after 12 days) (87), and that there is at least a 2-log decrease in cfu of *S. aureus* isolated from the kidney, joints, blood, and heart when mice were challenged with SrtA-deficient mutants (87-89). Attractively, *srtA*$^-$ strains of *S. aureus* do not exhibit impaired growth in culture medium outside of the host (48), suggesting that SrtA inhibitors will not kill the bacteria directly, but would prevent display of their virulence factors and allow the host to clear the infection. It has been suggested that this type of anti-infective agent might impose weaker selective pressure, thus retarding the emergence of resistant strains (50,90).

Realizing the potential of SrtA inhibitors, many efforts have been made to discover compounds that inhibit Sa-SrtA (49). These include screenings and structure activity relationship studies of natural product and compound libraries, and rationally designing peptidomimetics and small molecules (91-94). *In silico* experiments to screen for small molecule inhibitors or improve on previously discovered Sa-SrtA inhibitors have also been performed, which will be described in chapters 2 and 3. Currently, the most potent Sa-SrtA inhibitor is a pyridazinone class small molecule discovered in my dissertation work (chapter 3), which had an IC_{50} value of 21 nM. Like many other Sa-SrtA inhibitors, this compound is also capable of inhibiting sortase A from *B. anthracis* with a low IC_{50} value of 450 nM. Since studies of *srtA*$^-$ strains of several other Gram-positive pathogens such as *B. anthracis*, *L. monocytogenes*, and *S. pneumonia* have shown that they also had decreased infectivity (95-101), compounds designed for inhibiting Sa-SrtA, with some modification, may be used as a general sortase inhibitor to treat a broad range of Gram-positive bacterial infections.

1.8 Scope of dissertation

Staphylococcus aureus is the leading cause of hospital- and community-acquired infections in the United States. The rapid emergence of multidrug-resistant strains has created an urgent need for new antibiotics. *S. aureus* and other Gram-positive pathogens use sortase enzymes to display surface virulence factors. Compounds that inhibit sortase might therefore function as potent anti-infective agents. This dissertation describes the structure-based drug discovery methods that I used to look for potent *S. aureus* sortase A (Sa-SrtA) inhibitors, and more fundamental studies of the mechanism of catalysis. To begin, I performed an *in silico* screen using the molecular docking program GLIDE to look for compounds that should bind Sa-SrtA tightly. Unlike conventional docking methods, to increase the likelihood of finding true binders, I employed the "relaxed complex scheme," in which I docked the compounds to both the NMR structure and representative molecular dynamics simulated structures of Sa-SrtA. This allowed me to account for protein flexibility during docking, which is very important in finding Sa-SrtA inhibitors since the active site of Sa-SrtA contains two flexible loops. Details about the methods and results of this work are described in Chapter 2. In the second part of the quest to discover SrtA inhibitors, I used structural and computational methods to improve the potency of

ì □ Þ □ ì □ ì □

ì □ ì □ ì □ ì □ ì

against Sa-SrtA, first using GLIDE and then using experimental approaches (Chapter 3). This led to improved compounds that have IC_{50}'s in the nanomolar range, or ~70-fold more potent than the original compound. *S. aureus* SrtA inhibitors can potentially be developed into pan-sortase inhibitors to treat infections caused by a wide range of Gram-positive pathogens. To facilitate the rational design of pan-sortase inhibitors, and to gain a deeper understanding of substrate

recognition mechanism among different sortases, I solved the structure of *Bacillus anthracis* SrtA (Ba-SrtA) bound to a substrate analog using NMR methods (Chapter 4). Interestingly, the structure reveals that unlike Sa-SrtA, Ba-SrtA has a structured N-terminal extension that seems to regulate substrate entry. Moreover, I discovered that one of the active site loops undergoes disorder to order transition to create a new surface presumably for lipid II binding. This new structural information of Ba-SrtA will be useful in the rational design of potent sortase inhibitors that have broadened species specificity.

1.9 Figures

Figure 1.1. Overview of the cell wall structure. A general depiction of the cell walls of *Staphylococcus aureus* (left) and *Bacillus anthracis* (right) are shown. *S. aureus* cell wall is composed of peptidoglycan and a capsule on the surface. It contains lipoteichoic acids (LTA) which are bound to the membrane, and wall teichoic acids (WTA) that are tethered to the peptidoglycan. It also contains various membrane proteins, and cell wall-bound proteins that are anchored to the peptidoglycan by sortase enzymes. The structure of *B. anthracis* cell wall is similar, except that it does not contain WTA, but contains a layer of secondary cell wall polysaccharide (SCWP) on top of the peptidoglycan. SCWP serves as an anchor point for S-layer proteins to form a crystalline S-layer.

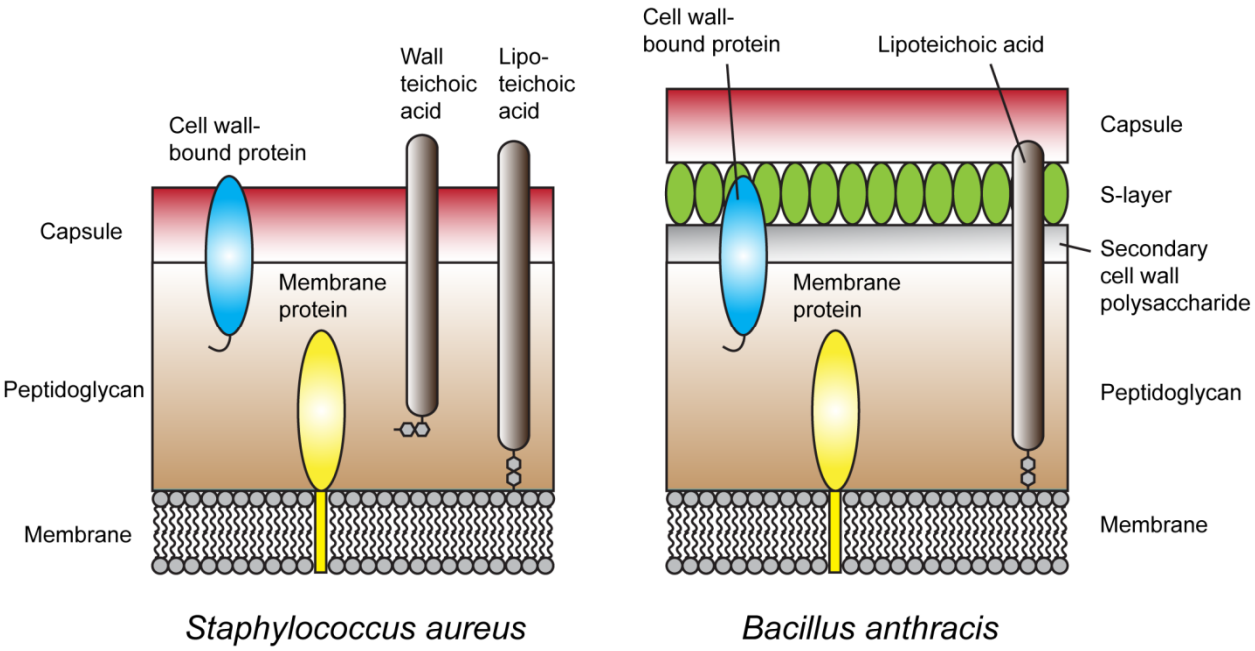


Figure 1.2. Peptidoglycan structure with cell wall binding proteins. Schematics of the *Staphylococcus aureus* (left) and *Bacillus anthracis* (right) peptidoglycan structures. Peptidoglycan is composed of glycan strands that contain *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) linked by β-1→4 bonds, and peptide stems that are connected to the glycan strands through MurNAc. Individual glycan strands are linked together by crosslinking between the third amino acid of one peptide stem and the fourth amino acid (D-Ala) of a different peptide stem from a different strand. Crosslinking may happen through a crossbridge (Gly_5 crossbridge within *S. aureus*) or directly through the side chain of the third amino acid (*meso*-diaminopimelic acid (*m*-DAP) in *B. anthracis*).

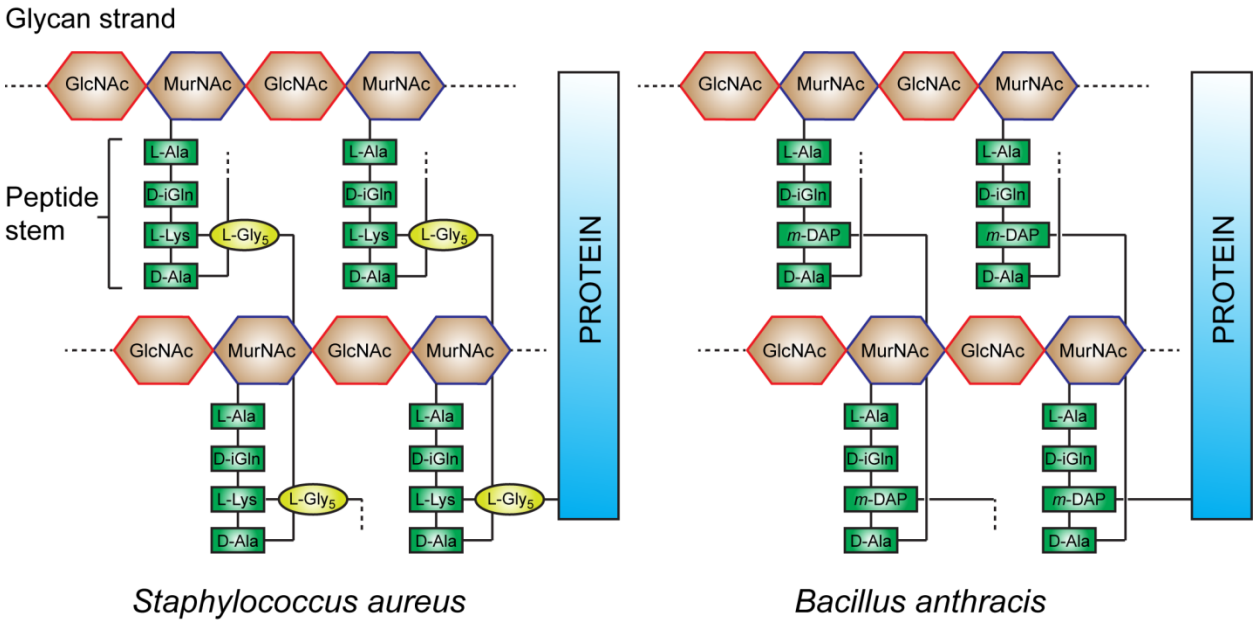


Figure 1.3. Overview of *S. aureus* sortase A-mediated cell wall protein anchoring. A precursor surface protein contains a signal peptide at the N-terminus, followed by its functional domain and a cell wall sorting signal (CWSS). It is 1) synthesized in the cytoplasm and directed to the Sec translocase, at which point the signal peptide is cleaved and the protein is inserted into the cell membrane; 2) sortase A recognizes the LPXTG motif within the CWSS and nucleophilically attacks the backbone carbonyl of Thr to 3) cleave the Thr-Gly bond and form a surface protein-sortase complex; 4) sortase recognizes peptidoglycan precursor lipid II and catalyzes the formation of a new peptide bond between Thr and the terminal amino group of the Gly_5 crossbridge of lipid II, 5) leading to formation of a surface protein-lipid II complex, 6) which is subsequently incorporated into the peptidoglycan during cell wall synthesis.

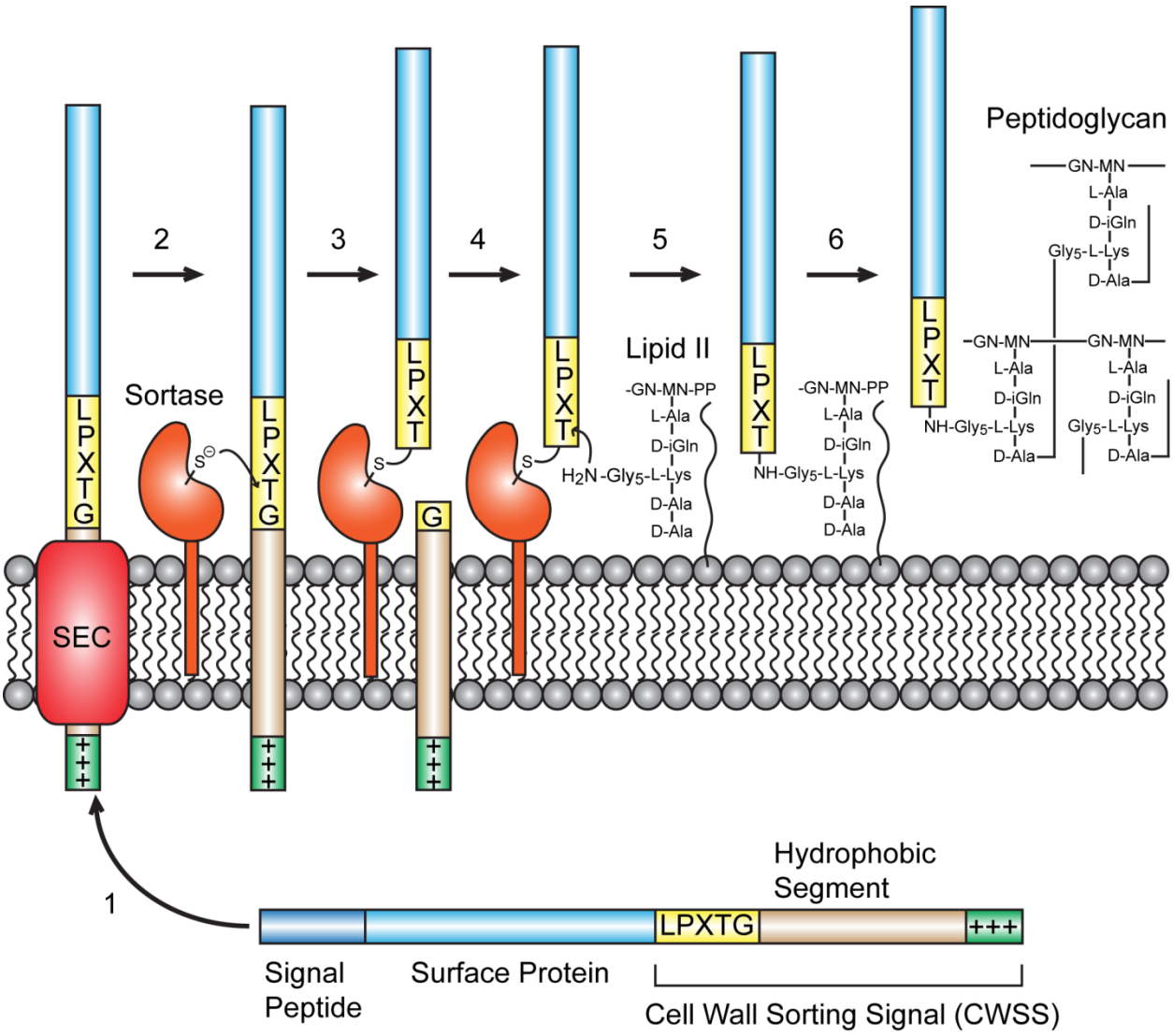

Peptidoglycan
Sortase
Lipid II
SEC
LPXTG
G
1
2
3
4
5
6
Hydrophobic
Segment
LPXTG
+++
Signal
Peptide
Surface Protein
Cell Wall Sorting Signal (CWSS)

Figure 1.4. Proposed chemical mechanism of the sortase A catalyzed transpeptidation reaction. This schematic shows the proposed catalytic mechanism of *S. aureus* sortase A catalyzing the formation of a peptide bond between the sorting signal substrate LPXTG and the Gly_5 crossbridge peptide in lipid II. 1) 0.06% sortase A exists in the catalytic competent state, in which 2) the thiolate of Cys184 nucleophilically attacks the backbone carbonyl of Thr, forming the first tetrahedral intermediate. 3) The tetrahedral intermediate rearranges into a more stable thioacyl intermediate when the Thr-Gly scissile bond is broken, a step that is thought to be assisted by protonation of the leaving group by His120. 4) The Gly_5 crossbridge peptide enters the active site, and its terminal amino group nucleophilically attacks the carbon in the thioacyl bond, a step that may be assisted by deprotonation by His120, forming a second tetrahedral intermediate. 5) The high energy tetrahedral intermediate rearranges to break the thioacyl bond between the peptide and the enzyme, resulting in the formation of a new peptide bond between the Thr and Gly_5. Since Arg197 is position close to Thr in the structure, Arg197 is thought to assist catalysis by stabilizing the high energy oxyanion that occurs in the transition states. (Figure modified from reference (102) with permission)

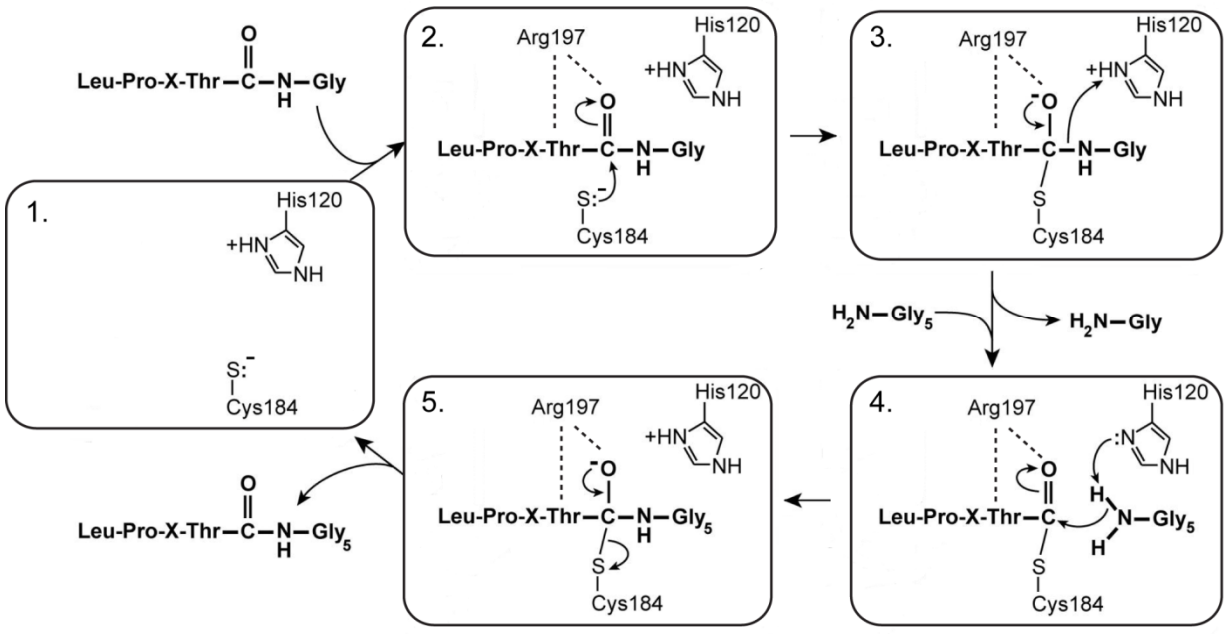
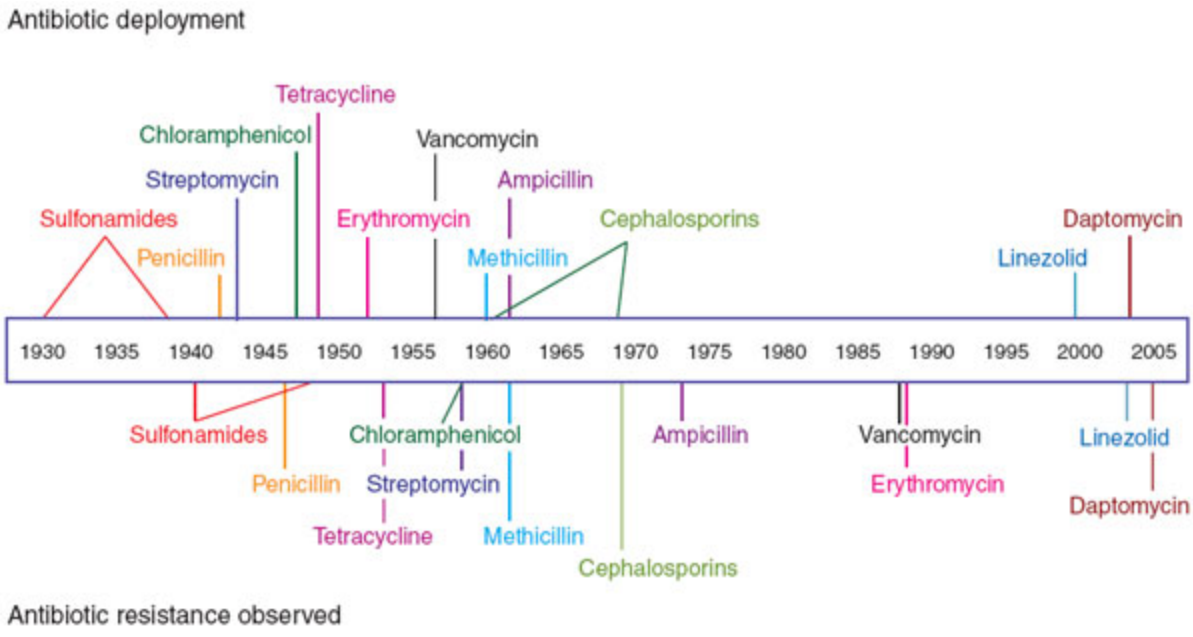

1.
His120
Cys184
Leu-Pro-X-Thr—C—N—Gly
2.
Arg197
His120
Leu-Pro-X-Thr—C—N—Gly
Cys184
3.
Arg197
His120
Leu-Pro-X-Thr—C—N—Gly
Cys184
H2N—Gly5
H2N—Gly
4.
Arg197
His120
Leu-Pro-X-Thr—C
N—Gly5
Cys184
5.
Arg197
His120
Leu-Pro-X-Thr—C—N—Gly5
Cys184
Leu-Pro-X-Thr—C—N—Gly5

Figure 1.5. Timeline of antibiotic deployment and the evolution of antibiotic resistance. The upper portion of the timeline shows the year each antibiotic was deployed, and the bottom portion shows the year resistance to each antibiotic was observed. This timeline also highlights the "Golden Age of discovery" between 1940s and 1960s, and that there were no new major classes of compound introduced since 1962 until 2000. (Figure adapted from reference (90) with permission)



1.10 References

1. Rice, L. B. (2008) *J Infect Dis* **197**, 1079-1081
2. Boucher, H. W., Talbot, G. H., Bradley, J. S., Edwards, J. E., Gilbert, D., Rice, L. B., Scheld, M., Spellberg, B., and Bartlett, J. (2009) *Clin Infect Dis* **48**, 1-12
3. Centers for Disease Control and Prevention. Active Bacterial Core Surveillance Report, Emerging Infections Program Network, Methicillin-Resistant Staphylococcus aureus, 2011. http://www.cdc.gov/abcs/reports-findings/survreports/mrsa11.html.
4. Goley, E. D., Comolli, L. R., Fero, K. E., Downing, K. H., and Shapiro, L. (2010) *Mol Microbiol* **77**, 56-73
5. Harold, F. M. (1990) *Microbiol Rev* **54**, 381-431
6. Matias, V. R., Al-Amoudi, A., Dubochet, J., and Beveridge, T. J. (2003) *J Bacteriol* **185**, 6112-6118
7. Matias, V. R., and Beveridge, T. J. (2005) *Mol Microbiol* **56**, 240-251
8. Matias, V. R., and Beveridge, T. J. (2006) *J Bacteriol* **188**, 1011-1021
9. Vollmer, W., Blanot, D., and de Pedro, M. A. (2008) *FEMS Microbiol Rev* **32**, 149-167
10. Weidenmaier, C., and Peschel, A. (2008) *Nat Rev Microbiol* **6**, 276-287
11. Brown, S., Santa Maria, J. P., Jr., and Walker, S. (2013) *Annu Rev Microbiol* **67**, 313-336
12. Weidenmaier, C., Kokai-Kun, J. F., Kristian, S. A., Chanturiya, T., Kalbacher, H., Gross, M., Nicholson, G., Neumeister, B., Mond, J. J., and Peschel, A. (2004) *Nat Med* **10**, 243-245
13. Weidenmaier, C., Peschel, A., Xiong, Y. Q., Kristian, S. A., Dietz, K., Yeaman, M. R., and Bayer, A. S. (2005) *J Infect Dis* **191**, 1771-1777
14. Heptinstall, S., Archibald, A. R., and Baddiley, J. (1970) *Nature* **225**, 519-521

15. Grundling, A., and Schneewind, O. (2007) *Proc Natl Acad Sci U S A* **104**, 8478-8483

16. Schneewind, O., and Missiakas, D. M. (2012) *Philos Trans R Soc Lond B Biol Sci* **367**, 1123-1139

17. Fagan, R. P., and Fairweather, N. F. (2014) *Nat Rev Microbiol* **12**, 211-222

18. Mignot, T., Mesnage, S., Couture-Tosi, E., Mock, M., and Fouet, A. (2002) *Mol Microbiol* **43**, 1615-1627

19. Ahn, J. S., Chandramohan, L., Liou, L. E., and Bayles, K. W. (2006) *Mol Microbiol* **62**, 1158-1169

20. Verbrugh, H. A., Hoidal, J. R., Nguyen, B. Y., Verhoef, J., Quie, P. G., and Peterson, P. K. (1982) *J Clin Invest* **69**, 63-74

21. Scorpio, A., Chabot, D. J., Day, W. A., O'Brien D, K., Vietri, N. J., Itoh, Y., Mohamadzadeh, M., and Friedlander, A. M. (2007) *Antimicrob Agents Chemother* **51**, 215-222

22. Buzzola, F. R., Alvarez, L. P., Tuchscherr, L. P., Barbagelata, M. S., Lattar, S. M., Calvinho, L., and Sordelli, D. O. (2007) *Infect Immun* **75**, 886-891

23. Tuchscherr, L. P., Buzzola, F. R., Alvarez, L. P., Caccuri, R. L., Lee, J. C., and Sordelli, D. O. (2005) *Infect Immun* **73**, 7932-7937

24. Freudl, R. (2013) *Res Microbiol* **164**, 664-674

25. du Plessis, D. J., Nouwen, N., and Driessen, A. J. (2011) *Biochim Biophys Acta* **1808**, 851-865

26. Lycklama, A. N. J. A., and Driessen, A. J. (2012) *Philos Trans R Soc Lond B Biol Sci* **367**, 1016-1028

27. Muller, J. P., Ozegowski, J., Vettermann, S., Swaving, J., Van Wely, K. H., and Driessen, A. J. (2000) *Biochem J* **348 Pt 2**, 367-373

28. Herbort, M., Klein, M., Manting, E. H., Driessen, A. J., and Freudl, R. (1999) *J Bacteriol* **181**, 493-500

29. Lyon, W. R., Gibson, C. M., and Caparon, M. G. (1998) *EMBO J* **17**, 6263-6275

30. Wu, S. C., Ye, R., Wu, X. C., Ng, S. C., and Wong, S. L. (1998) *J Bacteriol* **180**, 2830-2835

31. Rigel, N. W., and Braunstein, M. (2008) *Mol Microbiol* **69**, 291-302

32. Rigel, N. W., Gibbons, H. S., McCann, J. R., McDonough, J. A., Kurtz, S., and Braunstein, M. (2009) *J Biol Chem* **284**, 9927-9936

33. Archambaud, C., Nahori, M. A., Pizarro-Cerda, J., Cossart, P., and Dussurget, O. (2006) *J Biol Chem* **281**, 31812-31822

34. Bensing, B. A., Siboo, I. R., and Sullam, P. M. (2007) *J Bacteriol* **189**, 3846-3854

35. Takamatsu, D., Bensing, B. A., and Sullam, P. M. (2005) *J Bacteriol* **187**, 3878-3883

36. Goosens, V. J., Monteferrante, C. G., and van Dijl, J. M. (2013) *Biochim Biophys Acta* **1843**, 1698-1706

37. Biswas, L., Biswas, R., Nerz, C., Ohlsen, K., Schlag, M., Schafer, T., Lamkemeyer, T., Ziebandt, A. K., Hantke, K., Rosenstein, R., and Gotz, F. (2009) *J Bacteriol* **191**, 5921-5929

38. Widdick, D. A., Dilks, K., Chandra, G., Bottrill, A., Naldrett, M., Pohlschroder, M., and Palmer, T. (2006) *Proc Natl Acad Sci U S A* **103**, 17927-17932

39. Pallen, M. J. (2002) *Trends Microbiol* **10**, 209-212

40. Renshaw, P. S., Lightbody, K. L., Veverka, V., Muskett, F. W., Kelly, G., Frenkiel, T. A., Gordon, S. V., Hewinson, R. G., Burke, B., Norman, J., Williamson, R. A., and Carr, M. D. (2005) *EMBO J* **24**, 2491-2498

41. Buist, G., Steen, A., Kok, J., and Kuipers, O. P. (2008) *Mol Microbiol* **68**, 838-847

42. Bateman, A., and Bycroft, M. (2000) *J Mol Biol* **299**, 1113-1119

43. Scott, J. R., and Barnett, T. C. (2006) *Annu Rev Microbiol* **60**, 397-423

44. Desvaux, M., Dumas, E., Chafsey, I., and Hebraud, M. (2006) *FEMS Microbiol Lett* **256**, 1-15

45. Jonquieres, R., Bierne, H., Fiedler, F., Gounon, P., and Cossart, P. (1999) *Mol Microbiol* **34**, 902-914

46. Kern, J., Wilton, R., Zhang, R., Binkowski, T. A., Joachimiak, A., and Schneewind, O. (2011) *J Biol Chem* **286**, 26042-26049

47. Spirig, T., Weiner, E. M., and Clubb, R. T. (2011) *Mol Microbiol* **82**, 1044-1059

48. Mazmanian, S. K., Liu, G., Ton-That, H., and Schneewind, O. (1999) *Science* **285**, 760-763

49. Suree, N., Jung, M. E., and Clubb, R. T. (2007) *Mini Rev Med Chem* **7**, 991-1000

50. Maresso, A. W., and Schneewind, O. (2008) *Pharmacol Rev* **60**, 128-141

51. Popp, M. W., and Ploegh, H. L. (2011) *Angew Chem Int Ed Engl* **50**, 5024-5032

52. Comfort, D., and Clubb, R. T. (2004) *Infect Immun* **72**, 2710-2722

53. Pallen, M. J., Lam, A. C., Antonio, M., and Dunbar, K. (2001) *Trends Microbiol* **9**, 97-102

54. Marraffini, L. A., Dedent, A. C., and Schneewind, O. (2006) *Microbiol Mol Biol Rev* **70**, 192-221

55. Mazmanian, S. K., Skaar, E. P., Gaspar, A. H., Humayun, M., Gornicki, P., Jelenska, J., Joachmiak, A., Missiakas, D. M., and Schneewind, O. (2003) *Science* **299**, 906-909

56. Mazmanian, S. K., Ton-That, H., Su, K., and Schneewind, O. (2002) *Proc Natl Acad Sci U S A* **99**, 2293-2298

57. Marraffini, L. A., and Schneewind, O. (2005) *J Biol Chem* **280**, 16263-16271

58. Swaminathan, A., Mandlik, A., Swierczynski, A., Gaspar, A., Das, A., and Ton-That, H. (2007) *Mol Microbiol* **66**, 961-974

59. Marraffini, L. A., and Schneewind, O. (2006) *Mol Microbiol* **62**, 1402-1417

60. Marraffini, L. A., and Schneewind, O. (2007) *J Bacteriol* **189**, 6425-6436

61. Ton-That, H., and Schneewind, O. (2003) *Mol Microbiol* **50**, 1429-1438

62. Ton-That, H., Mazmanian, S. K., Alksne, L., and Schneewind, O. (2002) *J Biol Chem* **277**, 7447-7452

63. Marraffini, L. A., Ton-That, H., Zong, Y., Narayana, S. V., and Schneewind, O. (2004) *J Biol Chem* **279**, 37763-37770

64. Frankel, B. A., Tong, Y., Bentley, M. L., Fitzgerald, M. C., and McCafferty, D. G. (2007) *Biochemistry* **46**, 7269-7278

65. Huang, X., Aulabaugh, A., Ding, W., Kapoor, B., Alksne, L., Tabei, K., and Ellestad, G. (2003) *Biochemistry* **42**, 11307-11315

66. Frankel, B. A., Kruger, R. G., Robinson, D. E., Kelleher, N. L., and McCafferty, D. G. (2005) *Biochemistry* **44**, 11188-11200

67. Jacobitz, A. W., Wereszczynski, J., Yi, S. W., Amer, B. R., Huang, G. L., Nguyen, A. V., Sawaya, M. R., Jung, M. E., McCammon, J. A., and Clubb, R. T. (2014) *J Biol Chem* **289**, 8891-8902

68. Connolly, K. M., Smith, B. T., Pilpa, R., Ilangovan, U., Jung, M. E., and Clubb, R. T. (2003) *J Biol Chem* **278**, 34061-34065

69. Walsh, C. T., and Wencewicz, T. A. (2014) *J Antibiot (Tokyo)* **67**, 7-22

70. Kohanski, M. A., Dwyer, D. J., and Collins, J. J. (2010) *Nat Rev Microbiol* **8**, 423-435

71. Holtje, J. V. (1998) *Microbiol Mol Biol Rev* **62**, 181-203

72. Kahne, D., Leimkuhler, C., Lu, W., and Walsh, C. (2005) *Chem Rev* **105**, 425-448

73. Katz, L., and Ashley, G. W. (2005) *Chem Rev* **105**, 499-528

74. Mukhtar, T. A., and Wright, G. D. (2005) *Chem Rev* **105**, 529-542

75. Chopra, I., and Roberts, M. (2001) *Microbiol Mol Biol Rev* **65**, 232-260 ; second page, table of contents

76. Davies, J., Gorini, L., and Davis, B. D. (1965) *Mol Pharmacol* **1**, 93-106

77. Chen, C. R., Malik, M., Snyder, M., and Drlica, K. (1996) *J Mol Biol* **258**, 627-637

78. Campbell, E. A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., and Darst, S. A. (2001) *Cell* **104**, 901-912

79. Smilack, J. D. (1999) *Mayo Clin Proc* **74**, 730-734

80. Baltz, R. H. (2009) *Curr Opin Chem Biol* **13**, 144-151

81. Lowy, F. D. (1998) *N Engl J Med* **339**, 520-532

82. Ippolito, G., Leone, S., Lauria, F. N., Nicastri, E., and Wenzel, R. P. (2010) *Int J Infect Dis* **14 Suppl 4**, S7-11

83. Lee, B. Y., Singh, A., David, M. Z., Bartsch, S. M., Slayton, R. B., Huang, S. S., Zimmer, S. M., Potter, M. A., Macal, C. M., Lauderdale, D. S., Miller, L. G., and Daum, R. S. (2013) *Clin Microbiol Infect* **19**, 528-536

84. Sakoulas, G., and Moellering, R. C., Jr. (2008) *Clin Infect Dis* **46 Suppl 5**, S360-367

85. Marty, F. M., Yeh, W. W., Wennersten, C. B., Venkataraman, L., Albano, E., Alyea, E. P., Gold, H. S., Baden, L. R., and Pillai, S. K. (2006) *J Clin Microbiol* **44**, 595-597

86. Ikeda-Dantsuji, Y., Hanaki, H., Sakai, F., Tomono, K., Takesue, Y., Honda, J., Nonomiya, Y., Suwabe, A., Nagura, O., Yanagihara, K., Mikamo, H., Fukuchi, K., Kaku, M., Kohno, S., Yanagisawa, C., Nakae, T., Yoshida, K., and Niki, Y. (2011) *J Infect Chemother* **17**, 45-51

87. Jonsson, I. M., Mazmanian, S. K., Schneewind, O., Verdrengh, M., Bremell, T., and Tarkowski, A. (2002) *J Infect Dis* **185**, 1417-1424

88. Weiss, W. J., Lenoy, E., Murphy, T., Tardio, L., Burgio, P., Projan, S. J., Schneewind, O., and Alksne, L. (2004) *J Antimicrob Chemother* **53**, 480-486

89. Mazmanian, S. K., Liu, G., Jensen, E. R., Lenoy, E., and Schneewind, O. (2000) *Proc Natl Acad Sci U S A* **97**, 5510-5515

90. Clatworthy, A. E., Pierson, E., and Hung, D. T. (2007) *Nat Chem Biol* **3**, 541-548

91. Maresso, A. W., Wu, R., Kern, J. W., Zhang, R., Janik, D., Missiakas, D. M., Duban, M. E., Joachimiak, A., and Schneewind, O. (2007) *J Biol Chem* **282**, 23129-23139

92. Suree, N., Yi, S. W., Thieu, W., Marohn, M., Damoiseaux, R., Chan, A., Jung, M. E., and Clubb, R. T. (2009) *Bioorg Med Chem* **17**, 7174-7185

93. Oh, K. B., Kim, S. H., Lee, J., Cho, W. J., Lee, T., and Kim, S. (2004) *J Med Chem* **47**, 2418-2421

94. Oh, K. B., Nam, K. W., Ahn, H., Shin, J., Kim, S., and Mar, W. (2010) *Biochem Biophys Res Commun* **396**, 440-444

95. Zink, S. D., and Burns, D. L. (2005) *Infect Immun* **73**, 5222-5228

96. Bierne, H., Mazmanian, S. K., Trost, M., Pucciarelli, M. G., Liu, G., Dehoux, P., Jansch, L., Garcia-del Portillo, F., Schneewind, O., and Cossart, P. (2002) *Mol Microbiol* **43**, 869-881

97. Garandeau, C., Reglier-Poupet, H., Dubail, I., Beretti, J. L., Berche, P., and Charbit, A. (2002) *Infect Immun* **70**, 1382-1390

98. Paterson, G. K., and Mitchell, T. J. (2006) *Microbes Infect* **8**, 145-153

99. Chen, S., Paterson, G. K., Tong, H. H., Mitchell, T. J., and DeMaria, T. F. (2005) *FEMS Microbiol Lett* **253**, 151-154

100. Kharat, A. S., and Tomasz, A. (2003) *Infect Immun* **71**, 2758-2765

101. Heilbronner, S., Hanses, F., Monk, I. R., Speziale, P., and Foster, T. J. (2013) *Microbiology* **159**, 2141-2152

102. Suree, N., Liew, C. K., Villareal, V. A., Thieu, W., Fadeev, E. A., Clemens, J. J., Jung, M. E., and Clubb, R. T. (2009) *J Biol Chem* **284**, 24465-24477

Chapter 2

Discovery of *Staphylococcus aureus* Sortase A Inhibitors Using Virtual Screening and the Relaxed Complex Scheme

2.1 Overview

Staphylococcus aureus is the leading cause of hospital-acquired infections in the United States. The emergence of multi-drug resistant strains of *S. aureus* has created an urgent need for new antibiotics. *S. aureus* uses the sortase A (SrtA) enzyme to display surface virulence factors suggesting that compounds that inhibit its activity will function as potent anti-infective agents. This chapter presents the identification of several inhibitors of SrtA using virtual screening methods that employ the relaxed complex scheme (RCS), an advanced computer-docking methodology that accounts for protein receptor flexibility. Experimental testing validates that several compounds identified in the screen inhibit the activity of SrtA. A lead compound based on the 2-phenyl-2,3-dihydro-1*H*-perimidine scaffold is particularly promising and its binding mechanism was further investigated using molecular dynamics simulations and by conducting preliminary structure activity relationship studies. In the end this work is compared to a previous virtual screen conducted by a different group, which shows that our virtual screen that uses the RCS is more successful than traditional docking.

This chapter is a reformatted version of "Discovery of *Staphylococcus aureus* Sortase A Inhibitors Using Virtual Screening and the Relaxed Complex Scheme" (Chan, A.H., Wereszczynski, J., Amer, B.R., Yi, S.W., Jung, M.E., McCammon, J.A., and Clubb, R.T. (2013) *Chem Biol Drug Des* **82**, 418-28).

2.2 Introduction

Staphylococcus aureus is a leading cause of hospital- and community-acquired infections in the United States and produces a wide spectrum of diseases, ranging from minor skin

infections to osteomyelitis, meningitis, endocarditis, septicemia, and toxic shock syndrome (1,2). The widespread occurrence of methicillin-resistant *S. aureus* (MRSA), which is often resistant to many commonly used antibiotics in addition to methicillin (3), makes treatment difficult. In 2011, there were 80,000 cases of invasive MRSA infection in the United States, which resulted in more than 11,000 deaths (4). The effectiveness of vancomycin, which was once regarded as a drug of last resort to treat MRSA infections, has been marginalized by the emergence of vancomycin-resistant strains (5). Moreover, *S. aureus* resistance to newer generation drugs such as linezolid and daptomycin has also now been reported (6,7). This creates an urgent need for new therapeutic agents to treat MRSA infections, preferably ones that do not lead to rapid emergence of drug-resistant strains.

One potential attractive approach to treat infections caused by *S. aureus* and other pathogens is to use small molecules that effectively strip the bacteria of their surface proteins, which frequently function as virulence factors (8). *S. aureus* and many other Gram-positive pathogens use sortase enzymes to anchor surface proteins to their cell walls (9-11). In *S. aureus*, 21 distinct surface proteins are anchored to the cell wall by the extracellular sortase A (SrtA) enzyme (12). This cysteine transpeptidase catalyzes the formation of a peptide bond between a cell wall sorting signal located at the C-terminal end of the precursor surface protein and the cell wall precursor molecule lipid-II (10). The lipid-II linked protein product is then incorporated into the peptidoglycan by the transglycosylation and transpeptidation reactions that synthesize the cell wall (10). Many surface proteins attached to the cell wall by SrtA are virulence factors that play key roles in the infection process by promoting nutrient acquisition from the host, bacterial adhesion and immune evasion (12). Disrupting the display of these proteins by blocking the activity of SrtA using a small molecule could therefore effectively reduce bacterial virulence and

thus promote bacterial clearance by the host. Indeed, numerous animal model studies of *S. aureus* infection have shown that *srtA*$^-$ strains of *S. aureus* are significantly attenuated in their virulence, underscoring the therapeutic potential of a small molecule SrtA inhibitor (13-17). Attractively, SrtA inhibitors may also be less likely to induce selective pressures that lead to drug resistance, as *srtA*$^-$ strains do not exhibit impaired growth outside of their human host in culture medium (18).

A number of different strategies have been employed to search for sortase inhibitors (8,19). These include screening natural products (20-32) and small compound libraries (33-36), as well as synthesizing rationally designed peptidomimetics and small molecules (37-42). Structures of SrtA in its apo- and substrate-bound forms (43-45) have now been determined enabling pharmacophore and three dimensional quantitative structure-activity relationships to be established for a select number of inhibitors (46,47). Currently this structural information has been employed in one virtual screen for sortase inhibitors, which made use of the crystal structure of SrtA determined in its unbound state (48). However, virtual docking efforts were hindered because the structure used in this study exhibited significant conformational heterogeneity and mobility, presumably because the protein was not co-crystallized with its sorting signal substrate. In subsequent work, our group determined the three-dimensional structure of SrtA bound to its sorting signal substrate. This new structure may be better suited for virtual screening approaches as its active site becomes conformationally ordered, and undergoes substantial changes in its structure, upon binding the substrate (44,49-51). We therefore used it as a starting point for virtual screening effort in which the relaxed complex scheme (RCS) method was used to account for receptor and ligand flexibility during docking. Experimental testing of compounds identified in this analysis revealed that (2-(2,3-dihydro-1*H*-perimidin-2-

yl)-phenoxy)-acetic acid inhibits SrtA with an IC_{50} value of 47 ± 5.9 μM. Molecular dynamics simulations and a preliminary structure activity relationship study of this lead compound provide insight into its binding mechanism, and strategies to improve its activity.

2.3 Results and discussion

Virtual screening approaches are increasingly being used to identify lead molecules in drug discovery efforts (52). Typically, these campaigns make use of a single experimentally determined protein structure which is used by computational docking algorithms to predict the relative binding affinities and poses of a large number of small molecules (48). However, in solution, proteins are thought to adopt an ensemble of interchanging conformers (metastates), with the experimentally determined structure presumably representing an average of the low energy conformers sampled experimentally (53,54). In principle, small molecule binding to any one of the conformers in the ensemble might stabilize it and thereby shift the population equilibrium towards this conformation (55). Therefore, using only one, or a few, static experimentally determined protein structures in virtual screening may fail to discover high affinity binding small molecules that could be developed further into drugs. To account for protein flexibility in virtual screening, a number of techniques have been developed that in many instances allow for protein-side chain movement during the docking process (56). However, to account for full protein motion, docking to multiple structures obtained from x-ray crystallography, NMR, or molecular dynamics (MD) simulations is necessary (56,57). In this study we make use of the relaxed complex scheme (RCS), a virtual screening approach that combines the dynamic structural information afforded by MD simulations with docking algorithms. This method uses receptor snapshots generated from MD simulations to search for

small molecule binders via docking, therefore explicitly accounting for the flexibility of both the receptor and the ligands (58,59). A number of high affinity binders have been discovered through the RCS (60-62), including the FDA approved drug raltegravir which targets HIV-1 integrase (63-65).

NMR and MD studies have revealed that *S. aureus* SrtA is a highly dynamic protein suggesting that virtual screening approaches would benefit from the application of the RCS (44,49,50,66). In particular, two of its active site loops, the β6/β7 and β7/β8 loops, undergo major conformational changes when SrtA binds to its sorting signal substrate. The largest changes occur in the β6/β7 loop, which is unstructured and flexible in the apo-state, and transitions into a structured loop containing a 3_{10} helix when bound to the substrate analog (Fig. 2.1). Because structures generated from MD simulations are particularly well suited for improving the predictive power of docking results to flexible proteins (67), we therefore used the RCS method to conduct a virtual screen of compound libraries to identify inhibitors of the SrtA enzyme.

2.3.1 Virtual screening using the RCS

The procedures used for virtual screening are summarized in Fig. 2.2. A total of 33,161 compounds were downloaded from the ZINC database. 55,789 ligands were then obtained after accounting for their different charge states, stereoisomers and tautomerization states. We performed the screen in two stages, since it was computationally intractable to dock all 55,789 ligands to the NMR structure, as well as to numerous structures generated from MD calculations. In the first stage, all 55,789 ligands were docked to the substrate bound form of the enzyme determined by NMR (hereafter, called holo-SrtA). This structure was chosen for docking

because it presumably represents the enzymatically active form of the protein, and the atomic positions of the active site residues are well defined unlike structures of the enzyme determined in its apo-state. Moreover, previous studies have shown that the ligand bound structures of proteins which are generally less flexible, are more amenable to predictive docking experiments as compared to unliganded protein structures (68). Small molecules were docked to holo-SrtA after the *in silico* removal of the bound sorting signal. The top 500 compounds based on their docking scores were then chosen for the second round of screening using the RCS approach.

To prepare for the second round of screening that made use of the RCS, six 100 ns MD simulations of the NMR derived structure of the holo-SrtA were performed (49). A total of 4,800 snapshots from these calculations were clustered into 21 groups of related conformers using an RMSD-based clustering algorithm. The centroid member for each cluster was considered to be the best representative of each group and was used in subsequent analyses. As expected, an overlay of the 21 centroid structures reveals that most of the structural differences between the centroids occur in the β6/β7 loop (Fig. 2.3). In the second round of screening, each of the 21 centroid structures was docked to the top 500 ligands derived from the first screen. To evaluate the docking results, three approaches were used. First, the compounds were ranked by computing the average of the scores obtained from docking to the 21 centroid conformers (ensemble-average). Second, the compounds were ranked by their modified ensemble-average scores such that the number of conformers each centroid structure represents was taken into account (population-weighted ensemble-average). Third, the compounds were ranked by the best score they obtained from any of the docking calculations to the 21 centroid conformers (ensemble-best). For further analysis, the top 15 ligands in each ranking category were considered for

experimental testing, which after accounting for redundancy corresponded to 24 unique compounds.

2.3.2 Experimental screening of SrtA inhibition

A FRET-based assay was used to experimentally evaluate the inhibitory activity of lead compounds identified in the virtual screen. Of the 24 unique molecules, a total of 16 compounds were tested experimentally that were either purchased (14 total) or synthesized in house (2 total). The remaining eight compounds were not characterized experimentally as they were deemed too expensive to purchase, as well as too difficult to synthesize in house. However, of these eight compounds, a total of three closely related analogs were purchased and tested. Thus, of the initial 24 lead molecules identified in the virtual screen, a total of 19 lead molecules or closely related compounds were tested experimentally for their ability to inhibit SrtA. The FRET assay was used to evaluate 17 of the 19 compounds, while the remaining 2 molecules were fluorescent and needed to be tested with an HPLC assay.

Eight out of nineteen compounds tested had an IC_{50} between 47 and 368 μM (see Table 2.1). The most active compound identified from this screening is compound **1**, which had an IC_{50} value of 47.2 ± 5.9 μM. It is interesting to note that most of the experimentally determined inhibitory compounds that were deemed the best molecules using the RCS approach did not rank highly in the first stage of the virtual screen when they were docked only to the NMR structure. For example, compound **1**, which has the lowest IC_{50}, ranked 77^{th} when docked to the NMR structure, but it ranked 9^{th} when docked to the ensemble using the ensemble-best ranking method. This result illustrates the utility of the RCS method, since given limited resources, without application of the RCS method the initial low ranking of compound using conventional

approaches may have resulted in it not being tested experimentally. It is also interesting to note that each of the different ranking methods (ensemble-average, population-weighted ensemble-average or ensemble-best) produced a comparable number of experimentally verified hits, and that most of these verified potent molecules were detected by only one of the three ranking methods. This highlights the usefulness of using different methods to rank ligands docked to an ensemble of structures.

Our virtual screen using the RCS yielded a higher hit rate than previously reported virtual screen that made use of more traditional methods. Previously, a virtual screen for sortase inhibitors was reported that made use of the structure of apo-SrtA, the only structure that was available at that time. A total of ~150,000 compounds were virtually screened for binding (48). After experimental testing of the leads identified from the screen, 7.4% were inhibitory; a total of 8 out of 108 experimentally tested compounds in this study had IC_{50} values ranging between 75 to 400 μM (48). In contrast, 42.2% of the lead molecules we tested that were identified in our virtual screen were active; a total of 8 out of 19 molecules experimentally tested had IC_{50} values ranging between 47 to 368 μM. Because the virtual screens were performed by different research groups using different docking algorithms and virtual compound libraries, it is not possible to rigorously explain why we obtained a higher hit rate. However, there seems to be two likely reasons for this difference. First, we used the structure of holo-SrtA as the receptor in the initial docking calculations, which may yield better results than docking to apo-SrtA as its active site is more rigid and well-defined. Second, our analysis made use of the RCS, which accounts for protein motion by docking ligands to an ensemble of structures obtained from MD simulations.

2.3.3 Compound 1: Structure and dynamics of its predicted binding mode

Compound **1** was chosen for additional characterization as it has the lowest IC_{50} value and a number of derivatives of this molecule could readily be purchased. To further investigate the binding pose of compound **1**, it was redocked into the NMR structure using the "Induced Fit" workflow in Maestro, which combines both docking and protein rearrangement stages (see Methods for more details) (69,70). The structure of compound **1** is based on a 2-phenyl-2,3-dihydro-1*H*-perimidine scaffold. It contains a dihydroperimidine (DHP) group and a phenyl ring with an oxyacetic acid group attached at the *ortho* position. In the docking pose the molecule is positioned in the active site with the DHP group placed underneath the β6/β7 loop, and the phenyl ring projected towards the active site H120, C184, and R197 (Fig. 2.4*A*). Specificity for this orientation is achieved by interactions that originate from the carboxyl group of the small molecule, which simultaneously forms hydrogen bonds to the catalytically important residues R197 and H120 within the active site. A predicted hydrogen bond between the backbone of P163 and the amine of the DHP group in the small molecule also presumably stabilizes ligand binding. In the binding pose the naphthalene ring of the DHP group is wedged into a hydrophobic pocket formed by V166, I182, A118 and V161. This positioning orients the phenyl ring towards several potential hydrogen bonding groups within the enzyme's active site (e.g. the side chains of T183, C184, and the backbone of G119) suggesting that molecules in which this ring are appropriately modified could exhibit improved binding selectivity and affinity.

To gain insight into the dynamics of the bound state, a single, 50 ns MD simulation of SrtA:Compound **1** complex was performed. Over the course of the simulation, the structure of the protein resembled most closely several of the centroid structures, with the RMSD of the active site residues calculated to be as low as 1 Å. By the end of the simulation, the structure of

the complex was structurally most similar to several of the centroid structures (active site RMSD ~ 1.5 Å) and less similar to the NMR structure (active site RMSD ~ 2.5 Å). RMSD calculations of the ligand relative to the protein show that the molecule experiences motions that result in atomic displacements on the order of 2-3 Å relative to the initial pose. Interestingly, a major excursion from the binding mode can occur transiently which causes a >5 Å displacement from the initial binding pose, as well as a return to conformation that is very similar to the initial binding pose (less than 1.5 Å from the induced fit docking results) (Fig. 2.4*B*). This larger excursion is caused by movement of the naphthalene ring within the hydrophobic pocket formed underneath the β6/β7 loop. Presumably, the addition of nonpolar substituents to this ring to fill this pocket could further improve binding affinity. The side chain of R197 maintained hydrogen bond contacts with compound **1** for 53% of the simulation, primarily with atoms in the carboxylic acid group. Other contacts were more transient, with the most dominant interactions to compound **1** being between the NH in the DHP group and the backbone of A104 (11% of the simulation), the backbone of G167 (8% of the simulation), and the backbone of A92 (4% of the simulation). Early in the simulation the side chain of H120 flipped such that the hydrogen bonds between it and compound **1** were broken, and in the course of the simulation they did not reform. Overall, these results indicate that a reasonable strategy in lead development may be to create additional contacts to stabilize compound **1** in the binding site to increase the propensity of these hydrogen bonds.

2.3.4 Preliminary structure activity relationship study of compound 1

To develop compound **1** further, we performed a similarity search on the ChemBridge small molecule database. A total of 22 compounds with the 2-phenyl DHP scaffold were

identified. Based on the docking and MD calculations, 10 of these compounds were purchased and their inhibitory activity determined experimentally. These molecules contain polar substituents in the phenyl ring to facilitate hydrogen bonding to the active site, and include a smaller compound that only contains the 2-phenyl DHP scaffold (summarized in Table 2.2). Compounds containing naphthalene ring substituents may also exhibit improved binding, but were not tested in this study because they are not available for purchase from ChemBridge. The scaffold compound **1-1** did not inhibit SrtA, which is probably the result of missing hydrogen bonds to the active site R197 and H120, which underscores the importance of having polar groups on the phenyl ring. Compounds containing a nitro group (**1-5**) or chloro group (**1-6** and **1-7**) at the *para* position are the most active with IC_{50} values close to, or less than, 100 μM. The retention of activity after modification of the phenyl ring is presumably because these polar groups form favorable interactions with the side chains of R197 or H120 within the active site. Interestingly, compounds containing a substituent at the *meta* position of the phenyl ring (**1-2**, **1-4**, **1-8**, **1-9** and **1-10**) are less active or even inactive. This may be due to steric clashes at this site in the small molecules with residues projecting from the β2/α1 loop, or from the β7 and β8 strands. Lower activity was also observed when the nitro or chloro group at the *para* position was replaced by a smaller fluoro group (compare **1-3** with **1-5** and **1-6**). Unfortunately, none of the compounds inhibited SrtA better than our lead, probably because the substituents in the phenyl ring are not long enough to interact with both the active site R197 and H120. Future work will focus on synthesizing compounds with phenyl rings containing longer polar groups, and will explore different substituents on the naphthalene ring to increase contacts to the hydrophobic pocket.

2.4 Conclusion

A virtual screen identified molecules containing the 2-phenyl-2,3-dihydro-1*H*-perimidine scaffold as possible inhibitors of the *S. aureus* SrtA enzyme. A structure activity relationship analysis indicates that the best molecule in this class, (2-(2,3-dihydro-1*H*-perimidin-2-yl)-phenoxy)-acetic acid, inhibits the activity of SrtA with an IC_{50} value of 47.2 ± 5.9 μM. MD simulations of this molecule bound to SrtA provide insight into its binding mechanism and serve as the foundation for future structure guided studies to uncover analogs that might have increased potency. Our virtual screen made use of the RCS and had a significantly higher success rate in identifying inhibitor compounds of SrtA as compared to conventional methods, highlighting the improved predictive power of the ensemble docking approach (71,72). The 2-phenyl-2,3-dihydro-1*H*-perimidine-based lead compound discovered in this study is a promising candidate for further development into a therapeutically useful anti-infective agent that can be used to treat infections caused by MRSA, and other multi-drug resistant bacterial pathogens.

2.5 Materials and methods

2.5.1 Initial screen against the NMR structure

A 70% cluster of the clean lead-like library was obtained from the ZINC database and consists of 33,161 small molecules (73). The LigPrep program in Schrödinger Suite 2011 was used to prepare the ligands. Protonation states were assigned at pH 7.0 ± 2.0 with Epik (74,75). A total of 55,789 ligands were generated that had distinct structures, stereochemistries, and charge and tautomerisation states. All 55,789 ligands were docked into the lowest energy NMR structure of SrtA bound to a substrate analog (holo-SrtA, PDB ID: 2KID). The receptor was

processed using the default Protein Preparation Wizard, which employs a restrained, partial energy minimization. Grids were generated by Glide (76-78) with the grid box set around the substrate analog using default settings. The substrate analog was excluded in grid calculations. Docking was done with Glide using SP settings.

2.5.2 Molecular dynamics (MD) simulations and clustering

The MD simulations used in the current study have been described previously (49). Briefly, six 100-ns conventional MD simulations were performed on holo-SrtA using the AMBER99SB-ILDN force field with the simulation package NAMD (79,80). In three of these simulations the sorting signal remained in the active site, whereas in the other three the sorting signal adopted metastable states outside of the active site. The three simulations in which the sorting signal remained in the active site were chosen for clustering, as conformations from these simulations are likely to be more representative of the bound state than when the sorting signal was not bound near the catalytic triad. From the last 80 ns of each of these MD simulations, 1,600 frames at regularly spaced intervals were extracted, which yielded a total of 4,800 frames. These frames were aligned by the protein Cα atoms in the active site (residues 90-112, 120-130, 161-176 and 183-196) and clustered by root mean square deviation (RMSD) conformational clustering using the gromos algorithm as implemented in GROMACS 4.5 (81). Twenty-one clusters were obtained with an RMSD cutoff of 1.35 Å. The centroid member of each cluster was presumed to best represent members of the ensemble and was selected for subsequent docking studies.

2.5.3 Relaxed complex screen

Five hundred top scoring ligands from the initial screen using the NMR structure of holo-SrtA were docked into each of the 21 representative centroid structures. Procedures used for receptor preparation, grid generation and docking are the same as those described for the initial screen using the NMR structure. The compounds were ranked according to three ensemble-based criteria. First, the compounds were ranked by computing the average of the scores obtained from docking to the 21 centroid conformers (ensemble-average). Second, the compounds were ranked by the population-weighted ensemble-average scores, which were calculated according to Eq. 1:

$$\bar{E} = \frac{\sum_{i=1}^{21} w_i E_i}{\sum_{i=1}^{21} w_i} \quad (1)$$

, where $\bar{E}$ is the weighted ensemble-average score, w_i is the size of cluster i, and E_i is the docking score of the compound docked into the centroid of cluster i. Third, the compounds were ranked by the best score they obtained from any of the docking calculations to the 21 centroid conformers (ensemble-best).

2.5.4 Compounds and reagents

Select lead compounds identified from the docking calculations were purchased from ChemBridge Corp. (San Diego, CA, USA), Enamine Ltd. (Ukraine), Sigma-Aldrich Co. LLC (St. Louis, MO, USA) and Vitas-M Laboratory Ltd. (the Netherlands), or synthesized in house. The fluorogenic substrate used in the enzyme assays (Abz-LPETG-Dap(Dnp)-NH_2) was purchased from Pepnome Ltd. (China). All other reagents that were used were purchased from Sigma-Aldrich Co. LLC (St. Louis, MO, USA) or Thermo Fisher Scientific Inc. (Waltham, MA, USA), unless noted otherwise.

2.5.5 Enzymatic assays

Compounds were tested for SrtA enzymatic inhibition using an established Förster resonance energy transfer (FRET) assay. Work made use of $SrtA_{\Delta N59}$, which consists of residue 60-206. The purification and FRET assay protocols have been described previously (33,43). Briefly, 20 µL of SrtA (final assay concentration of 1 µM in FRET buffer: 20 mM HEPES, 5 mM $CaCl_2$, 0.05% v/v Tween-20, pH 7.5) was incubated with 1 µL of test compound solution (dissolved in Me_2SO, final assay concentration of 0.08-400 µM) for 1 hour at room temperature. Subsequently, 30 µL of substrate solution in FRET buffer (37.5 µM final assay concentration) was added to the mixture and the fluorescence was monitored using excitation and emission wavelengths of 335 and 420 nm, respectively. IC_{50} values were calculated by fitting three independent sets of data to Eq. 2 using SigmaPlot 6.0:

$$\frac{v_i}{v_0} = \frac{1}{1+\left(\frac{[I]}{IC_{50}}\right)^h} \quad (2)$$

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

The activities of fluorescent compounds that could not be reliably assayed by FRET were tested using a high performance liquid chromatography (HPLC) assay as previously described (38). Briefly, 1 µM SrtA was pre-incubated with inhibitors for 30 min at 37°C to account for any time-dependent inactivation. Reactions were performed in a total volume of 100 µL with all reagents dissolved in FRET buffer. The assay was started by adding to the enzyme a mixture containing 1 mM Abz-LPETG-Dap(Dnp)-NH_2 and 1 mM NH_2-Gly_3-OH (Sigma). After one hour, the reaction was quenched by adding 50 µL of 1 M HCl. A 100 µL of the quenched reaction mixture was then injected onto a reverse phase XSELECT™ HSS C18 5 µM 3.0 X 50

mm HPLC column (Waters Corp., Milford, MA) and its components separated using a linear gradient from 3-45% acetonitrile/0.1% trifluoroacetic acid applied over a period of 25 min. For each inhibitor the fractional activity remaining relative to uninhibited controls was calculated by measuring the difference in percent product formation (Abz-LPETGGG-OH) measured at 215 nm. IC_{50} values were calculated as described previously and are the average of three measurements.

2.5.6 Induced fit docking and molecular dynamics simulations

The compound determined experimentally to have the lowest IC_{50} value (hereafter called "Compound **1**") was computationally redocked to the NMR structure using the Schrödinger Induced Fit protocol (69,70). This protocol accounts for receptor flexibility using a three-step method that includes an initial docking calculation with Glide, refinement of residues within 5 Å of the small molecule's docked pose using Prime (82,83), and a redocking stage that uses Glide. In this protocol, Glide was used as previously described, and the default parameters were used for Prime.

Molecular dynamics simulations of compound **1** were performed using a combination of the AMBER99SB-ILDN force field for the protein, and GAFF for the small molecule (84,85). Partial charges for GAFF were determined from a RESP fit to quantum calculations at the HF/6-31G* level of theory. The complex was solvated in a triclinic box of TIP3P water molecules with sufficient sodium and calcium ions to create a neutral simulation box of approximately 150 mM NaCl. Following relaxation with the default protocol in Maestro, a 50 ns MD simulation was performed with Desmond (86). Hydrogen bond analysis was performed with the Hydrogen Bonds plugin in VMD using a donor-acceptor distance of 3.5 Å and an angle cutoff of 30° (87).

2.6 Figures

Figure 2.1. Structure of holo-SrtA. NMR structure of holo-SrtA (PDB ID: 2KID) shown in cartoon. Residues K67 to K206 are shown, while the sorting signal analog LPAT* is removed for clarity. The side chains of key active site residues are shown as sticks.

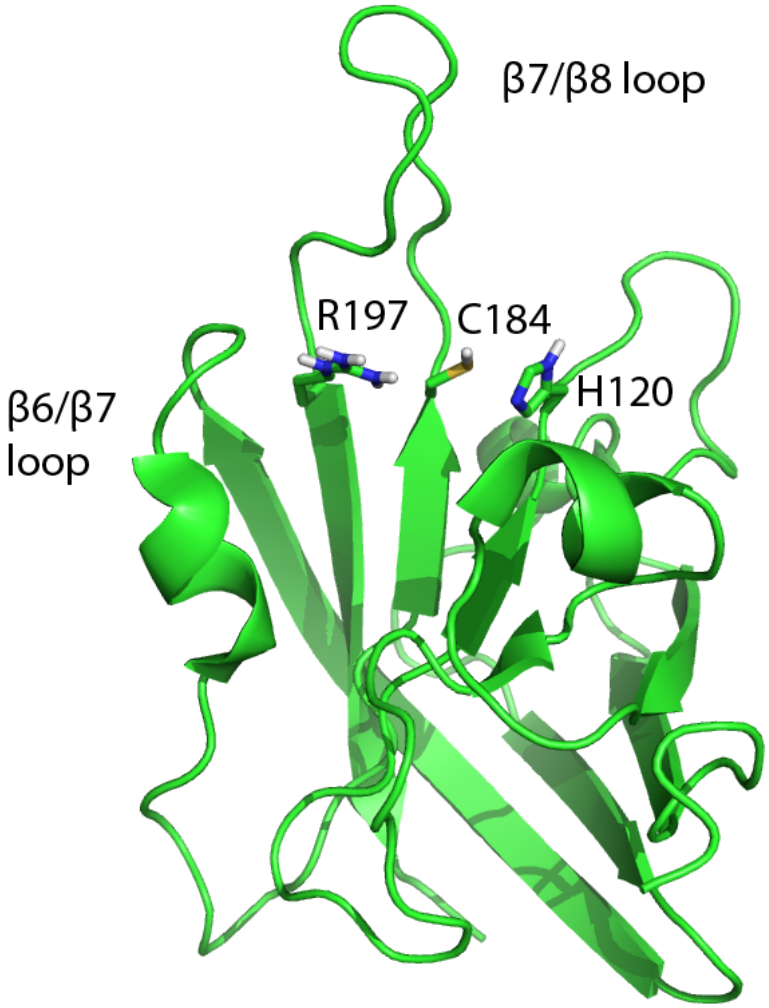


Figure 2.2. Overview of the two-staged virtual screening procedure that used the Relaxed Complex Scheme. In the first stage, small molecules from the ZINC compound library are docked using the program Glide to the NMR structure of substrate bound form of SrtA (PDB ID: 2KID). In the second stage, six 100 ns MD simulations of the NMR structure are performed and their snapshots are clustered by an RMSD-based algorithm, generating 21 clusters. The top 500 compounds obtained from the first screen are then docked using Glide to 21 centroid structures that represent each of the 21 clusters. Finally, the compounds are ranked by three different methods and the top 15 compounds in each ranking category are selected for experimental testing.

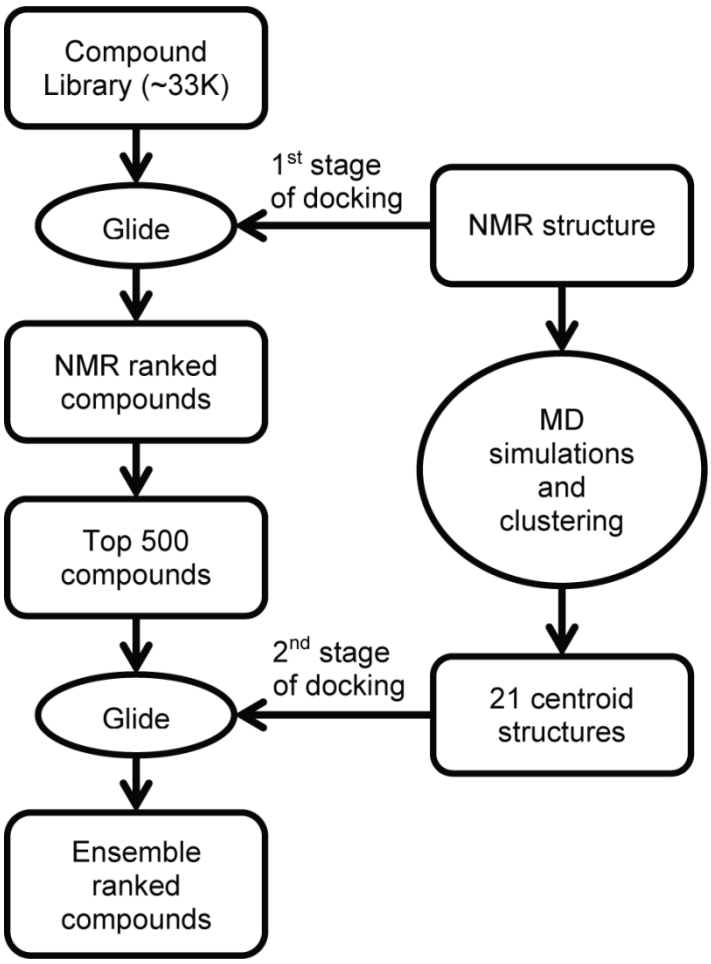


Figure 2.3. Representative MD simulated structures of holo-SrtA. Overlay of 21 centroid structures obtained from clustering the MD structures generated in the 6 x 100 ns simulation. Most of the changes occur within the β6/β7 loop and β7/β8 loop. These structures were aligned by the protein Cα atoms in the active site (residues 90-112, 120-130, 161-176 and 183-196).

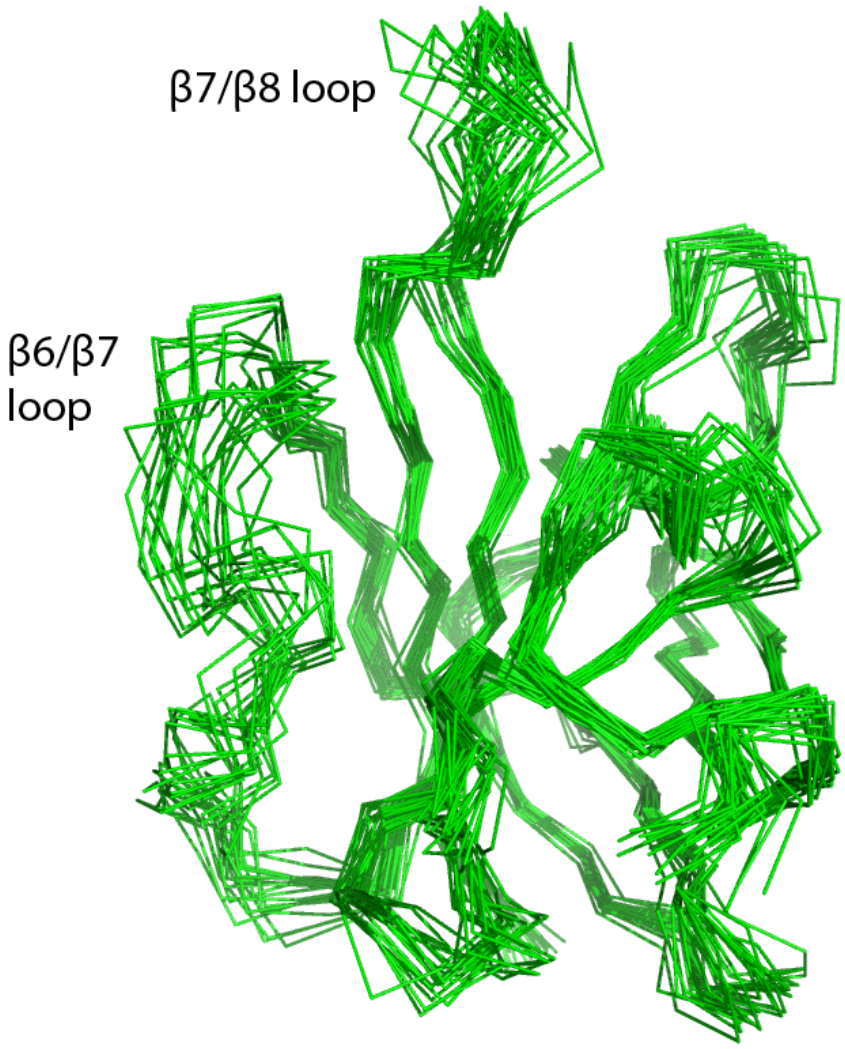
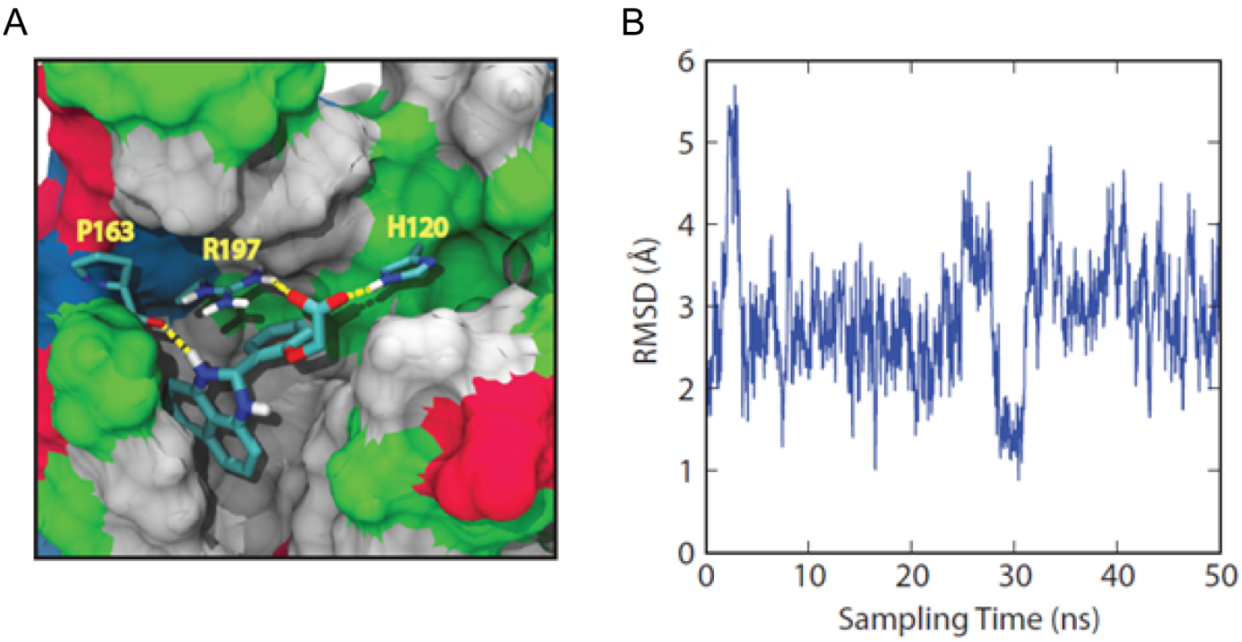


Figure 2.4. Docking pose of compound 1 and its coordinates RMSD during MD simulation. (A) Docking pose for compound **1** generated in "induced fit docking" calculations. Residues H120, P163, and R197 from SrtA are explicitly represented, along with their intermolecular hydrogen bonds to the small molecule (yellow lines). The remainder of SrtA is shown by a surface representation, with non-polar residues in grey, polar residues in green, acidic residues in red, and basic residues in blue. (B) Root-mean squared deviation (RMSD) of the position of the compound relative to its initial binding pose in SrtA at various time points during the 50 ns molecular dynamics simulation.



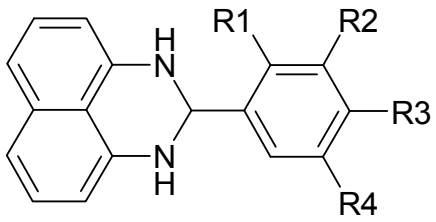
2.7 Tables

Table 2.1. Compounds identified from the virtual screen that inhibit SrtA

Compound	ZINC ID	NMR Rank	Ensm-Avg Rank	Weighted Ensm-Avg Rank	Ensm-Best Rank	IC_{50} (μM)
1	406572	77	27	24	9	47.2 ± 5.9
2	33733644[a]	145	8	4	13	98.9 ± 7.7
3	46093796	158	18	7	8	114 ± 13
4	41495051	468	4	10	19	132 ± 21
5	28294435[b]	124	7	8	47	189 ± 31
6	6538309	440	16	26	11	256 ± 21
7	6598689	148	19	12	77	276 ± 20
8	13610765	161	11	35	26	368 ± 29

Ensm-Avg refers to Ensemble-Average; Weighted Ensm-Avg refers to Population-weighted Ensemble-Average; Ensm-Best refers to Ensemble-Best; [a]ZINC33733644 identified in the virtual screen was not available for purchase and was deemed too technically difficult to synthesize in house. Therefore, its close analog ChemBridge 7253325 was tested instead; [b]ZINC28294435 identified in the virtual screen was not available for purchase and was deemed too technically difficult to synthesize in house. Therefore, its close analog ChemBridge 5303268 was tested instead.

Table 2.2. Preliminary structure activity relationship study of compound 1



Compound	R1	R2	R3	R4	IC_{50} (μM)
1 (lead)	OCH_2COOH	H	H	H	47.2 ± 5.9
1-1	H	H	H	H	>400
1-2	H	Cl	H	H	>400
1-3	H	H	F	H	276 ± 29
1-4	H	OH	OCH_3	H	>400
1-5	H	H	NO_2	H	80 ± 5
1-6	H	H	Cl	H	89 ± 12
1-7	Cl	H	Cl	H	111 ± 9
1-8	H	OCH_3	OH	H	231 ± 60
1-9	H	OCH_2CH_3	OH	H	>400
1-10	H	OCH_3	OH	Br	136 ± 20

2.8 References

1. Lowy, F. D. (1998) *N Engl J Med* **339**, 520-532
2. Ippolito, G., Leone, S., Lauria, F. N., Nicastri, E., and Wenzel, R. P. (2010) *Int J Infect Dis* **14 Suppl 4**, S7-11
3. Otto, M. (2012) *Cell Microbiol* **14**, 1513-1521
4. Centers for Disease Control and Prevention. Active Bacterial Core Surveillance Report, Emerging Infections Program Network, Methicillin-Resistant Staphylococcus aureus, 2011. http://www.cdc.gov/abcs/reports-findings/survreports/mrsa11.html.
5. Welsh, K. J., Abbott, A. N., Lewis, E. M., Gardiner, J. M., Kruzel, M. C., Lewis, C. T., Mohr, J. F., Wanger, A., and Armitige, L. Y. (2010) *J Clin Microbiol* **48**, 894-899
6. Marty, F. M., Yeh, W. W., Wennersten, C. B., Venkataraman, L., Albano, E., Alyea, E. P., Gold, H. S., Baden, L. R., and Pillai, S. K. (2006) *J Clin Microbiol* **44**, 595-597
7. Ikeda-Dantsuji, Y., Hanaki, H., Sakai, F., Tomono, K., Takesue, Y., Honda, J., Nonomiya, Y., Suwabe, A., Nagura, O., Yanagihara, K., Mikamo, H., Fukuchi, K., Kaku, M., Kohno, S., Yanagisawa, C., Nakae, T., Yoshida, K., and Niki, Y. (2011) *J Infect Chemother* **17**, 45-51
8. Maresso, A. W., and Schneewind, O. (2008) *Pharmacol Rev* **60**, 128-141
9. Schneewind, O., and Missiakas, D. M. (2012) *Philos Trans R Soc Lond B Biol Sci* **367**, 1123-1139
10. Spirig, T., Weiner, E. M., and Clubb, R. T. (2011) *Mol Microbiol* **82**, 1044-1059
11. Clancy, K. W., Melvin, J. A., and McCafferty, D. G. (2010) *Biopolymers* **94**, 385-396
12. Marraffini, L. A., Dedent, A. C., and Schneewind, O. (2006) *Microbiol Mol Biol Rev* **70**, 192-221

13. Paterson, G. K., and Mitchell, T. J. (2004) *Trends Microbiol* **12**, 89-95

14. Mazmanian, S. K., Liu, G., Jensen, E. R., Lenoy, E., and Schneewind, O. (2000) *Proc Natl Acad Sci U S A* **97**, 5510-5515

15. Jonsson, I. M., Mazmanian, S. K., Schneewind, O., Verdrengh, M., Bremell, T., and Tarkowski, A. (2002) *J Infect Dis* **185**, 1417-1424

16. Weiss, W. J., Lenoy, E., Murphy, T., Tardio, L., Burgio, P., Projan, S. J., Schneewind, O., and Alksne, L. (2004) *J Antimicrob Chemother* **53**, 480-486

17. Miyazaki, S., Matsumoto, Y., Sekimizu, K., and Kaito, C. (2012) *FEMS Microbiol Lett* **326**, 116-124

18. Mazmanian, S. K., Liu, G., Ton-That, H., and Schneewind, O. (1999) *Science* **285**, 760-763

19. Suree, N., Jung, M. E., and Clubb, R. T. (2007) *Mini Rev Med Chem* **7**, 991-1000

20. Kim, S. H., Shin, D. S., Oh, M. N., Chung, S. C., Lee, J. S., Chang, I. M., and Oh, K. B. (2003) *Biosci Biotechnol Biochem* **67**, 2477-2479

21. Kim, S. H., Shin, D. S., Oh, M. N., Chung, S. C., Lee, J. S., and Oh, K. B. (2004) *Biosci Biotechnol Biochem* **68**, 421-424

22. Kim, S. W., Chang, I. M., and Oh, K. B. (2002) *Biosci Biotechnol Biochem* **66**, 2751-2754

23. Oh, K. B., Mar, W., Kim, S., Kim, J. Y., Oh, M. N., Kim, J. G., Shin, D., Sim, C. J., and Shin, J. (2005) *Bioorg Med Chem Lett* **15**, 4927-4931

24. Jang, K. H., Chung, S. C., Shin, J., Lee, S. H., Kim, T. I., Lee, H. S., and Oh, K. B. (2007) *Bioorg Med Chem Lett* **17**, 5366-5369

25. Kang, S. S., Kim, J. G., Lee, T. H., and Oh, K. B. (2006) *Biol Pharm Bull* **29**, 1751-1755

26. Park, B. S., Kim, J. G., Kim, M. R., Lee, S. E., Takeoka, G. R., Oh, K. B., and Kim, J. H. (2005) *J Agric Food Chem* **53**, 9005-9009

27. Jeon, J. E., Na, Z., Jung, M., Lee, H. S., Sim, C. J., Nahm, K., Oh, K. B., and Shin, J. (2010) *J Nat Prod* **73**, 258-262

28. Lee, Y. J., Han, Y. R., Park, W., Nam, S. H., Oh, K. B., and Lee, H. S. (2010) *Bioorg Med Chem Lett* **20**, 6882-6885

29. Oh, I., Yang, W. Y., Chung, S. C., Kim, T. Y., Oh, K. B., and Shin, J. (2011) *Arch Pharm Res* **34**, 217-222

30. Won, T. H., Jeon, J. E., Kim, S. H., Lee, S. H., Rho, B. J., Oh, D. C., Oh, K. B., and Shin, J. (2012) *J Nat Prod* **75**, 2055-2061

31. Won, T. H., Jeon, J. E., Lee, S. H., Rho, B. J., Oh, K. B., and Shin, J. (2012) *Bioorg Med Chem* **20**, 4082-4087

32. Bae, J., Jeon, J. E., Lee, Y. J., Lee, H. S., Sim, C. J., Oh, K. B., and Shin, J. (2011) *J Nat Prod* **74**, 1805-1811

33. Suree, N., Yi, S. W., Thieu, W., Marohn, M., Damoiseaux, R., Chan, A., Jung, M. E., and Clubb, R. T. (2009) *Bioorg Med Chem* **17**, 7174-7185

34. Maresso, A. W., Wu, R., Kern, J. W., Zhang, R., Janik, D., Missiakas, D. M., Duban, M. E., Joachimiak, A., and Schneewind, O. (2007) *J Biol Chem* **282**, 23129-23139

35. Oh, K. B., Nam, K. W., Ahn, H., Shin, J., Kim, S., and Mar, W. (2010) *Biochem Biophys Res Commun* **396**, 440-444

36. Oh, K. B., Kim, S. H., Lee, J., Cho, W. J., Lee, T., and Kim, S. (2004) *J Med Chem* **47**, 2418-2421

37. Kruger, R. G., Barkallah, S., Frankel, B. A., and McCafferty, D. G. (2004) *Bioorg Med Chem* **12**, 3723-3729

38. Kudryavtsev, K. V., Bentley, M. L., and McCafferty, D. G. (2009) *Bioorg Med Chem* **17**, 2886-2893

39. Jung, M. E., Clemens, J. J., Suree, N., Liew, C. K., Pilpa, R., Campbell, D. O., and Clubb, R. T. (2005) *Bioorg Med Chem Lett* **15**, 5076-5079

40. Liew, C. K., Smith, B. T., Pilpa, R., Suree, N., Ilangovan, U., Connolly, K. M., Jung, M. E., and Clubb, R. T. (2004) *FEBS Lett* **571**, 221-226

41. Connolly, K. M., Smith, B. T., Pilpa, R., Ilangovan, U., Jung, M. E., and Clubb, R. T. (2003) *J Biol Chem* **278**, 34061-34065

42. Scott, C. J., McDowell, A., Martin, S. L., Lynas, J. F., Vandenbroeck, K., and Walker, B. (2002) *Biochem J* **366**, 953-958

43. Ilangovan, U., Ton-That, H., Iwahara, J., Schneewind, O., and Clubb, R. T. (2001) *Proc Natl Acad Sci U S A* **98**, 6056-6061

44. Suree, N., Liew, C. K., Villareal, V. A., Thieu, W., Fadeev, E. A., Clemens, J. J., Jung, M. E., and Clubb, R. T. (2009) *J Biol Chem* **284**, 24465-24477

45. Zong, Y., Bice, T. W., Ton-That, H., Schneewind, O., and Narayana, S. V. (2004) *J Biol Chem* **279**, 31383-31389

46. Uddin, R., Lodhi, M. U., and Ul-Haq, Z. (2012) *Chem Biol Drug Des* **80**, 300-314

47. Mehta, H., Khokra, S. L., Arora, K., and Kaushik, P. (2012) *Der Pharma Chemica* **4**, 1776-1784

48. Chenna, B. C., Shinkre, B. A., King, J. R., Lucius, A. L., Narayana, S. V., and Velu, S. E. (2008) *Bioorg Med Chem Lett* **18**, 380-385

49. Kappel, K., Wereszczynski, J., Clubb, R. T., and McCammon, J. A. (2012) *Protein Sci* **21**, 1858-1871

50. Moritsugu, K., Terada, T., and Kidera, A. (2012) *J Am Chem Soc* **134**, 7094-7101

51. Naik, M. T., Suree, N., Ilangovan, U., Liew, C. K., Thieu, W., Campbell, D. O., Clemens, J. J., Jung, M. E., and Clubb, R. T. (2006) *J Biol Chem* **281**, 1817-1826

52. Tanrikulu, Y., Kruger, B., and Proschak, E. (2013) *Drug Discov Today* **18**, 358-364

53. Onuchic, J. N., Luthey-Schulten, Z., and Wolynes, P. G. (1997) *Annu Rev Phys Chem* **48**, 545-600

54. Hardin, C., Eastwood, M. P., Prentiss, M., Luthey-Schulten, Z., and Wolynes, P. G. (2002) *J Comput Chem* **23**, 138-146

55. Ma, B., Shatsky, M., Wolfson, H. J., and Nussinov, R. (2002) *Protein Sci* **11**, 184-197

56. Durrant, J. D., and McCammon, J. A. (2010) *Curr Opin Pharmacol* **10**, 770-774

57. Amaro, R. E., and Li, W. W. (2010) *Curr Top Med Chem* **10**, 3-13

58. Lin, J. H., Perryman, A. L., Schames, J. R., and McCammon, J. A. (2003) *Biopolymers* **68**, 47-62

59. Amaro, R. E., Baron, R., and McCammon, J. A. (2008) *J Comput Aided Mol Des* **22**, 693-705

60. Amaro, R. E., Schnaufer, A., Interthal, H., Hol, W., Stuart, K. D., and McCammon, J. A. (2008) *Proc Natl Acad Sci U S A* **105**, 17278-17283

61. Durrant, J. D., Hall, L., Swift, R. V., Landon, M., Schnaufer, A., and Amaro, R. E. (2010) *PLoS Negl Trop Dis* **4**, e803

62. Durrant, J. D., Urbaniak, M. D., Ferguson, M. A., and McCammon, J. A. (2010) *J Med Chem* **53**, 5025-5032

63. Hazuda, D. J., Anthony, N. J., Gomez, R. P., Jolly, S. M., Wai, J. S., Zhuang, L., Fisher, T. E., Embrey, M., Guare, J. P., Jr., Egbertson, M. S., Vacca, J. P., Huff, J. R., Felock, P. J., Witmer, M. V., Stillmock, K. A., Danovich, R., Grobler, J., Miller, M. D., Espeseth, A. S., Jin, L., Chen, I. W., Lin, J. H., Kassahun, K., Ellis, J. D., Wong, B. K., Xu, W., Pearson, P. G., Schleif, W. A., Cortese, R., Emini, E., Summa, V., Holloway, M. K., and Young, S. D. (2004) *Proc Natl Acad Sci U S A* **101**, 11233-11238

64. Schames, J. R., Henchman, R. H., Siegel, J. S., Sotriffer, C. A., Ni, H., and McCammon, J. A. (2004) *J Med Chem* **47**, 1879-1881

65. Summa, V., Petrocchi, A., Bonelli, F., Crescenzi, B., Donghi, M., Ferrara, M., Fiore, F., Gardelli, C., Gonzalez Paz, O., Hazuda, D. J., Jones, P., Kinzel, O., Laufer, R., Monteagudo, E., Muraglia, E., Nizi, E., Orvieto, F., Pace, P., Pescatore, G., Scarpelli, R., Stillmock, K., Witmer, M. V., and Rowley, M. (2008) *J Med Chem* **51**, 5843-5855

66. Tian, B. X., and Eriksson, L. A. (2011) *J Phys Chem B* **115**, 13003-13011

67. Nichols, S. E., Baron, R., Ivetac, A., and McCammon, J. A. (2011) *J Chem Inf Model* **51**, 1439-1446

68. McGovern, S. L., and Shoichet, B. K. (2003) *J Med Chem* **46**, 2895-2907

69. Sherman, W., Beard, H. S., and Farid, R. (2006) *Chem Biol Drug Des* **67**, 83-84

70. Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A., and Farid, R. (2006) *J Med Chem* **49**, 534-553

71. Huang, S. Y., and Zou, X. (2007) *Proteins* **66**, 399-421

72. Rao, S., Sanschagrin, P. C., Greenwood, J. R., Repasky, M. P., Sherman, W., and Farid, R. (2008) *J Comput Aided Mol Des* **22**, 621-627

73. Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., and Coleman, R. G. (2012) *J Chem Inf Model* **52**, 1757-68

74. Shelley, J. C., Cholleti, A., Frye, L. L., Greenwood, J. R., Timlin, M. R., and Uchimaya, M. (2007) *J Comput Aided Mol Des* **21**, 681-691

75. Greenwood, J. R., Calkins, D., Sullivan, A. P., and Shelley, J. C. (2010) *J Comput Aided Mol Des* **24**, 591-604

76. Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., and Banks, J. L. (2004) *J Med Chem* **47**, 1750-1759

77. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004) *J Med Chem* **47**, 1739-1749

78. Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., and Mainz, D. T. (2006) *J Med Chem* **49**, 6177-6196

79. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. (2010) *Proteins* **78**, 1950-1958

80. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) *J Comput Chem* **26**, 1781-1802

81. Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B., and Lindahl, E. (2013) *Bioinformatics* **29**, 845-854

82. Jacobson, M. P., Friesner, R. A., Xiang, Z., and Honig, B. (2002) *J Mol Biol* **320**, 597-608

83. Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J., Honig, B., Shaw, D. E., and Friesner, R. A. (2004) *Proteins* **55**, 351-367

84. Wang, J., Wang, W., Kollman, P. A., and Case, D. A. (2006) *J Mol Graph Model* **25**, 247-260

85. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) *J Comput Chem* **25**, 1157-1174

86. Bowers, K. J., Chow, E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., Klepeis, J. L., Kolossvary, I., Moraes, M. A., Sacerdoti, F. D., Salmon, J. K., Shan, Y., and Shaw, D. E. (2006) *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)*, Tampa, Florida, November 11-17

87. Humphrey, W., Dalke, A., and Schulten, K. (1996) *J Mol Graph* **14**, 33-38, 27-38

Chapter 3

NMR Structure-based Optimization of Pyridazinone Class Sortase A Inhibitors

3.1 Overview

Previously, from a high throughput screen and structure-activity relationship study that followed, we reported a series of pyridazinone compounds that inhibit *Staphylococcus aureus* sortase A (Sa-SrtA) with IC_{50} values at single digit micromolar concentrations. However, their mode of inhibition remained unknown, and there was no structural information about how they bind to the enzyme. To guide inhibitor development, Dr. Ethan Weiner solved the solution structure of Sa-SrtA bound to a pyridazinone inhibitor using NMR methods. In this chapter, I summarize Dr. Weiner's findings and present my *in silico* and experimental work to identify even more potent pyridazinone compounds. Remarkably, a new compound that inhibits Sa-SrtA with an IC_{50} of 21 nM was discovered, which represents a ~70-fold improvement from the original pyridazinone compound. Further experimental results show that this compound also inhibits *B. anthracis* SrtA, but is not toxic to human cells. The mode and rate of inhibition were also determined.

This chapter is a modified version of a manuscript that will be submitted for publication in a peer-reviewed journal.

3.2 Introduction

The increasing occurrence of antibiotic-resistant pathogenic Gram-positive and Gram-negative bacteria is a growing medical concern (1,2). The prevalence of Methicillin-resistant *Staphylococcus aureus* (MRSA), a Gram-positive pathogen that can cause a wide spectrum of diseases (skin infections, endocarditis, pneumonia, etc.), is particularly concerning (3,4). In a

survey conducted by the CDC, in 2011 there were about 80,000 cases of invasive MRSA infections in the United States, which resulted in more than 11,000 deaths (5). Additionally, *S. aureus* strains resistant to vancomycin and newer generation drugs such as linezolid and daptomycin, as well as multiple drug-resistant strains, have been identified (6-8). As *S. aureus* develops resistance to currently available drugs due to selective pressure, it is important to continue to develop novel therapeutic agents to treat infections caused by this pathogen.

Most Gram-positive bacteria encode a sortase A (SrtA) enzyme, which covalently anchors target proteins to the bacterial cell wall (9-13). SrtA-displayed proteins have a variety of functions important for establishing a bacterial infection, such as evading the host immune response, acquiring nutrients, and allowing the bacteria to adhere to the host tissue (14). Studies have shown that *srtA*$^-$ strains of several pathogenic bacteria, such as *S. aureus*, *Bacillus anthracis*, *Listeria monocytogenes*, and *Streptococcus pneumoniae*, fail to attach many virulence factors to the cell surface, resulting in decreased infectivity (15-24). Interestingly, *srtA*$^-$ strains of *S. aureus* are still able to be cultured in laboratory media (10), suggesting inhibition of SrtA would not kill the bacteria directly, but would instead prevent it from displaying virulence factors and allow the host to clear the infection. Presumably, this type of anti-infective agent may reduce the selective pressure against such an inhibitor, as there would be no direct pressure as an antibiotic (25). Development of a SrtA inhibitor will hopefully result in a successful anti-virulent drug useful for treating infections resulting from various Gram-positive bacteria.

Several attempts to develop a *S. aureus* SrtA (Sa-SrtA) inhibitor have been made (25,26). These include screenings and structure activity relationship (SAR) studies of natural product (27-40) and compound libraries (41-44), and the construction of rationally designed peptidomimetics and small molecules (45-50). Moreover, virtual screens using the structures of the apo- and

substrate-bound forms of Sa-SrtA have been performed (51-53). Prior to this work, there is only one structure of the sortase enzyme bound by a small molecule inhibitor. As described in Maresso *et. al.*, covalent inhibitors of the aryl (β-amino)ethyl ketone class with IC_{50} values between 15 and 50 μM were discovered, and their binding mechanisms were studied by co-crystallizing the inhibitors with sortase B from *Bacillus anthracis* (Ba-SrtB) (41). However, the Ba-SrtB system may not be a good platform for the optimization of Sa-SrtA inhibitors, since sortase B recognizes a distinct sorting signal [NP(Q/K)(T/S)(N/G/S)(D/A)] and differs substantially in its substrate binding site architecture (54). In addition, inhibitor binding resulted in very minor structural rearrangements of Ba-SrtB, whereas Sa-SrtA has a more dynamic substrate-binding pocket (55-58). While many compounds have demonstrated an ability to inhibit the sortase reaction *in vitro*, a more potent, specific, and non-toxic inhibitor is necessary to be useful as a clinical therapeutic.

Recently, a symmetric disulfide dimer of 4-ethoxy-5-mercapto-2-phenyl-3(2H)-pyridazinone with an IC_{50} of 1.5 ± 0.4 μM was identified after a SAR study of a compound discovered using a high throughput screen of drug-like compounds (42). By fully understanding its mechanism of binding to Sa-SrtA, it may be possible to rationally design a more potent inhibitor by altering specific chemical groups within the molecule so as to improve its affinity and selectivity for Sa-SrtA. Unfortunately, *in silico* docking models of the pyridazinone lead molecule bound to Sa-SrtA cannot be calculated reliably because of the highly dynamic $\beta 6/\beta 7$ loop within the enzyme active site. It is therefore necessary to determine the structure of the inhibitor bound to the Sa-SrtA active site. In this work, a soluble sodium thiolate derivative of 4-ethoxy-5-mercapto-2-phenyl-3(2H)-pyridazinone was bound to Sa-SrtA, and the structure of the substrate binding pocket of the complex was determined by NMR. The structure reveals that the

enzyme adopts a hybrid of the apo and holo Sa-SrtA structures, with both the $\beta6/\beta7$ and $\beta7/\beta8$ loops playing crucial roles in inhibitor recognition. Guided by computational studies of the complex, a series of pyridazinone analogs were synthesized and their efficacy evaluated *in vitro*. This work led to the discovery of 4-(3-hydroxypropoxy)-5-mercapto-2-(3-fluorophenyl)-3(2H)-pyridazinone, which inhibits Sa-SrtA with an IC_{50} value of 21 ± 14 nM. It and related compounds are not cytotoxic against human cells and are found to inhibit SrtA at a rate similar to other known drugs such as Clavulanate and Tazobactam. These new compounds are therefore promising leads for further development into clinical therapeutic.

3.3 Results

3.3.1 Sa-SrtA binds a sodium thiolate version of the small molecule inhibitor

The small molecule Sa-SrtA inhibitor, 4-ethoxy-5-mercapto-2-phenyl-3(2H)-pyridazinone (hereafter referred as compound **2-10**), was previously discovered through a high-throughput screening and an SAR study (Fig. 3.1*A*) (42). Unfortunately, attempts to crystallize **2-10** bound to SrtA were not successful. We therefore employed NMR to elucidate its interactions with the enzyme. To increase its solubility in aqueous solvent, a sodium thiolate version of the small molecule inhibitor (hereafter referred as **2-salt**) was synthesized. It is soluble up to 75 mM and has an IC_{50} of 35 ± 7.4 μM. A stable Sa-SrtA:inhibitor complex was formed by incubating **2-salt** with Sa-SrtA overnight, and could be verified through a comparison of the ^{1}H-^{15}N HSQC spectra of the protein before and after incubation with **2-salt** (Fig. 3.3*A*). The inhibitor has a thiolate group that may form a disulfide bond with Cys^{184} in the active site of the protein. To investigate this possibility, **2-salt** was incubated with a C184A mutant of ^{15}N-labeled Sa-SrtA overnight. Compatible with disulfide formation, a C184A mutant failed to form a complex with

the inhibitor, as no chemical shift perturbation occurred in the ^{1}H-^{15}N HSQC spectrum. To further verify that **2-salt** forms a disulfide bond with Cys^{184}, we digested the Sa-SrtA:**2-salt** complex with trypsin, and used LC-MS to monitor the digestion products QLTLITCDDYNEK (*m/z* for $[M + 2H]^{2+}$ is 819.5) and QLTLITCDDYNEK:**2-salt** complex (*m/z* for $[M + 2H]^{2+}$ is 942.5) before and after adding DTT. Without DTT, a peak eluted at 6.86 min that corresponds to QLTLITCDDYNEK:**2-salt** complex was present (Fig. 3.2*A*,*B*). However, after adding DTT, the peak at 6.86 min was replaced by a peak eluted at 5.09 min that corresponds to the unmodified peptide QLTLITCDDYNEK (Fig. 3.2*A*,*C*). This confirms that **2-salt** forms a stable covalent complex with Sa-SrtA by forming a disulfide bond with the active site Cys^{184}.

3.3.2 NMR model of Sa-SrtA:inhibitor complex

The structure of the Sa-SrtA:inhibitor complex was determined using multidimensional heteronuclear NMR. The NMR spectra of the free and inhibitor bound forms of Sa-SrtA differ substantially (Fig. 3.3*A*), so the ^{1}H, ^{13}C and ^{15}N resonances of the complex were assigned *de novo*. During the course of structure calculations, backbone and side chain atoms of active site residues were allowed to move while all other Sa-SrtA atoms were held fixed in space using the NMR structure of apo Sa-SrtA as a template (see section 3.6.3 for details). A total of 229 experimental restraints were used to model **2-salt** into Sa-SrtA, including: 156 intramolecular distances, 20 intermolecular NOE distances, 43 ϕ and ψ dihedral angles, and 10 $^{3}J_{HN\text{-}H\alpha}$ couplings. An ensemble containing 20 conformers representing the structure of the complex is shown in Fig. 3.4*A*. The conformers exhibit good covalent geometry and have no NOE, dihedral angle, or scalar coupling violations greater than 0.5 Å, 5°, or 2 Hz, respectively. The ensemble of docked structures has a backbone and heavy atom coordinate root mean square deviations

(RMSD) to the mean structure of structured non-fixed atoms of 0.20 ± 0.07 and 0.69 ± 0.06 Å, respectively. Complete structure and restraint statistics are presented in Table 3.1.

The **2-salt** molecule binds to the LP*X*TG sorting signal binding site on Sa-SrtA (Fig. 3.4*B*). Residues in the β4 and β7 sheets form the floor of the binding groove, while the walls are formed by the β2/H1, β3/β4, β6/β7 and β7/β8 loops. The inhibitor is anchored to the catalytic Cys^{184} through a disulfide bond to the 5-mercapto group, and extended towards the β6/β7 loop and enzyme surface. Val^{166}, Val^{168}, and Leu^{169} within the β6/β7 loop, Val^{193} within the β7/β8 loop, and Ile^{182} on β6 make hydrophobic contacts with the 2-phenyl group of the inhibitor (Fig. 3.4*C*), as evidenced by intermolecular NOEs from Val^{166} Hα, Val^{166} Hγ, Val^{168} Hγ, Leu^{169} Hδ, and Val^{193} Hγ protons to the inhibitor Hε and Hδ protons (Fig. 3.3*B*). An NOE was also identified from Ile^{182} Hδ to the inhibitor Hδ, but not the inhibitor Hε, consistent with the structure as the 2-phenyl group is directed away from the β8 sheet towards the solvent. This configuration of the inhibitor allows the 2-phenyl group to be perpendicular to strands β7 and β8, and be partially buried within this hydrophobic pocket and partially exposed to the solvent sandwiched between the side chains of Val^{193} on the β7/β8 loop and Leu^{169} on the β6/β7 loop. The central pyridazinone ring of the inhibitor scaffold is structurally well positioned within the substrate-binding pocket of Sa-SrtA. The inhibitor is positioned in this region by the disulfide linkage to Cys^{184} and numerous NOE restraints from Hγ of the inhibitor to Thr^{183} Hα (strand β7) Cys^{184} Hα (strand β7), Trp^{194} Hβ (β7/β8 loop), Ile^{182} Hγ1 (strand β7), Ile^{182} Hδ (strand β7), and Val^{193} Hγ (β7/β8 loop) (Fig. 3.3*B*). This orients the central ring such that the carbonyl group is directed towards the solvent, while the Hγ of the inhibitor faces strand β7. The indole ring of Trp^{194} on the β7/β8 loop closes the inhibitor recognition groove, shielding the

disulfide bond between the inhibitor and Cys^{184} from the solvent. This is evidenced by NOEs between Hε1, Hζ2, and Hζ3 of the indole ring of Trp^{194} and Hδ methyl protons on Leu^{97} within helix H1. This demonstrates the β7/β8 loop of the **2-salt** bound enzyme is positioned in a manner more similar to the apo Sa-SrtA enzyme with the loop pressed against H1, and differs from the substrate bound form of the enzyme, as these elements are separated by ~13 Å (see section 3.4.1) (55,59). The 4-ethoxy group of the inhibitor sits approximately in a narrow groove formed by the β7/β8 loop and the β2/H1 loop, although the precise location of this group is difficult to place, as only one intermolecular NOE could be assigned in this region between Trp^{194} Hδ1 and the inhibitor Hα methyl.

3.3.3 Rational Design of Pyridazinone Compounds and Evaluation of Pyridazinone Derivatives with Docking

The surface plot of the Sa-SrtA:**2-salt** complex offers important clues for improving the potency of the pyridazinone inhibitor (Fig. 3.4*D*). As described earlier, the 2-phenyl group is partially buried within a hydrophobic pocket enclosed by Val^{166}, Val^{168}, Leu^{169}, Ile^{182} and Val^{193}. As shown in Fig. 3.4*D*, this pocket is not completely occupied by the phenyl group, and can potentially accommodate small groups branching off from the *ortho* or *meta* position of the benzene ring to increase hydrophobic contact. On the solvent-exposed side of the 2-phenyl group, a patch of electronegative surface can be found within 6 Å. This electronegative surface comprises several glutamate and aspartate residues on the β3/β4 loop and the C-terminal part of the β6/β7 loop, which together form a calcium ion binding site. Thus, polar or positively charged groups can be placed on the benzene ring to introduce hydrogen bonding or charge-charge interaction, and at the same time increase compound solubility and prevent the hydrophobic ring

from being exposed to the solvent. On the other side of the inhibitor, the 4-ethoxy group projects to the solvent along a channel bound by the β7/β8 loop and the β2/H1 loop. It is not immediately clear about how this group can be modified to increase the inhibitor binding affinity, other than to introduce a terminal polar group to reduce solvent exposure of the hydrocarbon chain and to allow potential hydrogen bonding.

To quantitatively predict which modifications will improve inhibitor binding the most, we computationally generated different derivatives of the pyridazinone compound and used molecular docking to evaluate their binding poses and energies. To validate our docking method, we initially docked **2-10** to the NMR structure of the protein in the Sa-SrtA:inhibitor complex, and were able to reproduce the experimentally determined binding pose (RMSD = 0.7 Å). The general docking workflow is shown in Fig. 3.5*A*, and details are described in section 3.6.4. In the first stage of docking, initially 623 analogs of **2-10** that vary substituents at seven sites (R2-R8) were docked to Sa-SrtA. This allowed us to identify a handful of the best substituents at each site, which we subsequently varied simultaneously to create 5,424 derivatives and docked again to the NMR structure. A few more substituents were eliminated after evaluation of the docking poses and scores, so that the total number of substituents was reduced to 22. At the second stage of docking, 432 compounds with all possible combinations of substituents were generated with CombiGlide and docked to the NMR structure. To account for protein flexibility during docking, we performed a 100 ns molecular dynamics simulation of the Sa-SrtA:**2-salt** structure, and used a RMSD-based clustering method to select 25 representative structures for docking. Thus, 432 compounds were docked to the NMR structure as well as 25 representative MD structures in the second stage of docking. To evaluate the docking results, compounds were ranked by the best docking score they obtained from any of the docking calculations to the NMR and MD simulated

structures. The top 43 compounds, or 10%, had docking scores ranging from -8.9 to -10.0. This indicates a significant improvement in binding when compare to the docking score of **2-10**, which was -4.08.

3.3.4 *In vitro* Testing of Pyridazinone Derivatives

The most favorable substituents among the top 43 compounds were considered for synthesis and experimental testing. They include six substituents at R2: $–O(CH_2)_2OH$, $–O(CH_2)_3OH$, $–OCH_2COOH$, $–O(CH_2)_2COOH$, $–OCH_2CONH_2$, and $–O(CH_2)_2CONH_2$; one substituent at R4: $–C(=NH)NH_2$ (hereafter referred as amidine); three substituents at R6: $–CH_3$, –F, and $–OCH_3$. A total of 12 molecules were synthesized and their activity determined experimentally using a FRET-based assay, and the results are shown in Table 3.2. Since **2-10** oxidizes easily to form a stable symmetric disulfide dimer **2-17** that has a significantly higher inhibitory activity (IC_{50} = 1.5 ± 0.4 μM), all of the pyridazinone derivatives were initially synthesized as disulfide dimers and their IC_{50}'s were compared to that of **2-17**. Initially, ten analogs (compounds **2-51** to **2-60**) were synthesized that vary only one substituent on the **2-10** molecule. As compared to **2-17**, some of these molecules have IC_{50} values that are 2 to 10-fold lower. Notable exceptions are with the carboxylic acid and amide substituents at R2 (**2-57** to **2-60**), which actually increased the IC_{50} by 2 to 15-fold. Modifying the R2 and R6 positions are most effective at improving the inhibitory activity of the pyridazinone compound. In contrast, adding an amidine group at the R4 position (**2-54**) improved IC_{50} by only 3-fold, even though docking results suggest that the positive charge should interact favorably with the acidic residues within the $\beta3/\beta4$ loop. A possible reason for this is discussed in section 3.4.2. To test whether we

can further improve the activity of pyridazinone derivatives by incorporating two substituents at the same time, we chose to incorporate the best substituents at the R2 and R6 sites. Interestingly, while incorporating $O(CH_2)_2OH$ (R2) and F (R6) (compound **2-61**) did not further improve the IC_{50}, combining $O(CH_2)_3OH$ (R2) and F (R6) (compound **2-62**) led to a significant improvement in inhibitory activity, with an IC_{50} value of 0.02 ± 0.01 μM. Their docking poses may explain this difference, and will be discussed in section 3.4.2.

The pyridazinone analogs are specific to sortases and non-toxic to human cells. To test specificity of the compounds, we tested them against the *Bacillus anthracis* class A sortase enzyme (Ba-SrtA) using the same FRET assay. Results show that these compounds are active against Ba-SrtA as well, with IC_{50} values in the low micromolar to high nanomolar range (Table 3.2). However, the most effective Sa-SrtA inhibitor (**2-62**) did not inhibit Ba-SrtA nearly as well (IC_{50} increased by 20-fold to 0.45 ± 0.29 μM), which might be due to subtle differences in the binding site architecture. Interestingly, the 3-F compound **2-51** also inhibited Ba-SrtA at a 10-fold higher IC_{50}, suggesting that 3-F is not well tolerated by Ba-SrtA. We tested the cytotoxicity of the most potent Sa-SrtA inhibitors against human cancer cell line HeLa cells. Results show minimal cytotoxic to human cells, with CC_{50} values of 100 μM or greater (Table 3.2). Compare to their IC_{50} values with Sa-SrtA, CC_{50} values are ~100 to 5000-fold higher. Remarkably, the CC_{50} values are correlated with the compounds' predicted octanol-water partition coefficient (cLogP) calculated from the ALOGPS 2.1 program (60,61). A negative LogP value indicates the compound is more soluble in water, while a positive LogP value indicates that it is more soluble in octanol, a mimic of the cell membrane. Table 3.2 shows that in general, compounds with cLogP values lower than 2.8 have CC_{50} values higher than 100 μM, and this will be discussed further in section 3.4.3.

3.3.5 Kinetics of enzyme inhibition

Because the compounds are covalent inhibitors, we measured the rate constant of inhibition of **2-62**, the most potent Sa-SrtA inhibitor (see section 3.6.5). Fig. 3.6 shows the observed rate constant for inhibition (k_{obs}) measured at different inhibitor concentrations, which yields a k_{inact} of 0.034 ± 0.007 s^{-1} and K_I of 12 ± 5 μM. The k_{inact} number describes the maximum k_{obs} when an infinite amount of inhibitor is present, and therefore is a measure of the rate in which the compound reacts covalently with the enzyme. On the other hand, K_I reflects how well the compound binds non-covalently to the enzyme. Since in all of our compounds the sulfur atom is attached to the same pyridazinone scaffold, we expect k_{inact} to be very similar among different pyridazinone derivatives. However, they are likely to have very different K_I, which contributed to the range of IC_{50} values we measured.

3.4 Discussion

3.4.1 Inhibitor bound Sa-SrtA adopts a hybrid structure of apo-Sa-SrtA and Sa-SrtA:LPAT* structures

The conformation of the inhibitor binding groove in the Sa-SrtA-inhibitor structure reflects a hybrid between the apo-Sa-SrtA and Sa-SrtA:LPAT* structures (Fig. 3.7*A*,*B*) (55,59). Based on previous biochemical and structural studies, the *β*6/*β*7 loop plays an important role in recognizing the LP*X*TG cell wall sorting signal substrate (48,55). NMR studies of the apo-SrtA have shown this loop to be structurally disordered from residues Thr^{156}-Asp^{176}. However, upon binding a substrate analog it rearranges into a "closed" conformation that contains a 3_{10}-helix. In the structure of the inhibitor complex the *β*6/*β*7 loop also closes around the inhibitor, although in

a manner distinct from the "closed" conformation identified in the Sa-SrtA:LPAT* structure (Fig. 3.7*B*). Once Sa-SrtA binds the LPAT* substrate analog, the $\beta6/\beta7$ loop forms a 3_{10}-helix from Val^{166}-Leu^{169} and forms a hydrophobic pocket around the substrate leucine side chain. A similar 3_{10}-helix also forms in the inhibitor complex that consists of Gly^{167}-Asp^{170}. It is identical to the helix in the substrate complex, but it is shifted by one residue in the primary sequence. Notably, the backbone of Val^{166} becomes extended in the inhibitor complex whereas in the Sa-SrtA:LPAT* complex it is in a helical conformation. This distinct conformation occurs because it repositions the Val^{166} side chain so as to enable contacts to the 2-phenyl group of the inhibitor, which is wedged in between the Val^{166} and Leu^{169} side chains (Fig. 3.7*D*). The position of the Val^{166} side chain is different in the Sa-SrtA:LPAT* structure, as it is rotated out away from the active site so as to partially cover the top of the substrate leucine side chain (Fig. 3.7*E*).

Although the inhibitor binds to the same site as the substrate, it does not cause similar dynamics and structural changes in the enzyme. The region of the substrate-binding pocket that recognizes the leucine residue of the cell wall sorting signal (CWSS) in the Sa-SrtA:LPAT* complex undergoes a disordered to ordered transition upon binding the analog (55). This region is formed by residues Val^{161}, Thr^{164}, and Val^{166} in the $\beta6/\beta7$ loop and Arg^{197} in strand $\beta8$. In its apo-form, these residues are disordered in both the NMR and crystal structures, while in the LPAT* complex they form a pocket that partially encapsulates the leucine side chain. In the structure of the inhibitor complex presented here, this portion of the protein remains mostly disordered after binding **2-salt** and thus resembles the apo-form. This is evidenced by the absence of NMR resonances for residues Thr^{156}-Val^{166} and Arg^{197}-Phe^{200} in the ^{1}H-^{15}N HSQC spectrum of the inhibitor complex (Fig. 3.7*C*). These signals are presumably broadened beyond detection because they undergo conformational rearrangements that occur on the micro- to milli-

second time scale. Motions probably involve only residues Thr^{156}-Val^{166} in the β6/β7 loop, which cause fluctuations in the magnetic environment of residues Arg^{197}-Phe^{200} in strand β8, which are presumably rigid. Similar mobility within this part of the β6/β7 loop has also been observed in the NMR structures of the apo-enzyme, as backbone amides for residues Ser^{157}-Asp^{160}, Lys^{162}, Thr^{164}, and Lys^{198}-Val^{201} are absent in the apo-enzyme spectrum. These motions are also likely to be present in the crystal structure, as B-factors within the crystal structure of apo-Sa-SrtA for the β6/β7 loop from residues Lys^{162}-Asp^{176} are elevated with some residues missing electron density entirely, further supporting the flexibility of the substrate leucine binding pocket. Combined, this indicates that the inhibitor does not completely mimic contacts originating from the leucine residue, causing the pocket to not form completely.

Unlike when the substrate analog binds Sa-SrtA, the β7/β8 loop is only slightly displaced upon binding of the inhibitor. Upon binding the signal analog, the β7/β8 loop of Sa-SrtA is displaced from the unbound state by ~13 Å. This rearrangement results in the formation of a new groove between the β7/β8 loop and helix H1, which has been hypothesized to form the binding site of the secondary substrate, the Gly_5 cross-bridge of lipid II. Binding of the inhibitor displaces the loop in a similar manner but at a much smaller amplitude of ~3.5 Å, such that the groove is not opened to the same extent. This is substantiated by NOEs in the spectra of the inhibitor complex between Hε1, Hζ2, and Hζ3 of the indole ring of Trp^{194} located in the β7/β8 loop and Hδ methyl protons on Leu^{97} within helix H1. Thus, in the structure of the inhibitor complex, the β7/β8 loop more closely resembles the structure of the loop in apo Sa-SrtA, retaining packing interactions to helix H1.

3.4.2 Relating docking poses to observed IC_{50} values

Through molecular docking and experimental testing, we identified a very potent compound **2-62** (4-(3-hydroxypropoxy)-5-mercapto-2-(3-fluorophenyl)-3(2H)-pyridazinone) that has an IC_{50} of 21 nM, which is ~70 times more potent than the parent compound **2-17**. To understand the molecular basis of improvement in binding, we examined its docking pose (Fig. 3.8*A*,*C*). It is predicted to bind the best to the NMR structure because it forms numerous specific contacts. On the R6 site, the fluoro group occupies the hydrophobic pocket created by residues in strand β6, the β6/β7 and β7/β8 loops to increase the number of van der Waals contact. On the R2 site, the terminal hydroxyl group on the $-O(CH_2)_3OH$ group becomes both a H-bond donor to the backbone carbonyl oxygen of Gly^{192} and a H-bond acceptor to the side chain hydroxyl group of Tyr^{187}, both located on the β7/β8 loop. In addition, the oxygen at the R2 site closest to the pyridazinone ring accepts a hydrogen bond from the εNH group in the indole ring of Trp^{194}. Overall, compared to the original compound **2-10**, this more potent pyridazinone derivative is predicted to form three extra H-bonds and have more van der Waals contact to the enzyme, resulting in increased binding affinity. Interestingly, compound **2-61,** which is identical to **2-62** but with the R2 group shortened by one carbon unit, is ~20 times less potent than **2-62**. Its docking pose (Fig. 3.8*B*,*D*) shows that instead of hydrogen bonding with Tyr^{187} and Gly^{192}, because of the shorter length of the R2 group, the terminal hydroxyl group is only capable of forming one H-bond with the backbone oxygen of Ala^{92} in the β2/H1 loop. The loss of one H-bond may be the reason that **2-61** did not inhibit Sa-SrtA as well as **2-62**.

Although including carboxylic acid and amide substituents at R2 (**2-57** to **2-60**) were predicted to improve binding of the pyridazinone compounds, their IC_{50} values are worse than the parent compound **2-17**. In their docking poses, both the carboxyl group and amide group

form hydrogen bonds with various residues within the β7/β8 and β2/H1 loops, similar to what has been observed for the hydroxyl group. However, since both carboxyl and amide groups are larger than a simple hydroxyl group, they may cause unfavorable steric clashes when the compound is bound to SrtA. It is important to note that the coordinates of the 4-ethoxy group in **2-salt** are not well defined in our NMR model due to limited number of intermolecular NOEs. This might in turn lead to inaccurate positioning of the side chains in the beginning of the β7/β8 loop and the later part of the β2/H1loop, preventing accurate prediction of potential steric clashes between the R2 substituents and these side chains.

In our NMR model of the Sa-SrtA:**salt-2** complex and our docking calculations, we expected a positively charged group at R4 to interact with the nearby acidic residues such as Glu^{105} in the β3/β4 loop. However, incorporating amidine at R4 (compound **2-54**) only lowered the IC_{50} value by ~3-fold. A possible reason for this is that although the amidine group-containing ligands scored very favorably when docked to the NMR structure, they did not score as well when docked to the MD simulated structures of Sa-SrtA, since in many of the MD structures the side chain of Glu^{105} points toward the C-terminal end of the β6/β7 loop / beginning of strand β7 such that it is not capable of forming a salt bridge with the amidine group on the ligand. Despite its modest improvement in IC_{50}, we are still interested in incorporating amidine to other pyridazinone derivatives in the future, since the amidine group should further improve the solubility of the compounds in water, and its positive charge could direct the compounds towards the negatively charged cell wall of *S. aureus*.

3.4.3 Increase in solubility likely reduces cytotoxicity in human cells

One of the goals in developing the pyridazinone compounds is to increase their solubility in aqueous buffer, since the previously published compounds are not very soluble in water, hindering structural studies as well as testing in animal models of *S. aureus* infections. We therefore created many polar substituents in attempt to increase solubility of the compounds. Other than at the R6 site, all substituents we screened are polar leading to 1-2 logs reduction in cLogP. Interestingly, we noticed that compounds with low cLogP values tend to be less toxic to human cells. Presumably this is because the more polar compounds are less likely to penetrate the cell membrane lipid bilayer, therefore reducing off-target inhibition that leads to their toxicity. Nevertheless, even our parent compound **2-17** which has a high cLogP is not very cytotoxic, perhaps because the predominantly reducing environment inside the cell prevents it from forming nonspecific disulfide bonds with the host proteins. Importantly, the catalytic domain of sortase is located outside of the cell membrane in the cell wall of the bacterial cell, so that it is accessible to both hydrophobic and polar compounds. The cell wall environment is also not as reducing as the cytoplasm so that our compounds can form disulfide bond with the enzyme to inhibit it.

3.4.4 Comparison of rate of inhibition between our compound and other covalent inhibitors

Our pyridazinone compounds inhibit Sa-SrtA at a rate similar to other known covalent inhibitors. To gain insight into the rate of inhibition, we measured k_{inact} and K_I of compound **2-62**. Our results indicate that **2-62** has k_{inact} of 0.034 ± 0.007 s^{-1} and K_I of 12 ± 5 μM, which are comparable to several mechanism-based β-lactamase inhibitors on the market such as

Clavulanate (k_{inact} = 0.034 ± 0.001 s^{-1}, K_I = 0.60 ± 0.08 μM) and Tazobactam (k_{inact} = 0.12 ± 0.01 s^{-1}, K_I = 1.6 ± 0.2 μM) (62). In particular, our k_{inact}, or rate of inhibition given large amount of the inhibitor, is as good as these FDA-approved drugs. Our K_I, which is related to the initial non-covalent binding step, is also within an order of magnitude of the K_I's of these two covalent inhibitors. Overall, our kinetic data indicate that compound **2-62** compares favorably to other known covalent drugs in terms of inhibiting its target.

3.5 Conclusion

To conclude, by modeling enzyme interactions of a previously discovered Sa-SrtA inhibitor using NMR, we were able to identify sites on the pyridazinone compound for modification, and to use molecular dynamics and docking to screen for the most favorable substituents. Experimental testing revealed that we have discovered a very potent compound **2-62** that inhibits Sa-SrtA with an IC_{50} of 21 nM, a ~70-fold improvement from the parent compound. Moreover, the new compounds we discovered are more soluble in water, non-toxic to human cells, and inhibit the enzyme at a rate similar to other drugs currently on the market. In the future we plan to test several of the best compounds in mouse model of systemic infection to investigate their efficacy as a therapeutic.

3.6 Materials and Methods

3.6.1 Sa-SrtA:inhibitor complex formation and NMR sample preparation

Wild-type or C184A SrtA protein from *S. aureus* (Sa-SrtA) containing amino acid residues Gln^{60}-Lys^{206} were produced as described previously (59). To form a Sa-SrtA:inhibitor complex, 0.5 mM uniformly ^{15}N- and ^{13}C-labeled Sa-SrtA was incubated with 5 mM of sodium

thiolate 4-ethoxy-5-mercapto-2-phenyl-3(2H)-pyridazinone (hereafter referred as **2-salt**) in 50 mM Tris, pH 6.4, 150 mM NaCl, 20 mM $CaCl_2$ for 72 h at room temperature. The fully formed complex was then dialyzed into fresh 50 mM Tris, pH 6.4, 150 mM NaCl, 20 mM $CaCl_2$ before being concentrated to 1.5 mM and adjusted to 7% D_2O. A D_2O sample was made by lyophilizing the sample before dissolving in 99.999% D_2O.

To verify that the inhibitor forms a disulfide bond with Cys^{184} in Sa-SrtA, the Sa-SrtA:**2-salt** complex was first digested with trypsin by incubating 20 μM $^{15}N^{13}C$-Sa-SrtA:**2-salt** complex with 5 μg/mL trypsin for 24 h at 37°C. The digestion reaction mixture was then split into two aliquots in which DTT was added to one of them to a final concentration of 5 mM. Both aliquots were subjected to LC-MS analysis to monitor the masses of the cleaved peptides.

3.6.2 LC-MS conditions

LC-MS experiments were carried out on a Waters Acquity UPLC connected to a Waters LCT-Premier XE Time of Flight Instrument controlled by MassLynx 4.1 software (Waters Corp., Milford, MA, USA). Digested protein samples in 20 mM HEPES, 5 mM $CaCl_2$ were separated using an Phenomenex Hydro-RP column (3.0 x 50 mm, 4 μm packing) and were eluted with a gradient of 2 – 50% solvent B over 10 min using a flow rate of 0.425 mL/min (solvent A: water, solvent B: acetonitrile, both with 0.3% formic acid (vol/vol)).

The mass spectrometer was equipped with a Multi-Mode Source operated in the electrospray mode. Mass spectra were recorded from a mass of 70 – 2,000 Daltons. Capillary voltage was set to 1700 V and the source/desolvation gas temperatures were 120°C/350°C, respectively. The ion abundance values for product ions at m/z 819.5, 942.5, 1638 and 1886 were monitored by generating extracted ion chromatograms using a 0.5 Da mass window and

integrating the peaks of interest. These ions had retention time values of 5.09, 6.86, 5.09 and 6.86 minutes, respectively.

3.6.3 NMR spectroscopy and structure determination

NMR spectra of Sa-SrtA:inhibitor complex were acquired at 298 K on Bruker Avance 500-, 600-, and 800-MHz spectrometers equipped with triple resonance cryogenic probes. NMR spectra were processed using NMRPipe (63) and analyzed using the PIPP (64) and CARA (version 1.8.4) (65) software packages. Chemical shift assignments (^{1}H, ^{13}C, ^{15}N) of Sa-SrtA were obtained by analyzing the following experiments: HNCA, HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO, ^{15}N-edited TOCSY, HNHA, HNHB, HBHA(CO)NH, HCCH-TOCSY, HCCH-COSY, (HB)CB(CGCDCE)HE, and (HB)CB(CGCD)HD (reviewed in (66,67)). Chemical shift assignments for the inhibitor were obtained by analyzing a two-dimensional (F2) ^{13}C-filtered NOESY spectrum. The majority of ϕ and ψ dihedral angle restraints were obtained by using the program TALOS+ (68). Additional backbone ϕ angle restraints were obtained by analyzing HNHA spectra (69). Intramolecular distance restraints were obtained from three-dimensional ^{15}N- and ^{13}C-edited NOESY spectra. Intermolecular distance restraints were obtained by analyzing the two-dimensional (F2) ^{13}C-filtered NOESY and ^{13}C-edited NOESY-HSQC spectra.

NOE assignments were initially obtained automatically using the programs ATNOS/CANDID controlled by UNIO (70,71) utilizing XPLOR-NIH (version 2.23) (72). NOE assignments were then verified for NOEs involving active site residues (Val^{161}-Asp^{175} [β6/β7 loop], Thr^{183}-Lys^{196} [β7/β8 loop], and side chains of Leu^{97}, Ser^{116}, His^{120}, Thr^{180}, Ile^{182}, and Val^{201}) by manually inspecting the NOESY data, and NOEs between non active site residues

were discarded. During inspection of the NOESY data, additional NOE restraints involving active site residues were identified and included in subsequent structure calculations. Final structures were calculated using XPLOR-NIH based on the lowest energy conformer of the previously solved apo Sa-SrtA NMR structure (PDB 1IJA) (59). During the course of the structure calculations, backbone and side chain atoms of active site residues Val^{161}-Asp^{175} ($\beta6/\beta7$ loop) and Thr^{183}-Lys^{196} ($\beta7/\beta8$ loop), as well as side chain atoms of Leu^{97}, Ser^{116}, His^{120}, Thr^{180}, Ile^{182}, Arg^{197}, Ile^{199}, and Val^{201}, were allowed to move while all other Sa-SrtA atoms were held fixed in space. Residues within the $\beta6/\beta7$ and $\beta7/\beta8$ loops were allowed to move because they form the surface surrounding the presumed inhibitor binding site, and previous crystal and NMR structural studies of apo and substrate (LPAT*) bound Sa-SrtA reveal these loops adopt distinct conformations as a result of substrate binding (55,59,73). Mobility of other active site residues was restricted to the side chains because while they make contacts with the sorting signal substrate in the Sa-SrtA:LPAT* structure, they do not demonstrate significant backbone rearrangement. Non-active site residues were kept rigid because they show little difference (RMSD = 0.8 Å) between the apo and substrate bound Sa-SrtA structures, and such structural information is of little value for structure-based drug design of a competitive inhibitor. The thiol group within the inhibitor was attached to Cys^{184} through a disulfide bond. A single hydrogen bond between the backbone carbonyl of Gly^{167} and the backbone amide of Asp^{170} was also used to stabilize the 3_{10}-helix within the $\beta6/\beta7$ loop, and was substantiated by characteristic NOE patterns from Gly^{167}-Asp^{170}. A total of 50 structures were calculated, of which 46 had no NOE, dihedral angle, or scalar coupling violations greater than 0.5 Å, 5°, or 2 Hz, respectively. Of these, 20 structures with the lowest overall energy were chosen to represent the structure of the

SrtA-inhibitor complex. The programs MOLMOL (74) and PyMOL (75) were used to generate figures.

3.6.4 *In silico* screening of pyridazinone derivatives

Ligand preparation, receptor preparation, grid generation and docking were all conducted with Schrödinger Suite 2011 (Schrödinger LLC, New York, NY, USA). Derivatives of 4-ethoxy-5-mercapto-2-phenyl-3(2H)-pyridazinone for docking experiment were generated with the CombiGlide application. In CombiGlide, 5-mercapto-2-phenyl-3(2H)-pyridazinone was defined as the core, and 225, 104, 104, 104, 32, 32, and 22 substituents were created at R2, R3, R4, R5, R6, R7, and R8 sites, respectively (see Fig. 3.5 for definition of the R sites). The substituents at the R2 site included various polar and non-polar groups of different sizes, as well as a series of substituents with the general formula $–O(CH_2)_nX$, where n = 1-4 and X is a polar group. The substituents at R3, R4 and R5 are predominantly polar or positively charged. R6, R7 and R8 contained mostly small hydrophobic substituents. The receptor was processed using Protein Preparation Wizard, which employs a restrained, partial energy minimization (76). The disulfide bond between SrtA and the inhibitor, and the side chain atoms of Cys^{187} were removed to prevent steric clashes during docking. Grids were generated by Glide with the grid box set around the inhibitor using default settings. The inhibitor was excluded in grid calculations. A docking restraint was set up such that the position of the 5-mercapto sulfur atom of the inhibitor derivatives was restricted within 1 Å of the inhibitor sulfur atom in the NMR structure. Docking was done with Glide using XP settings (77-79). Initially, a total of 623 ligands were docked to the NMR structure of the Sa-SrtA:inhibitor complex. After the first round of docking, the best substituents at each site were selected based on two criteria: 1) docking score was one standard

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

To account for protein flexibility during docking, Sa-SrtA motion was simulated using molecular dynamics (MD). Bond, angle, and torsion parameters for the inhibitor were derived from the Generalized Amber Force Field (GAFF), using the Antechamber program in Amber (80,81). Atomic partial charges were derived from RESP (82) fitting of Gaussian09 (83) calculated electrostatic potentials at the Hartree-Fock/6-31G* level. The procedures of MD simulations and clustering were the same as those described in (51). Briefly, a 100-ns conventional MD simulation was performed on the Sa-SrtA:inhibitor complex using the AMBER99SB-ILDN force field with the simulation package NAMD (84,85). Eight hundred frames at regularly spaced intervals were extracted from the last 80 ns of the MD simulation. These frames were aligned by the protein Cα atoms in the active site and clustered by root mean

square deviation (RMSD) conformational clustering using the GROMOS algorithm as implemented in GROMACS 4.5 (86). With an RMSD cutoff of 1.40 Å, 25 clusters were obtained, and the centroid member of each cluster was selected to represent each cluster. Subsequently, 432 compounds were docked to each of the 25 representative centroid structures, as well as the NMR structure. Procedures used for receptor preparation, grid generation and docking are the same as described above. To evaluate the docking results, compounds were ranked by the best docking score they obtained from any of the docking calculations to the NMR or 25 centroid conformers. The top 43 compounds were selected as candidates for experimental testing.

3.6.5 Enzymatic assays

Compounds were tested for SrtA enzymatic inhibition using an established Förster resonance energy transfer (FRET) assay. Both Sa-SrtA and Ba-SrtA were purified as described previously and used in the assay (42,59). Briefly, in the IC_{50} assay, 20 μL of Sa-SrtA (final assay concentration of 1 μM in FRET buffer: 20 mM HEPES, 5 mM $CaCl_2$, 0.05% v/v Tween-20, pH 7.5) was incubated with 1 μL of test compound solution (dissolved in 100% DMSO, final assay concentration of 0.08-400 μM) for 1 h at room temperature. Subsequently, 30 μL of substrate solution, which consists of the self-quenched fluorogenic peptide Abz-LPETG-Dap(Dnp)-NH_2 (32 μM final assay concentration) (Peptide 2.0 Inc., Chantilly, VA, USA) dissolved in FRET buffer, was added to the mixture. Fluorescence was read immediately using an Infinite® M1000 PRO (Tecan US Inc., Morrisville, NC, USA) plate reader with the excitation and emission wavelengths set at 335 and 420 nm, respectively. The IC_{50} assay used to test Ba-SrtA inhibition was similar with the following adjustments: final assay concentration of Ba-SrtA was 10 μM,

final assay concentration of substrate was 100 μM, and the FRET buffer was composed of 20 mM HEPES, 0.05% v/v Tween-20, pH 7.5. IC_{50} values were calculated by fitting three independent sets of data to equation 1 using SigmaPlot 6.0 (SPSS, Inc., Chicago, IL, USA):

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

, where v_i and v_0 are initial velocity of the reaction in the presence and absence of inhibitor at concentration [I], respectively. The term h is Hill's coefficient (87).

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

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

, where v_i and v_0 are initial velocity of the reaction in the presence and absence of inhibitor at concentration [I], respectively. $[E]_T$ is the total active enzyme concentration, and K_i^{app} is the apparent dissociation constant for the enzyme-inhibitor complex (87).

To test the rate of inhibition of Sa-SrtA by the compound **2-62**, the FRET assay we used for IC_{50} or K_i^{app} determination was modified such that the Abz-LPETG-Dap(Dnp)-NH_2 substrate and inhibitor of various concentrations (final concentration 2.5 – 14 μM) were added to the enzyme at the same time. Fluorescence was read every 8 s for a total of 30 min. The reaction progress curve was fit to equation 3 to determine k_{obs}, the rate constant for conversion from the initial velocity phase to full inhibition:

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

, where P is the baseline-corrected fluorescence value, v_i is the initial velocity, and t is time (87). These k_{obs} measurements were then plotted against inhibitor concentration and fit to equation 4 to obtain k_{inact} and K_I values.

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

, where k_{obs} is the observed rate constant of inhibition at inhibitor concentration [I], k_{inact} is the maximum rate of inhibition given an infinite concentration of inhibitor, and K_I is the concentration of inhibitor that yields a half-maximum rate of inhibition (87).

3.6.6 Cytotoxicity test

The cytotoxicity CC_{50} of each compound was determined using Promega CellTiter-Glo Luminescent Cell Viability Assay kit (Promega Corp., Madison, WI, USA) by measuring the total ATP levels to quantify the number of metabolically active cells upon drug treatment as described in (88). Briefly, the compounds were diluted in 384 plates (20 µl/well) in triplicate by a 10-point titration (195 nM to 100 µM) followed by the addition of 30 µl HeLa cells (2500 cells/well). The plates were incubated at 37°C. Three days later, 50 µl of CellTiter-Glo reagent was added to each well followed by a 2 min shaking and a 10 min incubation to lyse the cells. The relative luminescent intensity units (RLU) of each well was measured using an Infinite® M1000 PRO (Tecan US Inc., Morrisville, NC, USA) with its green filter and 1 s integration time. CC_{50} values were calculated by fitting the three independent sets of data to equation 1 using SigmaPlot 6.0 (SPSS, Inc., Chicago, IL, USA), by replacing IC_{50} with CC_{50}.

3.7 Figures

Figure 3.1. Structures of the pyridazinone compounds. *A*, structure of compound **2-10** (4-ethoxy-5-mercapto-2-phenyl-3(2H)-pyridazinone), a previously discovered Sa-SrtA inhibitor with IC_{50} of 13 μM or 1.5 μM when dimerized through a disulfide bond. *B*, structure of the sodium thiolate version of **2-10** (**2-salt**) used in LC-MS and NMR studies.

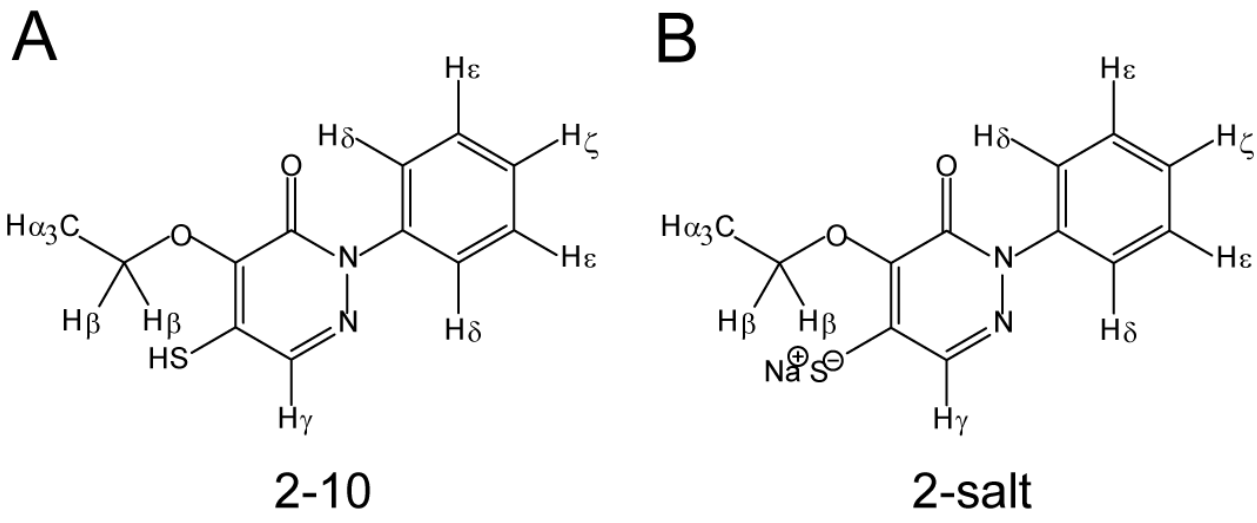


Figure 3.2. Pyridazinone compound 2-salt covalently modifies Sa-SrtA. *A*, overlay of the total ion traces of trypsin-digested 1:1 mixture of Sa-SrtA and **2-salt** before (blue) and after (red) addition of 5 mM DTT. Major differences occur at retention times 5.09 min and 6.86 min. *B*, chromatogram of trypsin-digested sample without addition of DTT. Only the traces of ions with *m/z* of 819.5 (red) and 942.5 (blue) are shown, which correspond to the $[M+2H]^{2+}$ ions of the unmodified Cys^{184}-containing peptide and **2-salt** modified version of the peptide, respectively. *C*, similar to (*B*), but the chromatogram shown is that of DTT added trypsin-digested sample.

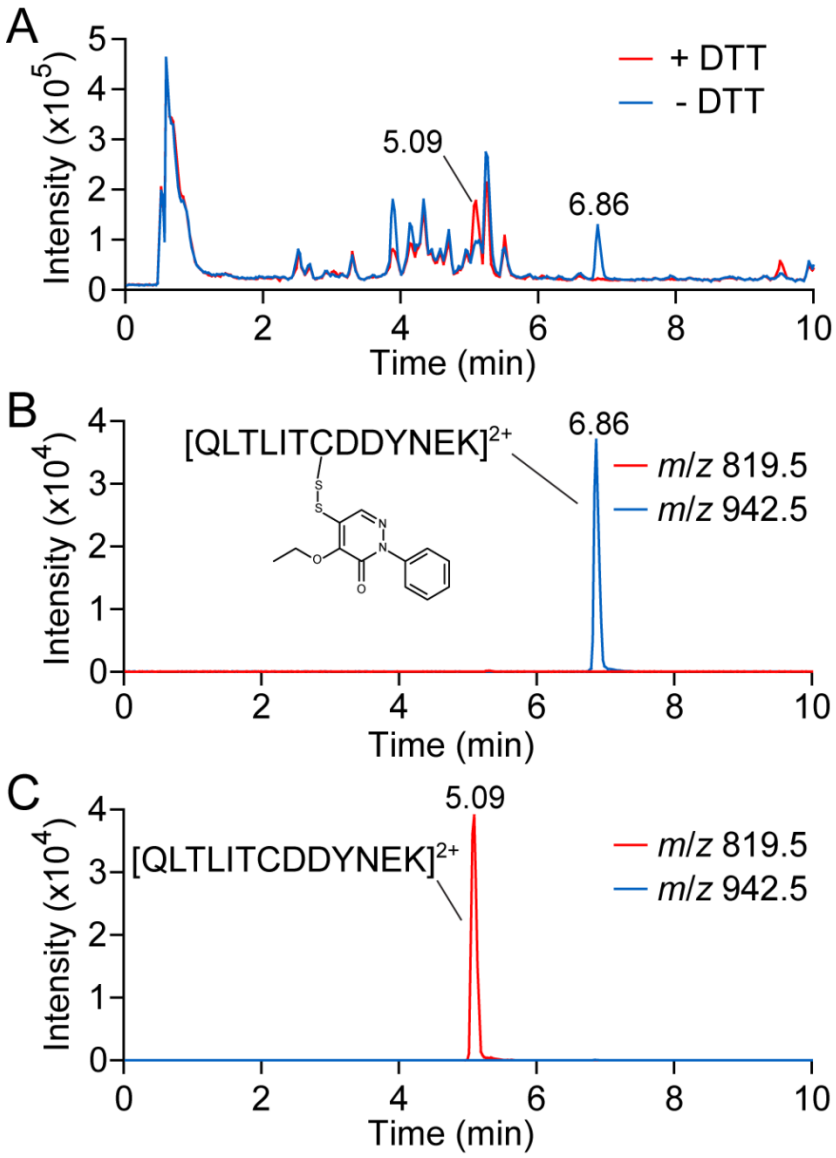


Figure 3.3. NMR spectra of Sa-SrtA:2-salt complex. *A*, ^{1}H-^{15}N HSQC of apo (red) and inhibitor bound (blue) Sa-SrtA after 24 hours of incubation. Peaks that had significant chemical shift changes ($\Delta\delta$) upon binding of the inhibitor are labeled. Significant chemical shift changes are defined as greater than the average Δδ + one standard deviation, where $\Delta\delta = \sqrt{(\Delta\delta_H)^2 + (\Delta\delta_N/6.49)^2}$, and $\Delta\delta_H$ and $\Delta\delta_N$ are the chemical shift changes in Hz for amide proton and amide nitrogen, respectively. *B*, selected panels showing intermolecular NOEs between Sa-SrtA and **2-salt**. 2D [F2] ^{13}C-filtered NOESY spectrum of the Sa-SrtA:inhibitor complex dissolved in $^{2}H_2O$. Identity and chemical shifts for each inhibitor proton are shown at the top and bottom of each panel, respectively. Each assigned cross peak is labeled with the corresponding proximal Sa-SrtA proton. Each assignment was verified in a 3D ^{13}C-editied NOESY by identifying a corresponding inhibitor cross peak. Not all cross peaks could be identified due to chemical shift ambiguities, including all NOESY cross peaks from the inhibitor Hβ protons.

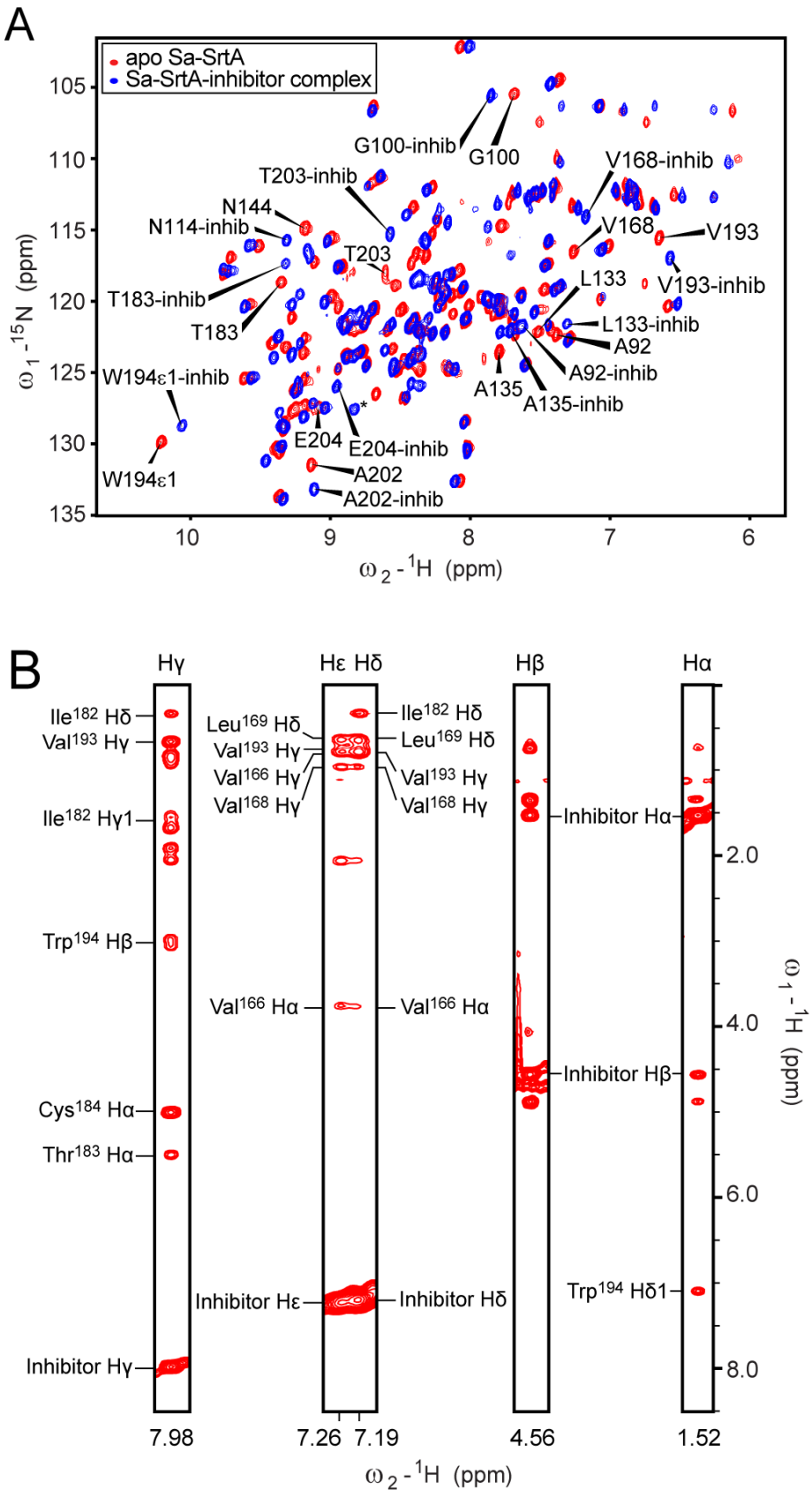
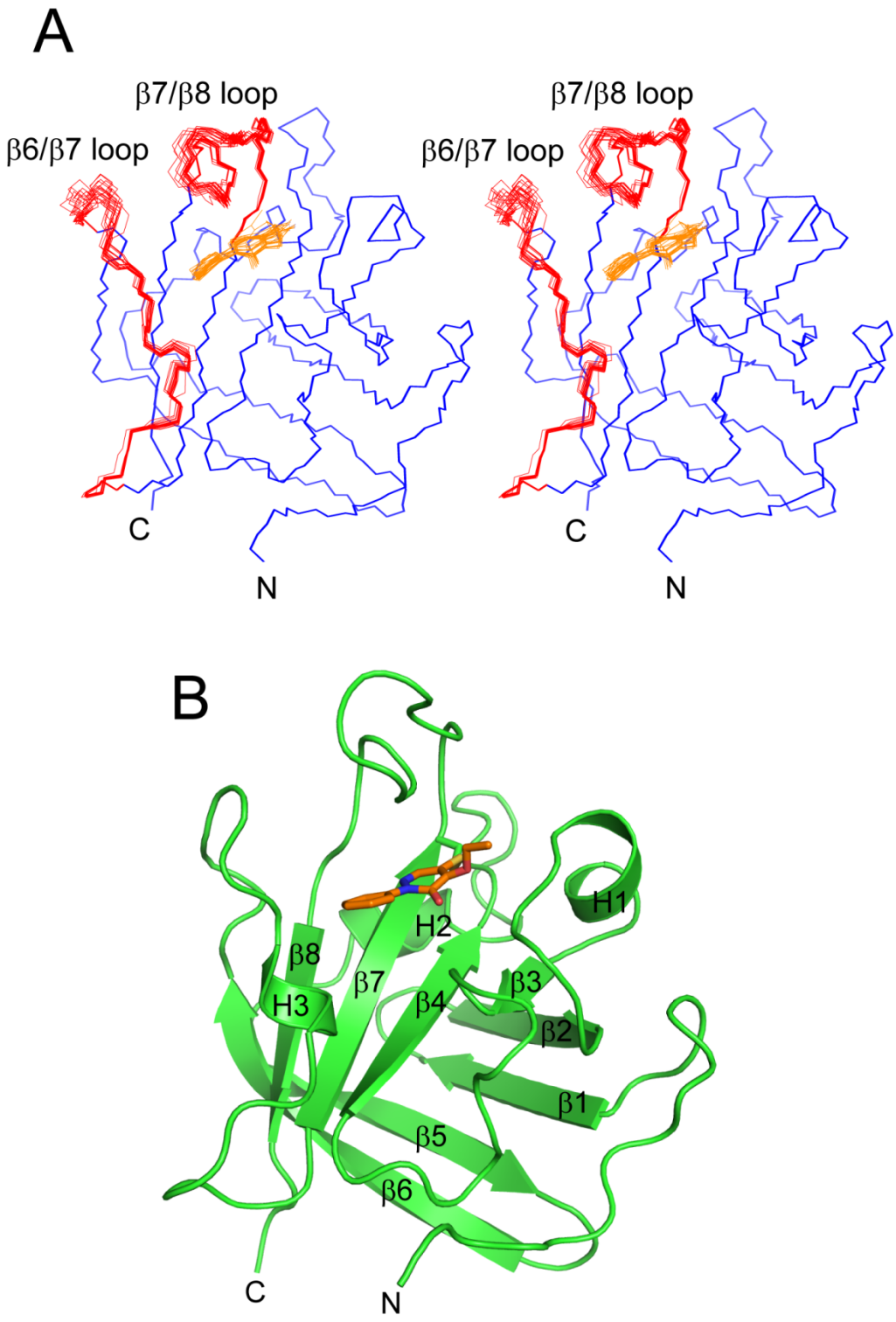
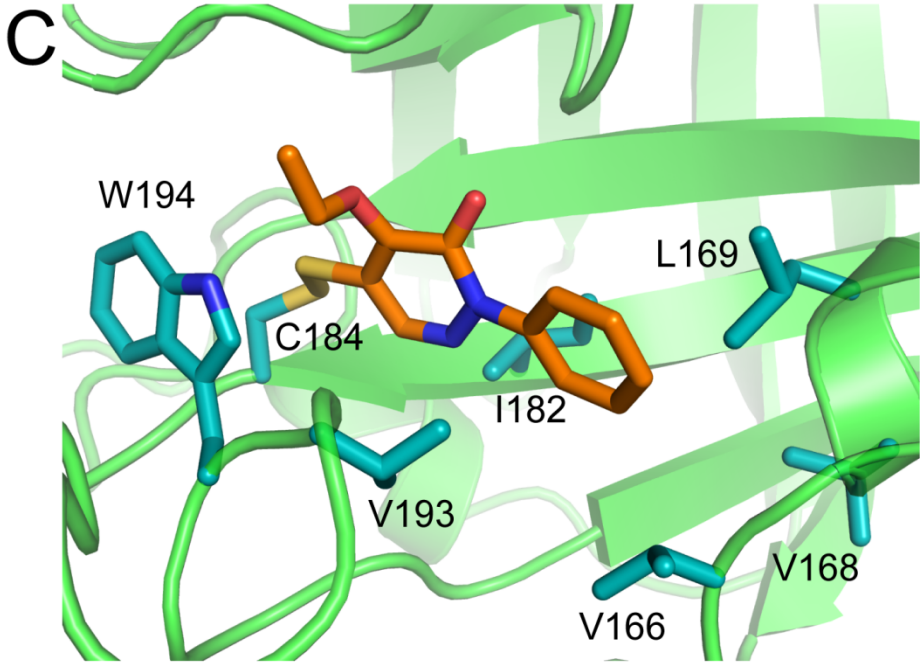

A
apo Sa-SrtA
Sa-SrtA-inhibitor complex
G100-inhib
G100
V168-inhib
T203-inhib
N144
N114-inhib
V168
V193
T203
L133
V193-inhib
T183-inhib
T183
L133-inhib
A92
A92-inhib
W194ε1-inhib
A135
A135-inhib
E204
E204-inhib
A202
W194ε1
A202-inhib
ω1 - 15N (ppm)
ω2 - 1H (ppm)
B
Hγ
Hε Hδ
Hβ
Hα
Ile182 Hδ
Val193 Hγ
Leu169 Hδ
Val193 Hγ
Val166 Hγ
Val168 Hγ
Ile182 Hγ1
Trp194 Hβ
Val166 Hα
Inhibitor Hα
Inhibitor Hβ
Cys184 Hα
Thr183 Hα
Inhibitor Hε
Inhibitor Hδ
Trp194 Hδ1
Inhibitor Hγ
7.98
7.26
7.19
4.56
1.52
ω2 - 1H (ppm)
ω1 - 1H (ppm)

Figure 3.4. NMR solution structure of the Sa-SrtA:2-salt complex. *A*, cross-eyed stereo image showing the ensemble of 20 lowest energy structures of the Sa-SrtA:**2-salt** complex. The bound compound **2-salt** is shown in orange, and protein backbone atoms allowed to move during molecular dynamics simulations are shown in red (Val^{161}-Lys^{175} and Thr^{183}-Lys^{196}). The bundle is aligned using all of the protein backbone heavy atoms and the heavy atoms from the inhibitor. *B*, ribbon structure of the Sa-SrtA:**2-salt** complex. The covalently bound inhibitor is shown in stick representation and colored orange. *C*, expanded view of the Sa-SrtA active site with the inhibitor bound. Side chains of residues exhibiting intermolecular NOEs to the inhibitor are shown in cyan sticks. A hydrophobic pocket formed by Val^{166}, Val^{168}, Leu^{169}, and Ile^{182} bind the 2-phenyl group of the inhibitor, while the remainder of the inhibitor is positioned over the β7 and β8 strands. *D*, expanded view of the Sa-SrtA active site with the protein represented by its solvent accessible surface. The surface is colored by its electrostatic properties from acidic (red) to basic (blue).


A
β7/β8 loop
β6/β7 loop
C
N
β7/β8 loop
β6/β7 loop
C
N
B
H1
H2
β8
β7
β3
H3
β4
β2
β1
β5
β6
C
N


C
W194
L169
C184
I182
V193
V168
V166

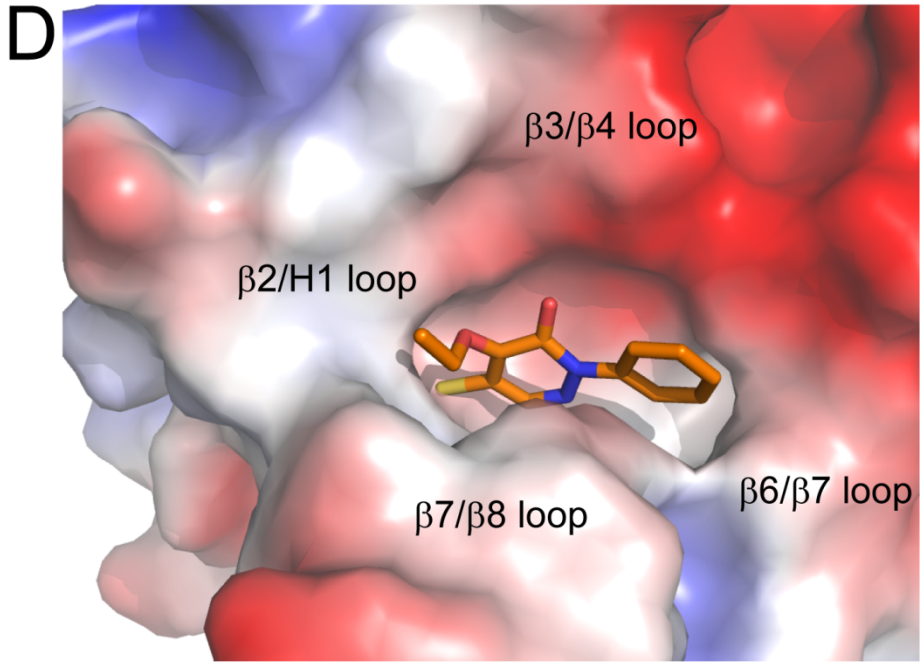

D
β3/β4 loop
β2/H1 loop
β7/β8 loop
β6/β7 loop

Figure 3.5. Overview of the molecular docking and experimental testing process. Descriptions of the docking, substituents selection, and experimental testing processes are shown on the right. The structure of the core of the pyridazinone compound is shown on the left. The number of substituents at each site is indicated in parenthesis.

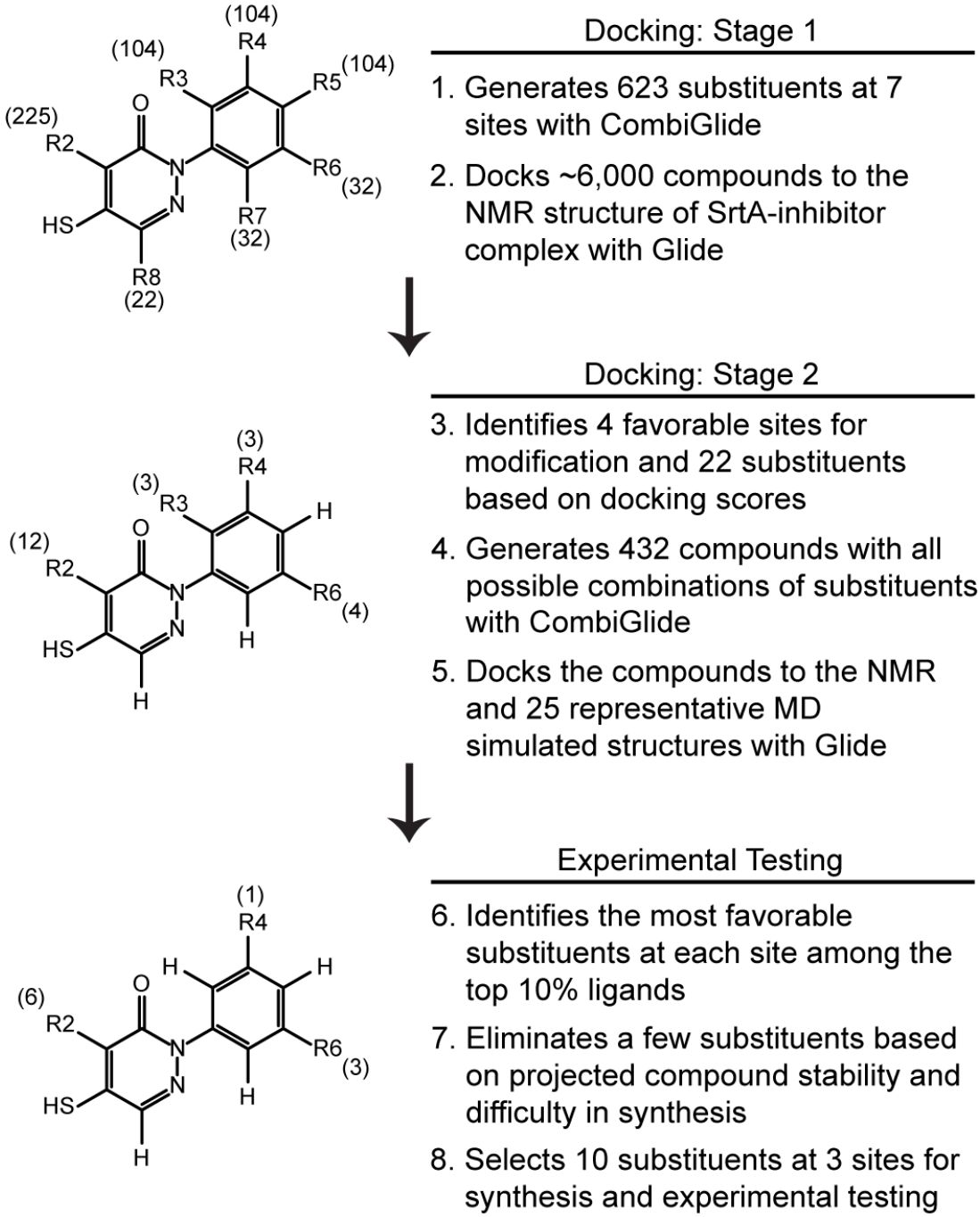


Figure 3.6. Determination of rate of Sa-SrtA inhibition by compound 2-62. The rate of Sa-SrtA inhibition by compound **2-62** was determined by calculating k_{obs} for each inhibitor concentration using equation 3 and subsequently calculating k_{inact} and K_I using equation 4 listed in section 3.6.5. This graph shows a representative curve of fitting the k_{obs} data with equation 4.

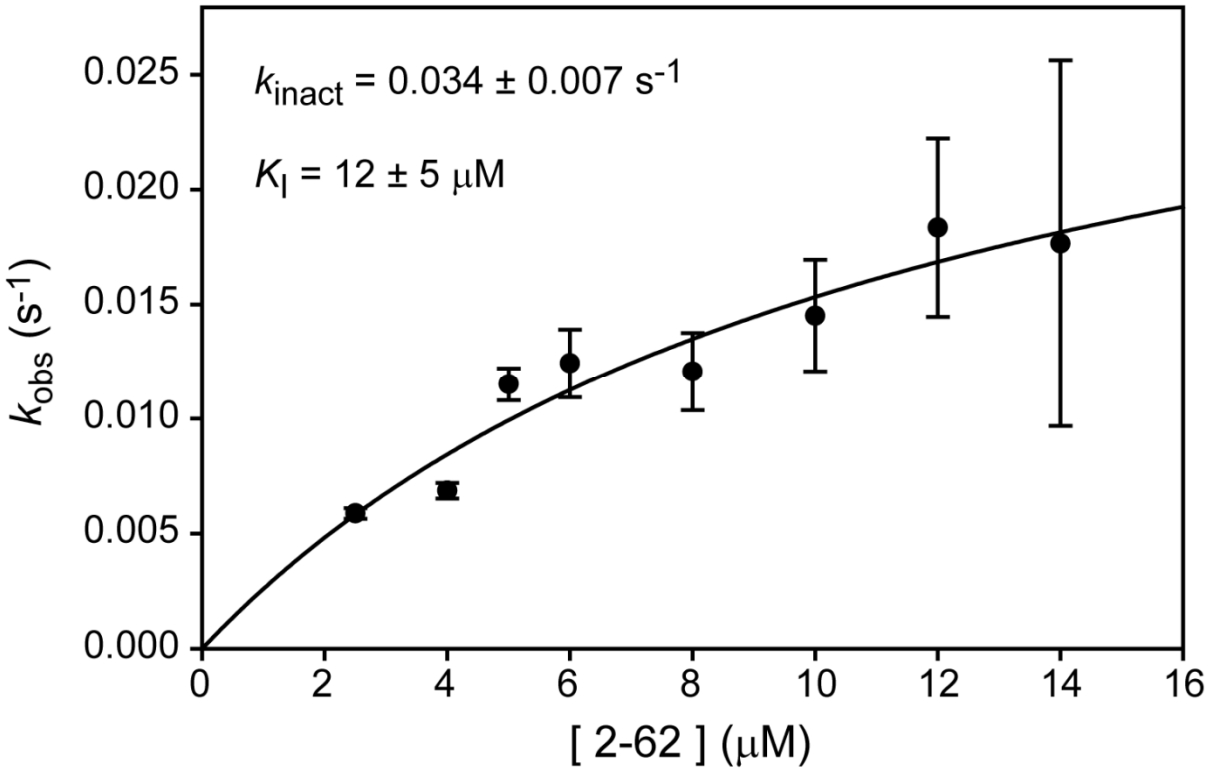


Figure 3.7. Comparison of the Sa-SrtA:2-salt structure with apo- and holo-Sa-SrtA structures. Sa-SrtA:**2-salt** complex (green) superimposed with (*A*) apo-Sa-SrtA (blue, PDB code 1IJA) and (*B*) Sa-SrtA:LPAT* complex with the substrate analog LPAT* removed (pink, PDB code 2KID). In apo-Sa-SrtA, residues Ser^{157}-Lys^{175} are unstructured in the $\beta6/\beta7$ loop. In inhibitor complex, a portion of this loop is immobilized. The $\beta6/\beta7$ loop in the Sa-SrtA:**2-salt** complex is unstructured from residues Thr^{156}-Asp^{165} similar to apo-Sa-SrtA, and adopts a single conformation from residues Val^{166}-Lys^{175}, as seen in the Sa-SrtA:LPAT* structure. The $\beta7/\beta8$ loop adopts a conformation more closely resembling the apo-Sa-SrtA structure. *C*, non-proline amide crosspeaks that were not detected in the ^{1}H-^{15}N HSQC spectra are shown in gray spheres. These residues include Met^{59}-Gln^{60}, Thr^{121}-Aps^{124}, Lys^{137}, Thr^{156}-Val^{166}, and Lys^{196}-Phe^{200}. *D*, expanded view of the $\beta6/\beta7$ loop 3_{10}-helix in the Sa-SrtA:**2-salt** complex with **2-salt** represented as orange sticks. *E*, expanded view of the $\beta6/\beta7$ loop 3_{10}-helix in the Sa-SrtA:LPAT* complex with LPAT* represented as yellow sticks.

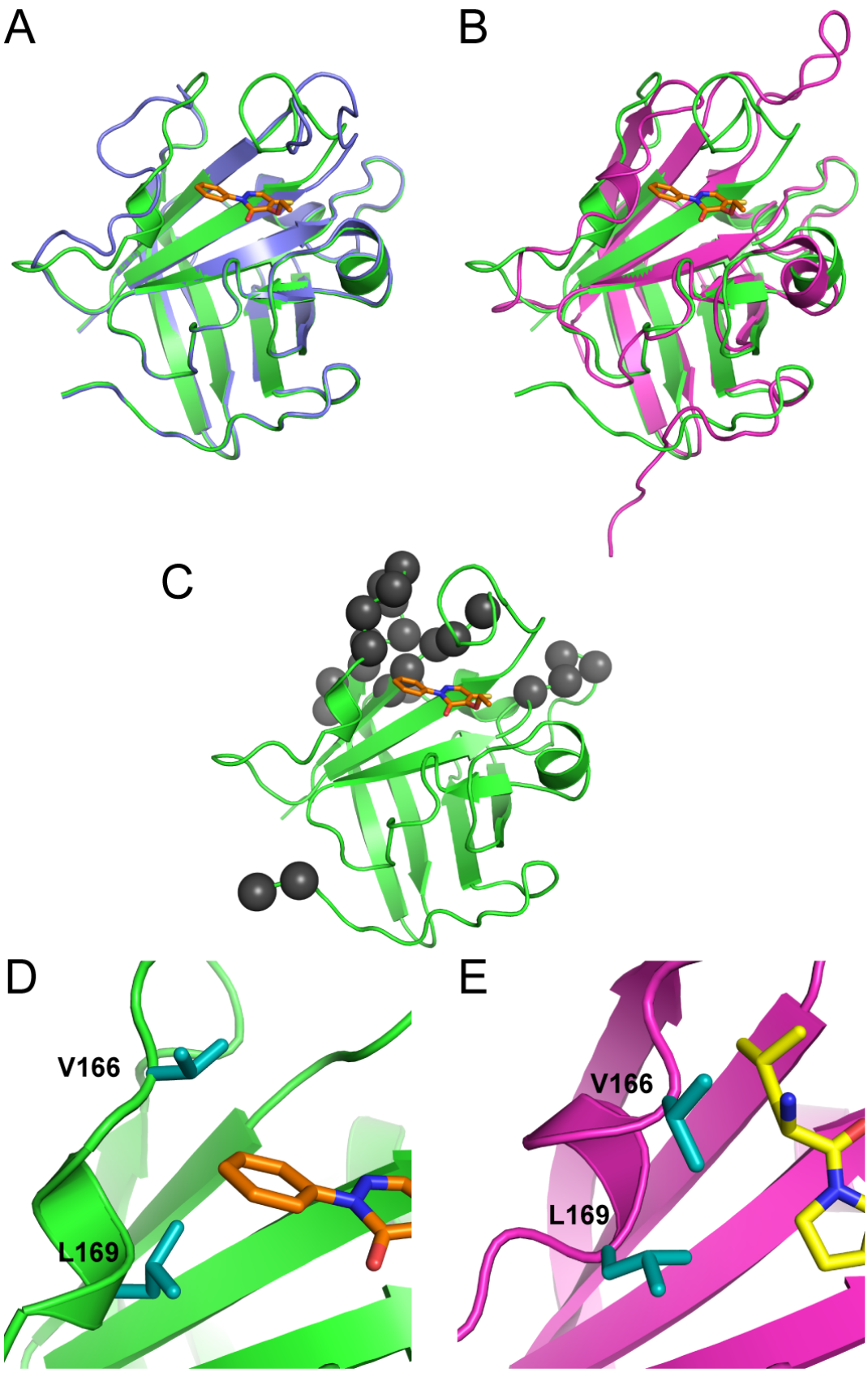
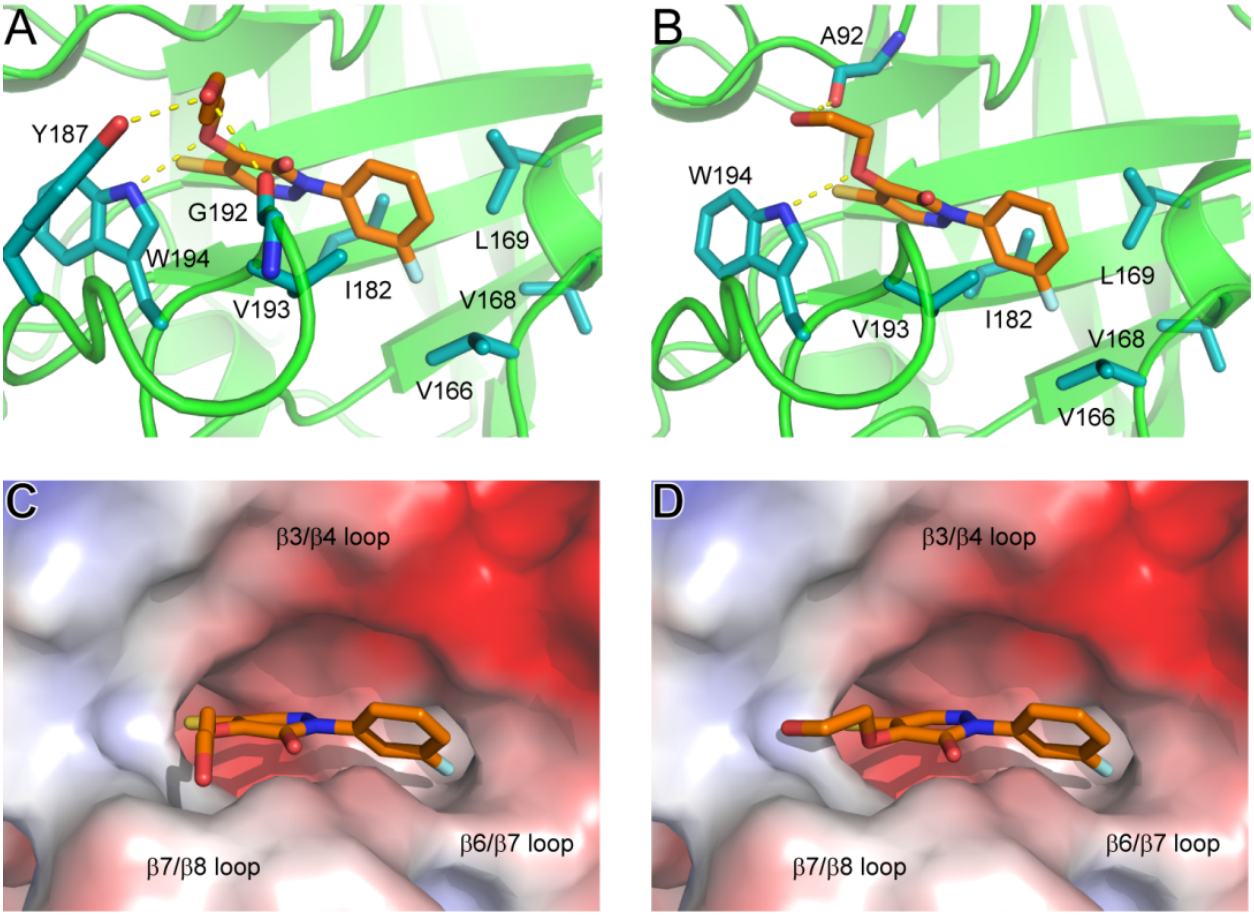

A
B
C
D
V166
L169
E
V166
L169

Figure 3.8. Docking poses of compounds 2-61 and 2-62. Expanded view of the active site showing interactions between the protein and compound **2-62** (*A*, *C*) and compound **2-61** (*B*, *D*). In *A* and *B*, residues that contact the compound are shown as cyan sticks. Hydrogen bonds are indicated by yellow dotted lines. In *C* and *D*, the protein solvent accessible surface is shown and colored by its electrostatic properties from acidic (red) to basic (blue).



3.8 Tables

Table 3.1 Statistics for the NMR modeled structure of Sa-SrtA bound to a pyridazinone inhibitor 2-salt

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

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

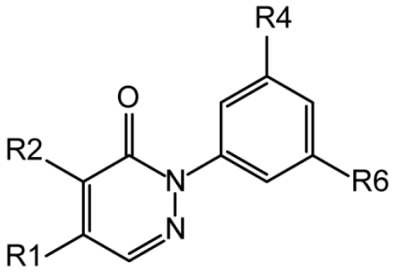
[b] Intramolecular NOE distance restraints include any NOE signals from any residue in the Sa-SrtA enzyme to residues Val^{166}-Lys^{175}, Thr^{183}-Lys^{196}, and the side chains from residues Leu^{97}, Ser^{116}, His^{120}, Thr^{180}, Ile^{182}, Arg^{197}, Ile^{199}, and Val^{201}.

[c] Experimental backbone dihedral angle restraints comprised 21 ϕ and 22 ψ angles within residues Val^{166}-Lys^{175} and Thr^{183}-Lys^{196}.

[d] PROCHECK-NMR data includes residues Val^{166}-Lys^{175} and Thr^{183}-Lys^{196} of the Sa-SrtA:inhibitor complex.

[e] The coordinate precision is defined as the average atomic r.m.s. deviation of the 20 individual simulated annealing structures and their mean coordinates. The reported values are for residues Val^{166}-Lys^{175} and Thr^{183}-Lys^{196} of the Sa-SrtA:inhibitor complex for the protein backbone r.m.s. deviation. In addition, residues Leu^{97}, Ser^{116}, His^{120}, Thr^{180}, Ile^{182}, Arg^{197}, Ile^{199}, and Val^{201} were included for calculating protein heavy atom r.m.s. deviation.

Table 3.2 SrtA inhibition of the pyridazinone compound derivatives



Compound	R1	R2	R4	R6	Sa-SrtA IC_{50} (μM)[a]	Ba-SrtA IC_{50} (μM)[b]	CC_{50} (μM)	cLogP
2-10	SH	OCH_2CH_3	H	H	13 ± 1[d]	3.2 ± 1.7[d]	n.d.[e]	3.03
2-17	S-SR[c]	OCH_2CH_3	H	H	1.5 ± 0.4[d]	1.2 ± 0.4[d]	~100	4.37
2-51	S-SR[c]	OCH_2CH_3	H	F	0.16 ± 0.04	1.34 ± 0.62	~100	4.55
2-52	S-SR[c]	OCH_2CH_3	H	CH_3	1.4 ± 0.3	0.49 ± 0.11	n.d.[e]	4.58
2-53	S-SR[c]	OCH_2CH_3	H	OCH_3	0.62 ± 0.18	0.146 ± 0.049	~100	4.46
2-54	S-SR[c]	OCH_2CH_3	$C(=NH)NH_2$	H	0.54 ± 0.14	2.08 ± 0.55	>100	2.72
2-55	S-SR[c]	$O(CH_2)_2OH$	H	H	0.219 ± 0.043	0.33 ± 0.14	>100	2.35
2-56	S-SR[c]	$O(CH_2)_3OH$	H	H	0.84 ± 0.06	1.86 ± 0.59	~100	2.80
2-57	S-SR[c]	OCH_2COOH	H	H	2.2 ± 0.2	0.49 ± 0.19	n.d.[e]	2.02
2-58	S-SR[c]	$O(CH_2)_2COOH$	H	H	6.1 ± 0.5	9.98 ± 0.37	>100	2.32
2-59	S-SR[c]	OCH_2CONH_2	H	H	22 ± 2	94 ± 23	n.d.[e]	2.11
2-60	S-SR[c]	$O(CH_2)_2CONH_2$	H	H	3.25 ± 0.34	11.9 ± 0.7	>100	2.06
2-61	S-SR[c]	$O(CH_2)_2OH$	H	F	0.41 ± 0.12	0.60 ± 0.23	>100	2.83
2-62	S-SR[c]	$O(CH_2)_3OH$	H	F	0.021 ± 0.014	0.45 ± 0.29	~100	3.43

[a] Or K_i^{app} for values that are lower than 0.5 μM as determined by the Morrison's equation

[b] Or K_i^{app} for values that are lower than 5 μM as determined by the Morrison's equation

[c] S-SR indicates that the compound is a disulfide bonded dimer

[d] Values taken from reference (42)

[e] n.d., no data

3.9 References

1. Rice, L. B. (2008) *J Infect Dis* **197**, 1079-1081
2. Boucher, H. W., Talbot, G. H., Bradley, J. S., Edwards, J. E., Gilbert, D., Rice, L. B., Scheld, M., Spellberg, B., and Bartlett, J. (2009) *Clin Infect Dis* **48**, 1-12
3. Lowy, F. D. (1998) *N Engl J Med* **339**, 520-532
4. Ippolito, G., Leone, S., Lauria, F. N., Nicastri, E., and Wenzel, R. P. (2010) *Int J Infect Dis* **14 Suppl 4**, S7-11
5. Centers for Disease Control and Prevention. Active Bacterial Core Surveillance Report, Emerging Infections Program Network, Methicillin-Resistant Staphylococcus aureus, 2011. http://www.cdc.gov/abcs/reports-findings/survreports/mrsa11.html.
6. Sakoulas, G., and Moellering, R. C., Jr. (2008) *Clin Infect Dis* **46 Suppl 5**, S360-367
7. Marty, F. M., Yeh, W. W., Wennersten, C. B., Venkataraman, L., Albano, E., Alyea, E. P., Gold, H. S., Baden, L. R., and Pillai, S. K. (2006) *J Clin Microbiol* **44**, 595-597
8. Ikeda-Dantsuji, Y., Hanaki, H., Sakai, F., Tomono, K., Takesue, Y., Honda, J., Nonomiya, Y., Suwabe, A., Nagura, O., Yanagihara, K., Mikamo, H., Fukuchi, K., Kaku, M., Kohno, S., Yanagisawa, C., Nakae, T., Yoshida, K., and Niki, Y. (2011) *J Infect Chemother* **17**, 45-51
9. Ton-That, H., Liu, G., Mazmanian, S. K., Faull, K. F., and Schneewind, O. (1999) *Proc Natl Acad Sci U S A* **96**, 12424-12429
10. Mazmanian, S. K., Liu, G., Ton-That, H., and Schneewind, O. (1999) *Science* **285**, 760-763
11. Schneewind, O., and Missiakas, D. M. (2012) *Philos Trans R Soc Lond B Biol Sci* **367**, 1123-1139

12. Spirig, T., Weiner, E. M., and Clubb, R. T. (2011) *Mol Microbiol* **82**, 1044-1059

13. Clancy, K. W., Melvin, J. A., and McCafferty, D. G. (2010) *Biopolymers* **94**, 385-396

14. Marraffini, L. A., Dedent, A. C., and Schneewind, O. (2006) *Microbiol Mol Biol Rev* **70**, 192-221

15. Mazmanian, S. K., Liu, G., Jensen, E. R., Lenoy, E., and Schneewind, O. (2000) *Proc Natl Acad Sci U S A* **97**, 5510-5515

16. Jonsson, I. M., Mazmanian, S. K., Schneewind, O., Verdrengh, M., Bremell, T., and Tarkowski, A. (2002) *J Infect Dis* **185**, 1417-1424

17. Weiss, W. J., Lenoy, E., Murphy, T., Tardio, L., Burgio, P., Projan, S. J., Schneewind, O., and Alksne, L. (2004) *J Antimicrob Chemother* **53**, 480-486

18. Zink, S. D., and Burns, D. L. (2005) *Infect Immun* **73**, 5222-5228

19. Bierne, H., Mazmanian, S. K., Trost, M., Pucciarelli, M. G., Liu, G., Dehoux, P., Jansch, L., Garcia-del Portillo, F., Schneewind, O., and Cossart, P. (2002) *Mol Microbiol* **43**, 869-881

20. Garandeau, C., Reglier-Poupet, H., Dubail, I., Beretti, J. L., Berche, P., and Charbit, A. (2002) *Infect Immun* **70**, 1382-1390

21. Paterson, G. K., and Mitchell, T. J. (2006) *Microbes Infect* **8**, 145-153

22. Chen, S., Paterson, G. K., Tong, H. H., Mitchell, T. J., and DeMaria, T. F. (2005) *FEMS Microbiol Lett* **253**, 151-154

23. Kharat, A. S., and Tomasz, A. (2003) *Infect Immun* **71**, 2758-2765

24. Heilbronner, S., Hanses, F., Monk, I. R., Speziale, P., and Foster, T. J. (2013) *Microbiology* **159**, 2141-2152

25. Maresso, A. W., and Schneewind, O. (2008) *Pharmacol Rev* **60**, 128-141

26. Suree, N., Jung, M. E., and Clubb, R. T. (2007) *Mini Rev Med Chem* **7**, 991-1000

27. Kim, S. H., Shin, D. S., Oh, M. N., Chung, S. C., Lee, J. S., Chang, I. M., and Oh, K. B. (2003) *Biosci Biotechnol Biochem* **67**, 2477-2479

28. Kim, S. H., Shin, D. S., Oh, M. N., Chung, S. C., Lee, J. S., and Oh, K. B. (2004) *Biosci Biotechnol Biochem* **68**, 421-424

29. Kim, S. W., Chang, I. M., and Oh, K. B. (2002) *Biosci Biotechnol Biochem* **66**, 2751-2754

30. Oh, K. B., Mar, W., Kim, S., Kim, J. Y., Oh, M. N., Kim, J. G., Shin, D., Sim, C. J., and Shin, J. (2005) *Bioorg Med Chem Lett* **15**, 4927-4931

31. Jang, K. H., Chung, S. C., Shin, J., Lee, S. H., Kim, T. I., Lee, H. S., and Oh, K. B. (2007) *Bioorg Med Chem Lett* **17**, 5366-5369

32. Kang, S. S., Kim, J. G., Lee, T. H., and Oh, K. B. (2006) *Biol Pharm Bull* **29**, 1751-1755

33. Park, B. S., Kim, J. G., Kim, M. R., Lee, S. E., Takeoka, G. R., Oh, K. B., and Kim, J. H. (2005) *J Agric Food Chem* **53**, 9005-9009

34. Jeon, J. E., Na, Z., Jung, M., Lee, H. S., Sim, C. J., Nahm, K., Oh, K. B., and Shin, J. (2010) *J Nat Prod* **73**, 258-262

35. Lee, Y. J., Han, Y. R., Park, W., Nam, S. H., Oh, K. B., and Lee, H. S. (2010) *Bioorg Med Chem Lett* **20**, 6882-6885

36. Oh, I., Yang, W. Y., Chung, S. C., Kim, T. Y., Oh, K. B., and Shin, J. (2011) *Arch Pharm Res* **34**, 217-222

37. Won, T. H., Jeon, J. E., Kim, S. H., Lee, S. H., Rho, B. J., Oh, D. C., Oh, K. B., and Shin, J. (2012) *J Nat Prod* **75**, 2055-2061

38. Won, T. H., Jeon, J. E., Lee, S. H., Rho, B. J., Oh, K. B., and Shin, J. (2012) *Bioorg Med Chem* **20**, 4082-4087

39. Bae, J., Jeon, J. E., Lee, Y. J., Lee, H. S., Sim, C. J., Oh, K. B., and Shin, J. (2011) *J Nat Prod* **74**, 1805-1811

40. Julianti, E., Lee, J. H., Liao, L., Park, W., Park, S., Oh, D. C., Oh, K. B., and Shin, J. (2013) *Org Lett* **15**, 1286-1289

41. Maresso, A. W., Wu, R., Kern, J. W., Zhang, R., Janik, D., Missiakas, D. M., Duban, M. E., Joachimiak, A., and Schneewind, O. (2007) *J Biol Chem* **282**, 23129-23139

42. Suree, N., Yi, S. W., Thieu, W., Marohn, M., Damoiseaux, R., Chan, A., Jung, M. E., and Clubb, R. T. (2009) *Bioorg Med Chem* **17**, 7174-7185

43. Oh, K. B., Kim, S. H., Lee, J., Cho, W. J., Lee, T., and Kim, S. (2004) *J Med Chem* **47**, 2418-2421

44. Oh, K. B., Nam, K. W., Ahn, H., Shin, J., Kim, S., and Mar, W. (2010) *Biochem Biophys Res Commun* **396**, 440-444

45. Kruger, R. G., Barkallah, S., Frankel, B. A., and McCafferty, D. G. (2004) *Bioorg Med Chem* **12**, 3723-3729

46. Kudryavtsev, K. V., Bentley, M. L., and McCafferty, D. G. (2009) *Bioorg Med Chem* **17**, 2886-2893

47. Jung, M. E., Clemens, J. J., Suree, N., Liew, C. K., Pilpa, R., Campbell, D. O., and Clubb, R. T. (2005) *Bioorg Med Chem Lett* **15**, 5076-5079

48. Liew, C. K., Smith, B. T., Pilpa, R., Suree, N., Ilangovan, U., Connolly, K. M., Jung, M. E., and Clubb, R. T. (2004) *FEBS Lett* **571**, 221-226

49. Connolly, K. M., Smith, B. T., Pilpa, R., Ilangovan, U., Jung, M. E., and Clubb, R. T. (2003) *J Biol Chem* **278**, 34061-34065

50. Scott, C. J., McDowell, A., Martin, S. L., Lynas, J. F., Vandenbroeck, K., and Walker, B. (2002) *Biochem J* **366**, 953-958

51. Chan, A. H., Wereszczynski, J., Amer, B. R., Yi, S. W., Jung, M. E., McCammon, J. A., and Clubb, R. T. (2013) *Chem Biol Drug Des*

52. Chenna, B. C., Shinkre, B. A., King, J. R., Lucius, A. L., Narayana, S. V., and Velu, S. E. (2008) *Bioorg Med Chem Lett* **18**, 380-385

53. Uddin, R., and Saeed, K. (2014) *Mol Biol Rep*

54. Zhang, R., Wu, R., Joachimiak, G., Mazmanian, S. K., Missiakas, D. M., Gornicki, P., Schneewind, O., and Joachimiak, A. (2004) *Structure* **12**, 1147-1156

55. Suree, N., Liew, C. K., Villareal, V. A., Thieu, W., Fadeev, E. A., Clemens, J. J., Jung, M. E., and Clubb, R. T. (2009) *J Biol Chem* **284**, 24465-24477

56. Kappel, K., Wereszczynski, J., Clubb, R. T., and McCammon, J. A. (2012) *Protein Sci* **21**, 1858-1871

57. Moritsugu, K., Terada, T., and Kidera, A. (2012) *J Am Chem Soc* **134**, 7094-7101

58. Naik, M. T., Suree, N., Ilangovan, U., Liew, C. K., Thieu, W., Campbell, D. O., Clemens, J. J., Jung, M. E., and Clubb, R. T. (2006) *J Biol Chem* **281**, 1817-1826

59. Ilangovan, U., Ton-That, H., Iwahara, J., Schneewind, O., and Clubb, R. T. (2001) *Proc Natl Acad Sci U S A* **98**, 6056-6061

60. Tetko, I. V., Gasteiger, J., Todeschini, R., Mauri, A., Livingstone, D., Ertl, P., Palyulin, V. A., Radchenko, E. V., Zefirov, N. S., Makarenko, A. S., Tanchuk, V. Y., and Prokopenko, V. V. (2005) *J Comput Aided Mol Des* **19**, 453-463

61. Tetko, I. V., and Tanchuk, V. Y. (2002) *J Chem Inf Comput Sci* **42**, 1136-1145

62. Bethel, C. R., Taracila, M., Shyr, T., Thomson, J. M., Distler, A. M., Hujer, K. M., Hujer, A. M., Endimiani, A., Papp-Wallace, K., Bonnet, R., and Bonomo, R. A. (2011) *Antimicrob Agents Chemother* **55**, 3465-3475

63. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J Biomol NMR* **6**, 277-293

64. Garrett, D. S., Powers, R., Gronenborn, A. M., and Clore, G. M. (1991) *J Magn Reson* **95**, 214-220

65. Keller, R. (2004) *The Computer Aided Resonance Assignment Tutorial*, CATINA Verlag, Goldau, Switzerland

66. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., and Skelton, N. J. (2006) *Protein NMR spectroscopy*, Elsevier Science and Technology, San Diego, CA

67. Teng, Q. (2005) *Structural Biology: Practical NMR Applications*, Springer Verlag, New York

68. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) *J Biomol NMR* **44**, 213-223

69. Vuister, G. W., and Bax, A. (1993) *J Am Chem Soc* **115**, 7772-7777

70. Herrmann, T., Guntert, P., and Wuthrich, K. (2002) *J Biomol NMR* **24**, 171-189

71. Herrmann, T., Guntert, P., and Wuthrich, K. (2002) *J Mol Biol* **319**, 209-227

72. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) *J Magn Reson* **160**, 65-73

73. Zong, Y., Bice, T. W., Ton-That, H., Schneewind, O., and Narayana, S. V. (2004) *J Biol Chem* **279**, 31383-31389

74. Koradi, R., Billeter, M., and Wuthrich, K. (1996) *J Mol Graph* **14**, 51-55, 29-32

75. DeLano, W. L. (2006) The PyMOL Molecular Graphics System. DeLano Scientific, LLC, Palo Alto, CA

76. Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013) *J Comput Aided Mol Des* **27**, 221-234

77. Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., and Banks, J. L. (2004) *J Med Chem* **47**, 1750-1759

78. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004) *J Med Chem* **47**, 1739-1749

79. Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., and Mainz, D. T. (2006) *J Med Chem* **49**, 6177-6196

80. Wang, J., Wang, W., Kollman, P. A., and Case, D. A. (2006) *J Mol Graph Model* **25**, 247-260

81. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) *J Comput Chem* **25**, 1157-1174

82. Cornell, W. D., Cieplak, P., Bayly, C. I., and Kollman, P. A. (1993) *Journal of the American Chemical Society* **115**, 9620-9631

83. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery Jr, J. A., Peralta, J. E., Ogliaro, F. o., Bearpark, M. J., Heyd, J., Brothers, E. N., Kudin, K. N., Staroverov, V. N.,

Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A. P., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, N. J., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, Ã. d. n., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. (2009) Gaussian 09. Gaussian, Inc., Wallingford, CT, USA

84. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. (2010) *Proteins* **78**, 1950-1958

85. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) *J Comput Chem* **26**, 1781-1802

86. Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B., and Lindahl, E. (2013) *Bioinformatics* **29**, 845-854

87. Copeland, R. A. (2005) *Evaluation of Enzyme Inhibitors in Drug Discoveries: A Guide for Medicinal Chemists and Pharmacologists*, John Wiley & Sons, New Jersey

88. Crouch, S. P., Kozlowski, R., Slater, K. J., and Fletcher, J. (1993) *J Immunol Methods* **160**, 81-88

Chapter 4

The Structure and Dynamics of the *Bacillus anthracis* Sortase-Substrate Complex Reveals a Regulatory Role of the N-terminal Extension in Substrate Entry

4.1 Overview

The housekeeping class A sortase is found in almost all Gram-positive pathogens. Previously, many biochemical, structural and inhibitor studies have focused on the class A *Staphylcoccus aureus* sortase A enzyme, while much less is known about sortases from other Gram-positive pathogens. To gain more insight into the substrate recognition and catalytic mechanism, recently we solved the structure of the sortase A enzyme from *Bacillus anthracis* and noticed several unusual active site features, including a structured N-terminal extension preceding the catalytic core that covers the active site, a seemingly preformed substrate-contacting β6/β7 loop, and an unstructured β7/β8 loop that connects the active site cysteine residue in strand β7 to the active site arginine residue in strand β8. To understand how substrate binding affects these special elements in *B. anthracis* sortase A (Ba-SrtA), I solved the structure of Ba-SrtA in complex with a covalently attached substrate analog using NMR methods. Interestingly, I found that substrate binding leads to the ordering of the β7/β8 loop and induces more frequent unlatching of the N-terminal extension from the active site, which I hypothesize play critical role on cell surface by modulating substrate entry into the enzyme active site.

This chapter is a modified version of a manuscript that will be submitted for publication in a peer-reviewed journal.

4.2 Introduction

Bacillus anthracis is a Gram-positive bacterium that causes lethal anthrax disease in humans and animals. The spores of the bacteria represent the infectious form of the pathogen,

which predominantly enter the human body through inhalation, ingestion or cuts in the skin (1). Cutaneous anthrax results from spores entering through cuts in the skin represents the most common form of the disease and is readily treated (1). In contrast, gastrointestinal and pulmonary anthrax caused by ingestion and inhalation of spores, respectively, are are often fatal (92% mortality rate) if untreated before the illness progresses to the fulminant phase (2). Regardless of the entry route, spores are initially taken up by macrophages in which they germinate and initiate expression of toxin genes (3). Despite being an extracellular bacterium, this intracellular step appears to be required for *B. anthracis* to establish infection (1). Because of the high mortality rate of anthrax and the relative ease of distribution of its spores, *B. anthracis* is an established bioterrorism agent. This has driven the search for new therapeutics to treat and prevent anthrax.

Many surface proteins in bacteria are considered virulence factors as they enable pathogens to adhere to sites of infection, acquire essential nutrients, enter host cells, and evade the immune response (4). In Gram-positive bacteria, many surface proteins are covalently attached to the cell wall by the sortase enzymes, which constitute a large family of membrane-associated transpeptidases (5,6). Sortases covalently join proteins containing a C-terminal cell wall sorting signal that contains a signal motif, followed by a hydrophobic transmembrane segment and a positively charged C-terminal tail, to the free amino group within the cell wall precursor lipid II (6). The proteins are then displayed on the cell surface after the lipid II linked protein products are incorporated into the peptidoglycan by the transpeptidation and transglycosylation reactions of cell wall synthesis (6). Some members of the sortase enzyme family can also polymerize protein subunits containing sorting signals and pilin motifs into pili, filamentous proteinaceous structures that promote bacterial adhesion. Because sortases can

attach proteins bearing specific sorting signal to a variety of biomolecules, they are valuable tools in molecular biology (7). Most importantly, since many Gram-positive pathogens require a functioning sortase to be fully virulent, sortase enzymes have become promising therapeutic targets for novel antibiotics development (8,9).

Sortases enzymes are widely distributed in Gram positive bacteria and can be classified into six subfamilies based on their primary sequences, called class A to F enzymes (6). Members of the different families frequently perform different functions in the cell and have evolved to recognize unique sorting signals. The genome of *B. anthracis* encodes for class A, B, and D sortases known as Ba-SrtA, Ba-SrtB, and Ba-SrtC, respectively. Ba-SrtA anchors seven proteins to the cell wall by joining the threonine of the LP*X*TG sorting signal at the C-terminus of the protein to the amine group of *meso*-diaminopimelic acid (*m*-DAP) within lipid II (10). Ba-SrtB anchors only one protein, IsdC, which is important for heme-binding (11). Ba-SrtC attaches two proteins, BasH and BasI, to the cell wall for proper spore formation (12). Ba-SrtA has been identified as a potential therapeutic target since it is required for *B. anthracis* to survive and replicate within alveolar macrophages, which is an important early step in the development of inhalation anthrax (13).

The archetypal *Staphylococcus aureus* sortase enzyme (called Sa-SrtA) was the first member of the family to be discovered and has been studied in detail (5,14). This class A sortase recognizes sorting signal motifs that contain the Leu-Pro-*X*-Thr-Gly (LP*X*TG, where *X* is any amino acid) sequence. Bioinformatic and mutagenesis studies have shown that all sortase enzymes contain three conserved active site residues that are critically important in the enzymatic activity: His^{120}, Cys^{184} and Arg^{197} (Sa-SrtA numbering) (15-17). Catalysis occurs

through a ping-pong mechanism that is initiated when the thiol group of Cys^{184} nucleophilically attacks the carbonyl carbon of the threonine residue within the LP*X*TG motif of the precursor surface protein (18,19). This forms the first transient tetrahedral intermediate which rearranges into a more stable thioacyl intermediate upon breakage of the threonine-glycine peptide bond. Sa-SrtA then joins the terminal amine group of the pentaglycine bridge of lipid II to the carbonyl carbon of the threonine, forming a second transient tetrahedral intermediate that is resolved into the lipid II-linked protein product. Although the role of Cys^{184} is well established, the functions of His^{120} and Arg^{197} are not as clear. Originally, His^{120} was proposed to activate Cys^{184} by forming an imidazolium-thiolate ion pair (15), but *p*Ka analysis revealed that both His^{120} and Cys^{184} are predominantly uncharged at physiological pH (19). This suggests that Sa-SrtA operates under reverse protonation mechanism in which only a small fraction (~0.06%) of Sa-SrtA is active, and that His^{120} functions as a general acid/base in catalysis (17,19). Arg^{197} plays an essential role in catalysis and has been proposed to stabilize the substrate (20-22) and oxyanion on the tetrahedral intermediate (17,23). Recent structural studies suggest that it interacts with the threonine residue in the substrate to form a substrate stabilized oxyanion hole (24).

How sortase enzymes recognize their sorting signal and nucleophile substrates is poorly understood. Although several structures of sortase enzymes have been determined revealing that members of this enzyme superfamily adopt an eight-stranded β-barrel fold, only two atomic-level structures of sortase enzymes covalently bound to their sorting signal substrates have thus been determined (20,24). It is therefore largely unknown how sortases have evolved to recognize a range of sorting signal substrates and nucleophiles, and how these enzymes regulate access of these substrates to their active sites, which has limited the design of substrate based inhibitors.

Here we present the NMR structure of Ba-SrtA covalently bound to its sorting signal substrate analog. We found that while the majority of the residues in Ba-SrtA undergo minimal structural changes after binding to the substrate analog, the N-terminal extension that used to contact the active site becomes more mobile, and a loop the connects strands β7 and β8 becomes more rigid and adopt a distinct structure. To explore the role of the N-terminal extension, we tested the kinetics of several N-terminal extension mutants, and found that it plays an inhibitory role in enzyme hydrolysis reaction.

4.3 Results

4.3.1 Structural basis of sorting signal recognition by the *B. anthracis* SrtA enzyme

To study the molecular basis of sorting signal recognition in *B. anthracis* SrtA, we synthesized a peptide analog of the sorting signal that covalently modifies the enzyme. The peptide, Boc-LPAT*, contains an N-terminal protecting group t-butoxycarbonyl (Boc) and a modified threonine (T*) that replaces the carbonyl group with -CH_2-SH (Fig. 4.1*A*). The thiol group on T* promotes the formation of a disulfide bond between the peptide and the active site Cys^{187} of the enzyme to generate a covalent complex that structurally mimics the thioacyl intermediate of catalysis. In this work, we used a Ba-SrtA construct that has its first 56 residues removed from the N-terminus (hereafter referred to as ^{Ba}SrtA), as we have shown previously that it provides good quality NMR spectra and has hydrolytic activity *in vitro* (25).

The solution structure of the complex was determined using multidimensional heteronuclear NMR and simulated annealing methods. A total of 2,689 experimental restraints were used to determine the structure of the complex, including: 1,802 intraprotein distances, 59

intermolecular distances, 318 dihedral angles, 77 $^{3}J_{HN\alpha}$, 282 ^{13}C secondary shifts, and 151 residual dipolar coupling restraints. An ensemble containing 20 conformers representing the structure of the complex is shown in Fig. 4.2*A*. These conformers possess good covalent geometry with no NOE, dihedral angle, or scalar coupling violations greater than 0.5 Å, 5°, or 2 Hz, respectively. In the ensemble, amino acids Asp^{57}-Lys^{210} in the protein and the bound peptide are well defined; the backbone and heavy atom coordinate root mean square deviation (r.m.s. deviation) to the average structure are 0.42 ± 0.06 and 0.89 ± 0.06 Å, respectively. Complete structure and restraint statistics are presented in Table 4.1.

In the structure of the complex, the enzyme recognizes the sorting signal through a large groove that leads into the active site (Fig. 4.2*B*). The peptide rests on top of a floor formed by residues in strands β4 and β7, and is bound by the walls formed by surface loops that connect strand β6 to β7 (β6/β7 loop), strand β3 to strand β4 (β3/β4 loop), and strand β2 to helix H2 (β2/H2 loop). In addition, the top of the peptide is covered by residues in the beginning of the N-terminal loop (N-terminal extension). Specifically, the leucine residue of the peptide is inserted into a pocket created by the β6/β7 loop, strand β6 and strand β8. It is contacted by the α-protons of Pro^{168}, Asp^{169} and Lys^{170} on the top of the β6/β7 loop, and the α-proton of Trp^{171} and side chains of Val^{173} and Val^{174} in the 3_{10}-helix H4 (Fig. 4.2*D*). The proline residue, which forms a ~90° kink to redirect the trajectory of the rest of the peptide toward the active site Cys^{187}, is enclosed by the side chains projected from the floor of the groove (Ala^{124} and Ile^{185} in strands β4 and β7, respectively), the walls (Val^{174}, Val^{110} and Ser^{98} in β6/β7 loop, β3/β4 loop and β2/H2 loop, respectively), and Ser^{59} in the N-terminal extension. The alanine residue in the peptide mimics the *X* position of the LP*X*TG motif. Its side chain methyl points toward helix H2 and is sandwiched between Leu^{103} in helix H2 and Ile^{61} in the N-terminal extension. This is consistent

with the documented promiscuity of SrtA for this site within the sorting signal since a larger residue, such as a lysine in the SrtA substrate BasD, can project its side chain through the gap between the N-terminal extension and helix H2 into the solvent. The side chain of the modified threonine is buried in a pocket formed by the side chains of Ile^{185} in strand β7, Ala^{124} in strand β4, Leu^{103} in helix H2 and the active site His^{126}. In particular, its γ-methyl group is packed underneath the imidazole ring of the active site His^{126}. This interaction is substantiated by the intermolecular NOE crosspeaks between the γ-methyl group of the modified threonine and the ε1- and α-protons of His^{126} (Fig. 4.1*B*).

4.3.2 Unique N-terminal extension interacts with the bound sorting signal

The catalytic domain of ^{Ba}SrtA is preceded by a uniquely structured N-terminal polypeptide segment that partially shields the sorting signal peptide and active site His^{126} residue from the solvent. This N-terminal extension encompasses residues Asp^{57} to Val^{79}. The first eight residues, Asp^{57} to Pro^{64}, cover the peptide and active site histidine, and are wedged between helix H2, the β4/β5 and β7/β8 loops. The remainder of the N-terminal extension wraps around surface of the protein so as to contact helix H2 and β2/H2 loop, and contains a short helix (H1, residues Leu^{65} to Asn^{71}) that occupies the space between helix H2 and the beginning of β4/β5 loop (Fig. 4.2*B*). Two residues, Ser^{59} and Ile^{61}, within the N-terminal extension interact with the bound peptide (Fig. 4.2*C*). The non-polar side chain of Ile^{61} interacts with the "*X*" residue in the bound peptide (alanine in the ^{Ba}SrtA-LPAT* complex) and partially shields it from the solvent. It also forms extensive contacts to the catalytically essential active site histidine so as to limit its solvent exposure. The side chain of Ser^{59} is also poised to interact with the peptide in the

complex. An inspection of the ensemble suggests that its hydroxyl group is appropriately positioned to donate a hydrogen bond to the backbone carbonyl oxygen of proline in the bound peptide (this interaction can be inferred from the distances between potential hydrogen bond donors and acceptors in 30% of the conformers in the NMR ensemble). Several intermolecular NOEs position the N-terminal extension near the peptide. These include NOE crosspeaks between α-proton of Ser^{59} and β-methyl group of alanine in the peptide, β-protons of Ser^{59} and β- and γ-protons of proline in the peptide, and α-proton of Ile^{61} and β-methyl group of alanine in the peptide (Fig. 4.1*B*). The capping interactions observed in the ^{Ba}SrtA-LPAT* structure have not been observed in structures of other class A sortase enzymes in their apo form or substrate bound form, and suggest that ^{Ba}SrtA have evolved to use residues outside of the catalytic domain to modulate catalysis.

4.3.3 Positioning of the active site loops

In the previously reported NMR structure of the apo form of the enzyme, the substrate contacting β6/β7 loop adopted an ordered structure, whereas residues in the β7/β8 loop that bridges the active site cysteine and arginine were poorly defined by the NMR data. In the complex, the structure and dynamics of the rigid β6/β7 loop is largely unaffected by substrate binding as the position of the loop remained unchanged and well-defined, and dynamics data suggest it remains immobile (see below). In contrast, significant changes in the structure and dynamics of the β7/β8 loop are observed. In the Ba-$SrtA_{\Delta 56}$-LPAT* structure, the β7/β8 loop is structured as suggested by the tight bundle in the 20 structures ensemble (Fig. 4.2*A*). The shape of the loop was defined by 15 long range NOEs, 8 intraloop NOEs, and 14 dihedral angle

restraints predicted from the backbone atom chemical shifts. The β7/β8 loop does not interact with the sorting signal peptide as there was no intermolecular NOE to support their interaction, but the N-terminal side of the β7/β8 loop contacts the beginning of the β4/β5 loop as indicated by several long range NOEs. The two loops are separated by a Cα-Cα distance of ~5.5 Å, and most of the contacts are made between Val^{190} within the β7/β8 loop, and Met^{128} and Ser^{129} within the β4/β5 loop. The restructuring of the β7/β8 loop might have a role in binding the secondary substrate lipid II, and will be discussed in section 4.4.4.

4.3.4 Model of the thioacyl intermediate

To gain insight into how the *B. anthracis* sortase anchors proteins to the cell wall, we modeled a thioacyl intermediate of catalysis. The model was constructed using the structure of the ^{Ba}SrtA-LPAT* complex by replacing the coordinates of methylene group and sulfur atom in the modified threonine in the bound peptide with a carbonyl group, and by replacing the peptide-enzyme disulfide linkage with a thioacyl bond. The new coordinates were then energy minimized (see section 4.6.4 for details). Because the orientation of the thioacyl bond is critical for defining the proper mechanism of catalysis, two starting models were energy minimized in which this group either pointed towards or away from the active site arginine residue (Arg^{196}). After energy minimization, both starting models converged to a similar final structure in which the carbonyl oxygen of the thioacyl bond points pointed towards Arg^{196}. The role of active site arginine residue in catalysis has been controversial. In the model of the thioacyl intermediate, the side chain of Arg^{196} makes extensive contacts with the LPAT peptide (Fig. 4.3). Its side chain is wedged into a pocket enclosed by the L-shaped backbone of the peptide, with its guanidino

group forming hydrogen bonds with the backbone carbonyl oxygen of proline and the side chain γ-hydroxyl group of threonine. The cationic guanidino group is positioned near the carbonyl oxygen within the thioacyl bond. It is therefore poised to stabilize the oxyanion presented in tetrahedral intermediates that form before or after the thioacyl intermediate, since we estimated that the position of the oxygen atom should not change by more than 0.5 Å during the transition between sp^2 and sp^3 hybridization of the carbonyl carbon atom. In the ensemble of NMR structures of the ^{Ba}SrtA-LPAT* complex, the side chain of Arg^{196} adopts multiple conformations and in some conformers its side chain points away from the bound peptide. Importantly, the starting conformation of the arginine side chain prior to energy minimization does not affect its positioning in the model of the thioacyl intermediate, since when one of these structures was used to generate the thioacyl intermediate model, the side chain in the energy refined model is also poised to hydrogen bond with the oxygen within the thioacyl bond and by extension the oxyanion of the tetrahedral intermediates.

4.3.5 ^{15}N Relaxation measurements: the substrate contacting N-terminal appendage transiently detaches from the enzyme

To investigate backbone motions within ^{Ba}SrtA, we measured $\{^1H\}$-^{15}N NOEs, spin-lattice (R_1) and spin-spin (R_2) relaxation rates of the protein backbone nitrogen atoms, and used the Model-free formalism to interpret the data (52,53). Of the 151 non-proline amino acids in the protein, complete quantifiable relaxation information (NOE, R_1, R_2) can be obtained from 119 residues, 98 of which can be analyzed using the Model-free approach. This analysis yields three parameters that describe the motion of the protein backbone: the general order parameter S^2, the

effective correlation time for internal motion τ_e, and R_{ex}. S^2 reports the backbone amide's mobility on the picosecond time scale and is characterized by the internal correlation time τ_e. The value of S^2 ranges from 0 to 1, with a value of 1 indicating that the amide bond is completely immobilized. R_{ex} is the chemical exchange contribution to R_2 and is diagnostic for the presence of slower micro- to millisecond time scale motions. The complete Model-free analysis results can be found in supplemental materials (Table 4.4).

The dynamics data are compatible with the N-terminal extension exhibiting elevated mobility on the picosecond time scale as a result of detaching from the enzyme. The backbone of the catalytic core of ^{Ba}SrtA (formed by residues in strands β1 to β8) is generally rigid on the fast picosecond time scale as evidenced by average S^2 values of 0.93. In contrast, residues in the N-terminal extension are more mobile and its residues have significantly lower S^2 values (less than 0.51) (Fig. 4.4*A*). Since numerous NOEs are observed between the appendage and the catalytic core of the protein, this suggests that the extension adopts two conformations: a more populated ordered conformation observed in the NMR structure in which the residues in the extension contact the body of the protein and the sorting signal peptide, and a less populated disordered state in which the extension is detached from the body of the protein. The existence of attached and detached states for the appendage is also compatible with our inability to fit the relaxation data of residues Ala^{70}-Val^{79} in the appendage, which reside at the junction point with the catalytic domain and may undergo more complex hinge motions that are too complicated to be adequately fitted using the Model-free formalism (55). Moreover, it is consistent with the fact that several residues within the catalytic domain that reside at the surface which contacts the appendage exhibit resonance line-broadening that precluded their chemical shift assignment. The resonances of these residues may be broadened by the transient detachment of the N-terminal

appendage and include: Ser^{97} within the β2/H2 loop, His^{126} and Asn^{127} in the β4/β5 loop, Cys^{187} in the end of strand β7, and residues Val^{188} Asn^{193} and Lys^{195} in the β7/β8 loop (Fig. 4.4*B*).

The substrate binding site in ^{Ba}SrtA adopts a generally rigid structure. In particular, both the aforementioned β7/β8 loop and the substrate contacting β6/β7 loop are rigid on the picosecond time scale according to their S^2 data (S^2 values for residues in these regions are > 0.87). Overall, few residues exhibit significant R_{ex} values when the data is interpreted using the Model-free approach suggesting an absence of significant motions in the active site when the substrate is bound. Significant R_{ex} values are observed for backbone amide atoms of Asp^{169} and Glu^{172} within the β6/β7 loop, but they likely originate from motions in the proximal Boc N-teminal protecting group of the bound peptide that protrudes into the solvent and is not recognized by ^{Ba}SrtA (Fig. 4.4*B*). It is also notable that the amides of His^{126}, Asn^{127}, Cys^{187} and Val^{188} are broadened beyond detection. These residues reside within the active site and may be broadened by changes in the tautomerization state of the proximal His^{126} side chain. Because the Model-free analysis only indirectly identifies R_{ex} values that arise from slower motions that might be expected to accompany detaching of the appendage, we attempted to directly measure R_{ex} by performing constant-relaxation time Carr-Purcell-Meiboom-Gill (CPMG) experiments (59), which are suitable for probing motions in the 0.1-10 ms time window (60-62). However, the ^{15}N-CPMG relaxation dispersion curves are flat for the residues in the N-terminal extension, indicating that latching motions associated with the N-terminal extension occur at times faster than 100 μs (62). Interestingly, a comparison of relaxation data with the previously reported values of the apo-form enzyme indicates that the substrate binding acts to destabilize the N-terminal tail (described in the section 4.4.1).

4.3.6 Enzyme kinetics: the N-terminal extension inhibits substrate hydrolysis

To understand the role of the N-terminal extension in enzyme catalysis, a FRET-based assay was used to monitor the ability of ^{Ba}SrtA to hydrolyze a self-quenched fluorescent peptide substrate Abz-LPETG-Dap(Dnp)-NH_2 as previously described. In this assay, a water molecule acts as a nucleophile to resolve the Abz-LPET:enzyme complex instead of the canonical substrate lipid II. This assay thus monitors the initial acylation steps of catalysis that occur during transpeptation. The kinetics of hydrolysis were determined for the wild type enzyme (^{BA}SrtA) and a series of mutants that probed the importance of N-terminal extension, which include S59G and I61A mutants that alter interactions between the extension and the substrate, a deletion mutant that eliminates the first eight amino acids of the extension ($^{Ba}SrtA_{\Delta 64}$, which removes Asp^{57}-Pro^{64} from ^{Ba}SrtA) and a C187A mutant that eliminates the active site cysteine residue. Fig. 4.5 shows representative data fitted to Michaelis-Menten curves, and the fitted kinetic parameters are summarized in Table 4.2.

The results of the kinetics studies suggest that presence of the N-terminal extension inhibits hydrolysis of the substrate. Wild-type ^{Ba}SrtA cleaves the peptide with k_{cat} and K_m values of $3.6 \times 10^{-5} \pm 2 \times 10^{-6}$ s^{-1} and 306 ± 23 μM, respectively. When the entire N-terminal extension that covers the peptide is removed ($^{Ba}SrtA_{\Delta 64}$), a ~60% increase in k_{cat} and ~43% decrease in K_m is observed. This suggests that the extension inhibits the hydrolysis reaction, as well as prevents the peptide substrate from accessing the binding site. Interestingly, single residue mutations (S59G and I61A) did not lead to a significantly lower K_m, so it seems that all of the residues within the extension collectively interfere with peptide binding. The improvement in signal peptide binding after removal of the extension suggests that even though it contacts the bound

peptide in the structure of the complex, the extension does not play a significant role in substrate recognition. This is consistent with our finding based on NMR relaxation studies that substrate binding destabilizes the N-terminal extension, since the extension-substrate interaction is likely transient and not contributing much to substrate recognition. Contrary to K_m, we noticed that mutating Ile^{61} to Ala was sufficient for increasing k_{cat} to the same level as removing the first eight amino acids of the extension. This is presumably because in addition to contacting the sorting signal peptide, Ile^{61} forms extensive contacts with the active site His^{126} which might allow it to directly interfere with catalysis. Unfortunately, we were not able to study the role of the N-terminal extension in transpeptidation reaction by performing FRET assay using Abz-LPETG-Dnp(Dap)NH_2 and up to a 100-fold molar excess of mimics of lipid II (*m*-DAP, L-Ala-D-iGln-*m*-DAP or GlcNAc-MurNAc-L-Ala-D-iGln-m-DAP). It is likely that the intact lipid II or additional components of the cell wall is needed for Ba-SrtA to complete the transpeptidation reaction.

4.4 Discussion

Sortase enzymes are promising drug targets because they are used by Gram-positive bacteria to covalently attach protein virulence factors to the cell surface (6,63). Our understanding of the mechanisms of catalysis and substrate recognition is limited as structures of only two members of this enzyme superfamily have been determined when they are bound to their substrates. The first structure of the sortase-sorting signal complex was reported in 2004, which was determined with the sorting signal peptide LPETG soaked into the crystal of apo *S. aureus* $SrtA_{\Delta 59}$ containing an active site cysteine to alanine mutation (64). However, we believe

the peptide is bound nonspecifically as its interactions with the enzyme are incompatible with sequence conservation and mutagenesis results (20). In 2009, we published the solution structure of *S. aureus* $SrtA_{\Delta 59}$ covalently attached to a novel sulfhydryl containing analog of the sorting signal (Cbz-LPAT*) (20). The structure of ^{Sa}SrtA-LPAT* is compatible with biochemical results, and revealed that the enzyme binds to the substrate using an induced fit mechanism. Most recently, we reported the structure and molecular dynamics of *S. aureus* $SrtB_{\Delta 30}$ in complex with a Cbz-NPQT* revealing the mode of recognition by class B enzymes and the presence of a substrate-stabilized oxyanion hole that we proposed is used by all sortase enzymes to stabilize the tetrahedral catalytic intermediates of catalysis (24). Here we report the structure of SrtA enzyme from *B. anthracis* bound to its sorting signal, the ^{Ba}SrtA-LPAT* complex. ^{Ba}SrtA is a potential drug target as its activity has been shown to be required for *B. anthracis* to escape from macrophages, which is a key step in the progress of lethal anthrax disease (13). It is the first structure of a sortase-sorting signal complex determined for a non-*S. aureus* sortase enzyme and shares limited primary sequence homology to ^{Sa}SrtA (29% identity) or ^{Sa}SrtB (18% identity). Unlike these enzymes, it contains a structured N-terminal extension that contacts the active site and bound substrate, as well as a unique disordered active site loop that becomes ordered upon signal substrate binding.

4.4.1 Comparing the apo and substrate bound structures of Ba-SrtA

The active site of Ba-SrtA undergoes various changes in its structure and dynamics as a result of signal binding. The main surface on the enzyme that contacts the signal forms a rigid preformed binding site for the signal and only undergoes limited changes upon substrate binding,

whereas the β7/β8 loop becomes more structurally ordered when the peptide binds even though it does not directly interact with the peptide. Overall, the structures of apo ^{Ba}SrtA and ^{Ba}SrtA-LPAT* overlay almost identically, with the backbone atoms in the structured regions of the apo enzyme (Ile^{61}-Thr^{134}, Arg^{196}-Lys^{210}) aligned to the same regions in ^{Ba}SrtA-LPAT* with an r.m.s. deviation of 0.73 Å (Fig. 4.6*A,B*). The rigid binding surface is formed by residues in the β6/β7 loop (Val^{166}-Val^{174}), strand β4 (Ala^{124}), strand β7 (Ile^{185}), β3/β4 loop (Val^{110}), β2/H2 loop (Ser^{98}) and helix H2 (Leu^{103}). The backbone coordinates of these residues are very similar between the apo and substrate bound enzyme, with an r.m.s. deviation of 0.76 Å. The similarity in structure is supported by the very similar backbone amide resonances for most of the residues, as the majority of their crosspeaks in the ^{1}H-^{15}N HSQC spectra can be overlaid on top of each other (Fig. 4.6*C*). A histogram plot of the chemical shift differences in the presence and absence of the signal substrate analog reveals that the biggest chemical shift changes occur in residues within the β6/β7 loop, Ile^{185} in strand β7 and the active site Arg^{196}, followed by the residues in the N-terminal extension, helix H2, strand β4 and the beginning of the β4/β5 loop (Fig. 4.6*D*). All of these regions contact the sorting substrate analog as mentioned in Results, which explains their significant changes in chemical shifts.

Substrate binding causes residues in the β7/β8 loop to become more conformationally ordered (Fig. 4.6*A*). In our previous study of the apo-form of the enzyme, the backbone amide cross peaks for nearly all residues in the loop exhibited substantial resonance line-broadening in the NMR spectra of ^{Ba}SrtA, precluding their chemical shift assignment (only the backbone amide of Lys^{196} was assignable). Their resonance line broadening indicates that they undergo motions on the micro to millisecond time scales, and resulted in the coordinates being poorly defined in the NMR structure of the apo-enzyme. In contrast, in the NMR spectra of the sorting signal

bound enzyme the NMR resonances of the loop residues are observable and thus it can be concluded that peptide binding acts to rigidify the loop. Unlike the apo-form of the enzyme, chemical shift assignments were obtainable for the majority of the loop residues in the complex (only the backbone amides of Val^{188}, Asn^{193} and Lys^{195} could not be assigned), and numerous NOEs were identified to define its structure. To confirm that the β7/β8 loop is disordered in the absence of the sorting signal analog, we reexamined the NOESY spectra of apo ^{Ba}SrtA, and found that the long range or intraloop NOE crosspeaks either do not exist or are at least 6-fold less intense than those in the NOESY spectra of ^{Ba}SrtA-LPAT*. Interestingly, the substrate-induced ordering of the loop is not caused by direct interactions with the bound peptide, but instead appears to be caused by peptide-induced displacement of the β7/β8 loop that occurs as a result of disulfide bond formation with the active site Cys^{187} residue, which is located at the end of strand β7 leading into the β7/β8 loop and displaced by 7 Å relative to its positioning in the apo-form of the enzyme. This displacement appears to facilitate interaction between the β7/β8 and β4/β5 loops, which pack against each other only in the complexed form of the enzyme via contacts between Val^{190} within the β7/β8 loop, and Met^{128} and Ser^{129} within the β4/β5 loop. Substrate-induced restructuring of the β7/β8 loop has been observed in ^{Sa}SrtA and may have a role in binding the secondary substrate lipid II, since it creates a new surface between the β7/β8 loop and β4/H3 loop that leads into the active site.

The data suggest that the unique N-terminal extension in BaSrtA unlatches from the active site more frequently when the substrate analog is bound, possibly to allow lipid II to gain access to the active site to complete the transpeptidation reaction, as we will show below that the latched N-terminal extension covers the potential lipid II binding site. Comparison of the S^2 data of ^{Ba}SrtA-LPAT* to apo ^{Ba}SrtA shows that the protein has become more ordered overall with the

exception of the N-terminal extension (Fig. 4.6*E*). Although the beginning of the N-terminal extension bundles more tightly in our ^{Ba}SrtA-LPAT* structure ensemble compare to the apo ^{Ba}SrtA structure ensemble, we think it is because we were able to include intermolecular distance restraints obtained from the three-dimensional (F1) ^{13}C,^{15}N-filtered (F2) ^{13}C-edited NOESY-HSQC experiment, whereas no long range intramolecular NOEs could be unambiguously assigned in ^{Ba}SrtA-LPAT* or apo ^{Ba}SrtA from the three-dimensional ^{15}N- and ^{13}C-edited NOESY spectra due to severe signal overlap. The lower S^2 values of the residues in the N-terminal extension in the presence of the substrate analog is supported by a 3-fold drop in intensity of the long range NOE crosspeaks between residues in the N-terminal extension and the rest of the protein, which suggests that there is less population of the N-terminal extension latched onto the protein when the sorting signal peptide is bound. Lastly, analysis of the random coil index (RCI), a method to measure protein flexibility by comparing backbone chemical shifts to reference random coil shifts (65,66), shows that the N-terminal extension has significantly higher RCI when the LPAT* substrate analog is bound (Fig. 4.6*F*), especially at the regions flanking helix H1. This suggests that when LPAT* binds to the enzyme, a significant portion of the N-terminal extension detaches from the core of the protein and, with the exception of helix H1, adopt a random coil conformation.

4.4.2 The regulatory role of N-terminal extension resembles the function of the "lid" in class C sortases

Since Sa-SrtA does not contain a structured N-terminal extension, we are interested in understanding the role of the N-terminal extension in catalysis. We found that without the

peptide substrate, the extension, specifically Ile^{61}, occupies the active site (25). However, when the substrate analog LPAT* is present, a significant population of the N-terminal extension is unlatched from the core of the protein, although the majority of it remains latch on to the core as evidence by intra- and intermolecular NOEs. Interestingly, mutating Ile^{61} within the N-terminal extension or removing the first eight residues from the N-terminus led to larger catalytic turnover and tighter binding of the LPETG peptide substrate, ultimately leading to higher hydrolysis efficiency. This suggests that the N-terminal extension reduces substrate hydrolysis rate by physically occluding the binding site, and by interacting with the active site His^{126}. This feature is similar to the N-terminal extension found in the pili-anchoring class C sortases (SrtC), which acts as a lid to block access of active site, but with important differences: the lid contains a conserved DP(W/Y/F) motif that makes specific contacts with the enzyme (salt bridge between Asp and active site Arg, aromatic-sulfur interaction between Trp/Tyr/Phe and active site Cys), and the N-terminal extension on SrtC approaches the active site from the groove between helix H2, β3/β4 loop and β6/β7 loop instead of between helix H2 and β4/H3 loop (67). Moreover, the lid inhibition effect is stronger, with a 2- to 7-fold increase in k_{cat} in hydrolysis reaction after truncation of the lid-containing N-terminal extension (68,69). Unfortunately, without a functional transpeptidation assay, we cannot determine whether the catalysis inhibition effect extends to transpeptidation. It is possible that the N-terminal extension in Ba-SrtA only retards catalysis when lipid II is absent to regulate futile hydrolysis reaction.

4.4.3 Comparing ^{Ba}SrtA-LPAT* with other sortase enzyme-substrate complexes

In all of the sortase enzyme-substrate analog complex structures available to date, the signal substrate analogs adopt very similar conformations in backbone but with important differences in side chain positioning (Fig. 4.7*A-D*). To facilitate a discussion of the molecular basis of substrate recognition, we shall use the nomenclature developed by Schechter and Berger, where P and P' refer to amino acids on the N- and C-terminal side of the scissile peptide bond of the sorting signal, respectively (70). For the LPATG sorting signal, the N-terminal leucine is P4, proline is P3, alanine is P2, threonine is P1, and the first amino acid after the scissile bond, glycine, is P1'. As shown in Fig. 4.7*D* and Table 4.3, the three signal substrate analogs can be superimposed at positions P4 and P3 very well because they have very similar dihedral angles, but they do not overlay very well at positions P2 and P1. Residue at P4 position is primarily recognized by residues within the β6/β7 loop. For SrtA substrate in which P4 is a leucine, the side chain of leucine is inserted deep into the hydrophobic pocket created by the β6/β7 loop, strand β6 and strand β8 (Fig. 4.7*A,B*). In contrast, for SrtB substrate in which P4 is an asparagine, the side chain of asparagine is not inserted into the hydrophobic pocket because of its polar amide group. Instead, it forms a hydrogen bond with the backbone carbonyl of Thr^{177} within the β6/β7 loop. The P3 position is occupied by a proline in *trans* conformation in all three structures. It produces a ~90° kink to redirect the trajectory of P2 and P1 to the active site cysteine. Since proline is present at the P3 position in ~90% of all predicted sortase sorting signal substrates (71), it is likely that most sorting signals adopt very similar L-shaped conformations when bound to sortase. The side chain of the P2 residue orients differently when comparing the three complexes. Since many different types of amino acid can be found at the P2 position of sortase substrates (71), it is conceivable that the side chain of P2 projects to the

solution instead of being recognized by specific residues in the enzyme, such that many types of side chain can be accommodated without causing steric clashes. This is true in ^{Ba}SrtA-LPAT* and ^{Sa}SrtB-NPQT*, where the P2 position is occupied by an alanine and glutamine with a negative φ angle at -40° and -72°, respectively (Table 4.3). Interestingly, the P2 alanine in ^{Sa}SrtA-LPAT* has a positive φ angle at 49°, which orients the side chain toward strands β4 and β7 of the enzyme. This suggests that when P2 is occupied by an amino acid with a shorter side chain, its side chain can point into the substrate binding pocket without causing significant destabilization of the complex. The side chain of threonine found at the P1 position also orients differently in the three complexes. In general, we can classify the orientation as either "in" or "out," as described in (24). In ^{Sa}SrtA-LPAT*, the side chain of threonine points "out" towards the solvent as indicated by several NOEs between the methyl protons on the modified threonine and the Hε1 proton of Trp^{194} within the β7/β8 loop. In ^{Ba}SrtA-LPAT* and ^{Sa}SrtB-NPQT*, the threonine points "in" towards the floor of the substrate binding site instead. However, there are differences between the two "in" conformations. In ^{Ba}SrtA-LPAT*, the threonine γ-methyl is buried in a pocket enclosed by His^{126} as indicated by several NOEs, and has a pseudo-φ angle at 67°. In ^{Sa}SrtB-NPQT*, the threonine does not interact with the active site His^{130} as His^{130} is blocked by the side chain of Leu^{96}. Instead, the side chain swings towards the β7/β8 loop with a pseudo-φ angle at -112°, which allows the γ-hydroxyl group to form a hydrogen bond with the active site Arg^{233}. Importantly, MD simulations of the thioacyl intermediate suggest that the threonine-"in" conformation is accessible to all three complexes (see Results and (24)), and in all three cases the active site Arg can interact with the γ-hydroxyl group and carbonyl oxygen of the P1 threonine. This supports the hypothesis that Arg serves as an oxyanion hole to stabilize the

high energy tetrahedral catalytic intermediate as we and others have proposed before (17,19,21,24).

4.4.4 Predicting potential lipid II binding site

Although the primary substrate binding site has been well defined in ^{Sa}SrtA, ^{Sa}SrtB and ^{Ba}SrtA, it remains unclear where the lipid II binding site is. It has been suggested in the past that the binding site might be at the β7/β8 loop from a crystal structure of Sa-SrtB with lipid II mimic triglycine soaked in (23), or at a groove located between the β7/β8 loop and β2/β3 loop from a chemical shift perturbation study with triglycine titrated into ^{Sa}SrtA_LPAT* (20). However, both studies were performed with high concentration of triglycine since lipid II is expected to bind weakly to the enzyme (K_m for the pentaglycine binding is 140 μM (72)), so it remains possible that the triglycine is binding non-specifically in the above two cases. Nevertheless, in ^{Ba}SrtA, both predicted binding sites are occluded by the N-terminal extension when it is latched on to the active site (Fig. 4.8*A*). It is therefore not surprising to see an increase in the N-terminal extension mobility when the sorting signal substrate analog is bound to the enzyme, such that lipid II can come into the active site to complete transpeptidation. Depends on how much of the N-terminal extension is unlatched, there seems to be two possible channels for lipid II to come into the active site. The first channel is between the β7/β8 loop and β4/H3 loop when only Asp^{57}-Pro^{64} are unlatched (Fig. 4.8*B*), and the second channel is between the β4/H3 loop and H2 helix when additional residues are unlatched (Fig. 4.8*C*). Both channels lead to His^{126}, which could function as a general base to deprotonate lipid II. However, we think it is more likely that the β7/β8 loop is involved in lipid II binding, since it changes from mobile to rigid upon LPAT* binding. As we

could not perform molecular docking or chemical shift perturbation experiments to locate the lipid II binding site since a functional lipid II mimic for Ba-SrtA has not been identified, we used FTSite to predict the ligand binding pockets within ^{Ba}SrtA-LPAT*. FTSite uses multiple small molecular probes to search for ligand binding site, and is capable of identifying the correct binding site in 94% of apo proteins (73,74). Interestingly, the binding site is predicted to be located between the β7/β8 loop and β4/H3 loop only, even when the entire N-terminal extension is removed (Fig. 4.8*B,C*). This, together with the rigidification of the β7/β8 loop, suggests that lipid II binds to the channel between the β7/β8 loop and β4/H3 loop.

4.5 Conclusion

To summarize, we solved the structure of *B. anthracis* SrtA bound to a substrate analog LPAT*, and discovered that ^{Ba}SrtA binds the substrate with a preformed binding site. However, substrate binding induces two important changes to the protein: 1) the N-terminal extension unlatches from the active site more frequently, and 2) the β7/β8 loop becomes more rigid and contacts the β4/β5 loop. From mutagenesis and enzyme kinetics studies, we found that the N-terminal extension plays a regulatory role in inhibiting substrate entry and catalysis in the absence of the secondary substrate lipid II, which may be a way for the enzyme to regulate the futile hydrolysis reaction. We also found that the ordering of the β7/β8 loop creates a new surface which may become the binding site for lipid II. The new structural information of ^{Ba}SrtA may be combined with the structure of *S. aureus* SrtA to rationally design compounds that would inhibit both enzymes well as a first step to design pan-sortase inhibitors.

4.6 Materials and Methods

4.6.1 Preparation of the covalent complex for NMR studies

Wild-type SrtA protein from *B. anthracis* containing amino acid residues Asp^{57}-Lys^{210} (^{Ba}SrtA) was produced as described previously (25). Uniformly ^{15}N- and ^{13}C-labeled ^{Ba}SrtA protein was covalently attached to an analog of the LP*X*TG sorting signal, Boc-LPAT* (where T* is (2*R*,3*S*)-3-amino-4-mercapto-2-butanol, and Boc is a *tert*-butoxycarbonyl protecting group), by incubating 200 μM of ^{Ba}SrtA with 2 mM Boc-LPAT* in modification buffer (50 mM Tris-HCl, pH 8.0, 100 mM NaCl) for 48 hours at room temperature. MALDI-TOF mass spectrometry was used to confirm the production of stable covalent complex. Two samples of the complex dissolved in NMR buffer (50mM NaPO4, pH 6.0, 0.01% NaN_3) were studied by NMR: 1) 2.6 mM [^{15}N,^{13}C] ^{Ba}SrtA bound to the unlabeled peptide dissolved in NMR buffer containing 8% D_2O; 2) 2.6 mM [^{15}N,^{13}C] ^{Ba}SrtA bound to the unlabeled peptide dissolved in deuterated NMR buffer.

4.6.2 NMR Spectroscopy and Structure Determination

NMR spectra were acquired at 298 K on Bruker Avance 500-, 600-, and 800-MHz spectrometers equipped with triple resonance cryogenic probes. NMR spectra were processed using NMRPipe (26) and analyzed using PIPP (27) and CARA (version 1.8.4) (28). ^{1}H, ^{13}C, and ^{15}N protein chemical shift assignments were obtained by analyzing the following experiments: ^{15}N-HSQC, ^{13}C-HSQC, HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO, HBHA(CO)NH, HNHA, HNHB, CC(CO)NH, H(CC)(CO)NH, HCCH-TOCSY, and HCCH-COSY (29,30). Chemical shift assignments of the unlabeled Boc-LPAT* were obtained by analyzing a two-

dimensional (F1,F2) ^{13}C-filtered NOESY spectrum (31). Distance restraints to define the structure of the protein were obtained from three-dimensional ^{15}N- and ^{13}C-edited NOESY spectra (mixing time at 120 ms), and intermolecular distance restraints between ^{Ba}SrtA and Boc-LPAT* were identified in three-dimensional (F1) ^{13}C,^{15}N-filtered (F2) ^{13}C-edited NOESY-HSQC and (F1) ^{13}C,^{15}N-filtered (F2) ^{15}N-edited NOESY-HSQC spectra (32), and in a two-dimensional (F1) ^{13}C-filtered NOESY spectrum (31). The majority of φ and ψ dihedral angle restraints were obtained using the program TALOS (33). Additional φ angle restraints were obtained by measuring $^{3}J_{HN\alpha}$ values from HNHA spectrum (34), and additional ψ angle restraints were obtained by analyzing three-dimensional ^{15}N-edited NOESY spectrum (35). χ1 angles for dihedral angle restraints and stereotypical chemical shift assignments of β-methylene protons were obtained by analyzing ^{15}N-TOCSY-HSQC, HNHB, HN(CO)HB, and ^{15}N-ROESY spectra (36,37).

NOE assignments were obtained automatically using the ATNOS and CANDID algorithms in UNIO'10 (38,39), and structure calculations were performed using XPLOR-NIH (40). The three-dimensional ^{15}N-edited NOESY-HSQC spectrum of the complex dissolved in H_2O, and three-dimensional ^{13}C-edited NOESY-HSQC spectrum of the complex dissolved in D_2O were used as input for UNIO'10. These data were supplemented with dihedral angle restraints and carbon chemical shifts for secondary structure predictions. Seven cycles of ATNOS/CANDID and XPLOR-NIH calculations yielded a converged ensemble of the protein in the complex. All of the NOE assignments made by CANDID were subsequently verified by manually inspecting the NOESY data. In the structure refinement process, additional intra- and intermolecular distance restraints identified manually, $^{3}J_{HN\alpha}$ couplings and hydrogen bond restraints were included in structure calculations. Hydrogen bonds were identified by inspecting

the NOESY data for characteristic patterns along with deuterium exchange experiments, and restraints for hydrogen bonds were implemented with the HBDB algorithm (41). At the final refinement stage, $^{1}D_{NH}$ and $^{1}D_{NCO}$ residual dipolar couplings were included, which were measured using protein samples partially aligned in PEG C12E5/hexanol (42), using two-dimensional ^{15}N-coupled IPAP ^{1}H-^{15}N HSQC and two-dimensional carbonyl-coupled ^{1}H-^{15}N HSQC experiments, respectively (43,44). Axial and rhombic components of the alignment tensor were fitted using Module (45). In the final set of calculations, a total of 200 structures were generated, of which 94 had no NOE, dihedral angle, or scalar coupling violations greater than 0.5 Å, 5°, or 2 Hz, respectively. Of these, 20 structures with the lowest overall energy were chosen to represent the structure of ^{Ba}SrtA-LPAT* and have been deposited in the protein data bank (PDB code 2RUI). The programs MOLMOL (46) and PyMOL (47) were used to generate figures.

4.6.3 Backbone Dynamics of ^{Ba}SrtA-LPAT* Determined from ^{15}N Relaxation Data

The ^{15}N relaxation data were collected using the ^{15}N- and ^{13}C- labeled sample of the complex dissolved in H_2O on a Bruker Avance 600-MHz NMR spectrometer equipped with a triple resonance cryogenic probe. Data were analyzed using SPARKY (48) and included: 15N longitudinal relaxation rates (R_1), transverse relaxation rates (R_2), and $\{^{1}H\}$-^{15}N heteronuclear NOEs. Complete quantifiable relaxation information (R_1, R_2 and NOE) could be measured for 119 of 151 backbone amides. The average R_1, R_2 and NOE values for the complex are 1.47 ± 0.16 s^{-1}, 13.14 ± 3.97 s^{-1}, and 0.71 ± 0.25 s^{-1}, respectively. The relaxation data were analyzed using the suite of analysis programs kindly provided by Prof. Arthur G. Palmer III, and the

strategy used to analyze the relaxation data has been described previously (49). Briefly, Pdbinertia was used to calculate the principal moments of inertia and yielded relative moments of 1.00:0.91:0.68. R2R1_tm was used to calculate an approximate correlation time of 9.3 ± 0.4 ns for rigid residues using R_2/R_1 ratios. To ensure only rigid residues were used in the correlation time analysis, only R_2/R_1 ratios that met the following criteria were used in the analysis: 1) the residue had a {1H}-15N NOE value > 0.65, and 2) they were within one standard deviation of the average. The tensor parameters were then calculated using Quadric_Diffusion (50,51), which indicated that the isotropic model is statistically preferred for the complex over the axially symmetric or anisotropic models of tumbling. The relaxation data were then interpreted using the Lipari-Szabo Model-free formalism (52,53) using the program FAST-Modelfree (54) to iteratively run the program Modelfree 4.20 (55). Of the 119 amino acids that gave complete quantifiable relaxation information, data from 98 residues could be fit satisfactorily. Forty-five residues were fit to Model 1 (S^2 only), 3 residues were fit to Model 2 (S^2 and τ_e), 29 residues were fit to Model 3 (S^2 and R_{ex}), 5 residues were fit to Model 4 (S^2, τ_e, and R_{ex}), and 16 residues were fit to Model 5 (S^2_f, S^2_s, and τ_e). S^2 parameters could be measured for backbone amides within all regions of the protein, with the exception of residues in the loop that connects helix H1 and strand β1 (residues 70-79), and the latter half of the β7/β8 loop (residues 192-195).

4.6.4 Computational Modeling of the Thioacyl Intermediate

The procedures of generating energy minimized models of the thioacyl intermediate had been described previously (24). Briefly, the methylene group and sulfur of the modified threonine were removed in PyMOL (47), and a carbonyl group was placed in between the Cα of

threonine and the sulfur of Cys^{187}, with the oxygen on the carbonyl pointed towards or away from the Arg^{196}. The rough models of the thioacyl intermediate were solvated in a periodic water box with a solvent distance of 10 Å, and parameterized in tLEAP (56). Restraints were placed on the initial atom positions of heavy atoms within the protein and the LPAT peptide, except for the side chain atoms of Arg^{196}. Restraints were not placed on the initial coordinates of Arg^{196} because the position of its side chain was not well defined in the ensemble due to peak broadening. Models were then energy minimized and equilibrated in NAMD (57) by slowly removing restraints over 1 ns with 2 fs steps. Details about NAMD set up can be found in (24).

4.6.5 Site-directed Mutagenesis and Enzyme Kinetics Measurements

Three single amino acid mutants (S59G, I61A and C187A) of ^{Ba}SrtA were produced for kinetic measurements. Mutations were made using the QuikChange® method (Strategene) and confirmed by DNA sequencing. Mutants, wild-type $^{Ba}SrtA_{\Delta 56}$ and wild-type $^{Ba}SrtA_{\Delta 64}$ were purified as described previously (25). *In vitro* substrate hydrolysis assay was performed as previously described (58). Briefly, the cleavage of a self-quenched fluorescent peptide, Abz-LPETG-Dap(Dnp)-NH_2 (Peptide 2.0, Inc.), was monitored by excitation at 335 nm and recording emission at 420 nm on a Infinite® M1000 PRO spectrofluorometer (Tecan). Assay conditions consisted of 20 mM HEPES, pH 7.5, and 10 μM enzyme. Abz-LPETG-Dap(Dnp)-NH_2 concentrations between 5 and 320 μM were added to the reaction for a total reaction volume of 50 μL. Fluorescence was recorded for 12 h at 10-min intervals. Initial velocities were calculated in terms of relative fluorescence unit per second (RFU/s). Standard curves were used to account for inner filter effects and convert RFU to product concentration (18). K_m and k_{cat} values were

calculated by direct nonlinear fitting three independent sets of data to the Michaelis-Menten equation using SigmaPlot 6.0 (SPSS).

4.7 Figures

Figure 4.1. The Boc-LPAT* modifier and NMR data of its complex with Ba-SrtA. *A*, chemical structure of the Boc-LPAT* peptide analog, where T* is (2R,3S)-3-amino-4-mercapto-2-butanol, and Boc is a tert-butyloxycarbonyl protecting group. *B*, selected panels showing intermolecular NOEs between the Ba-SrtA protein and Boc-LPAT*. The panels are taken from a three-dimensional (F1) ^{13}C, ^{15}N-filtered, (F2) ^{13}C-edited NOESY-HSQC spectrum of the Ba-SrtA-LPAT* complex dissolved in deuterated buffer. The identity of the proton from Ba-SrtA that gives rise to the set of NOEs and its chemical shift are shown at the *top* and *bottom* of each panel, respectively. On the *right* side of each cross-peak the sorting signal peptide proton that is proximal to the protein is indicated.

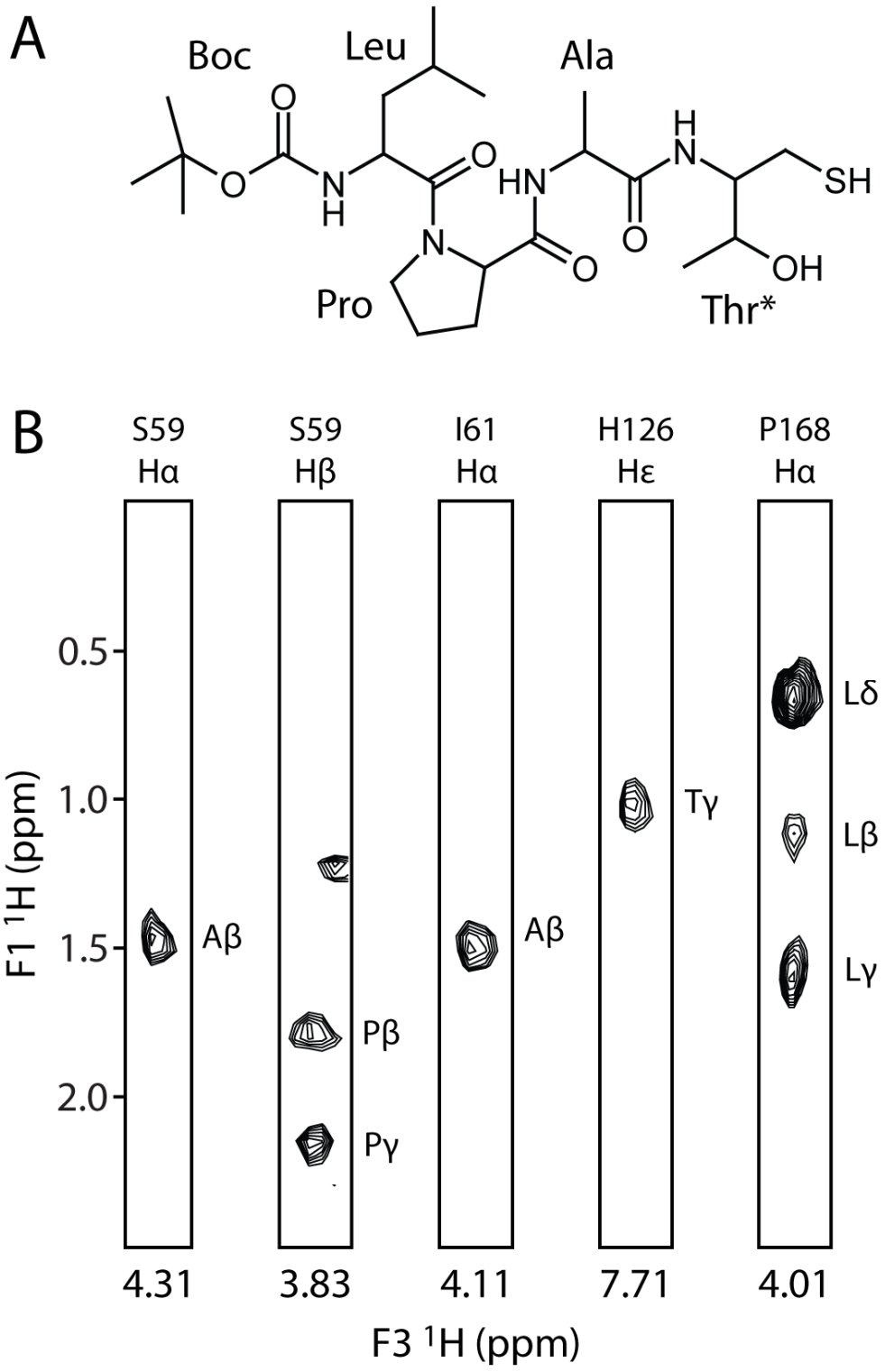

A
Boc
Leu
Ala
Pro
Thr*
SH
OH
B
S59 Hα
S59 Hβ
I61 Hα
H126 Hε
P168 Hα
Aβ
Pβ
Pγ
Aβ
Tγ
Lδ
Lβ
Lγ
0.5
1.0
1.5
2.0
F1 1H (ppm)
4.31
3.83
4.11
7.71
4.01
F3 1H (ppm)

Figure 4.2. NMR solution structure of the Ba-SrtA-LPAT* complex. *A*, cross-eyes stereo image showing the ensemble of 20 lowest energy structures of the Ba-SrtA-LPAT* complex. The protein backbone heavy atoms (blue) and the covalently linked peptide (red) are show. The N-terminal extension is colored green. The coordinates were superimposed by aligning the backbone N, Cα, and C atoms of Asp^{57}-Lys^{210} and Leu, Pro, Ala, Thr* of the Boc-LPAT* peptide. The backbone and heavy atom coordinates of these residues have a root mean square deviation (r.m.s. deviation) to the average structure of 0.42 and 0.89 Å, respectively. *B*, ribbon drawing of the energy minimized average structure of the Ba-SrtA-LPAT* complex. The covalently bound peptide is shown in a red ball-and-stick representation with its amino acids labeled. The N-terminal tail is colored cyan. *C*, an expanded view of the active site showing how the N-terminal tail positions over the sorting signal peptide. The side chains (orange) that participate in the interaction are labeled and shown as sticks. *D*, an expanded view of the active site showing how the sorting signal peptide is recognized. The side chains (orange) that participate in the interaction are labeled and shown as sticks.

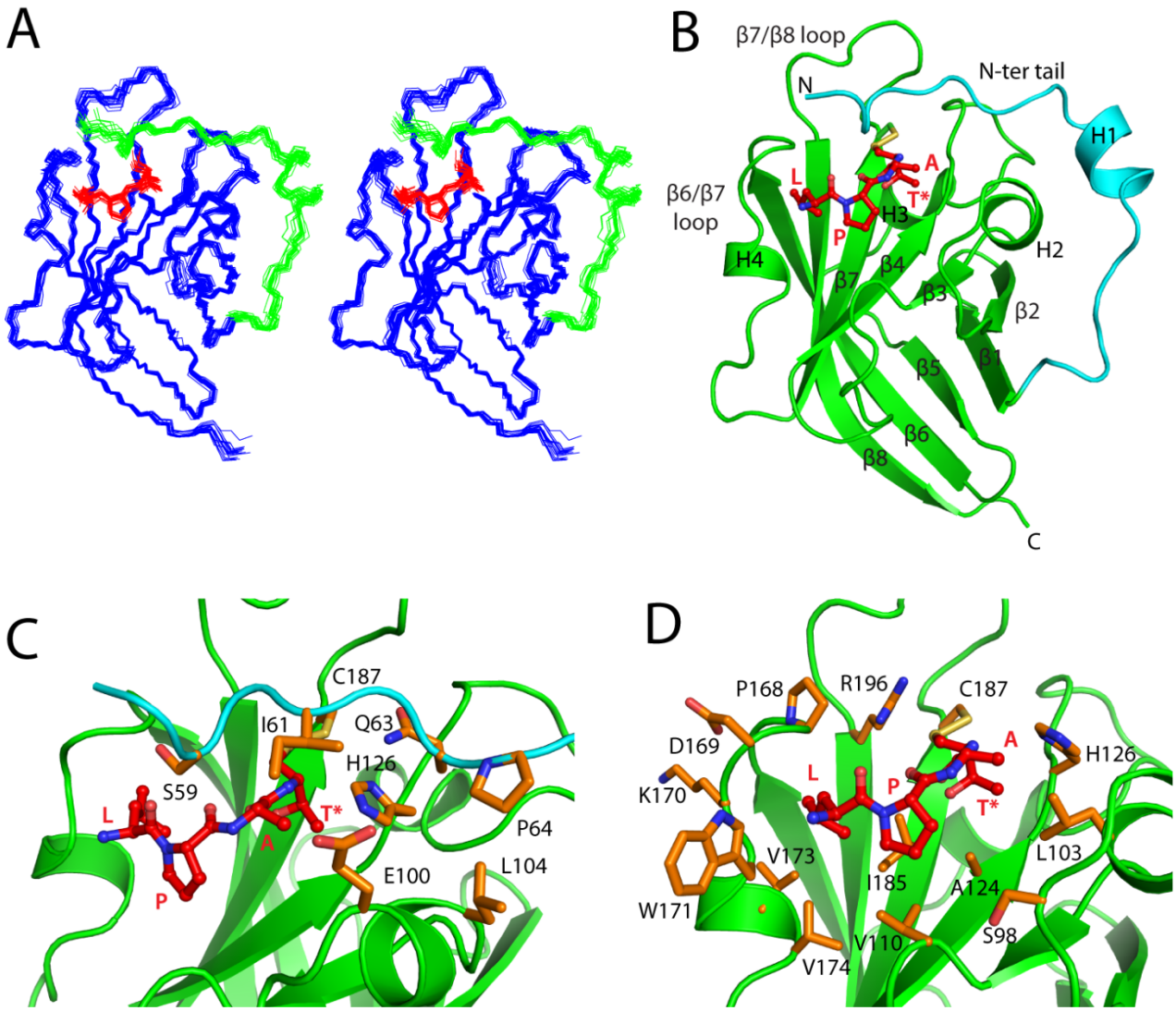

A
B
β7/β8 loop
N
N-ter tail
H1
β6/β7
loop
L
A
T*
H3
P
H4
β4
H2
β7
β3
β2
β1
β5
β6
β8
C
C
C187
I61
Q63
H126
S59
L
T*
A
P64
P
E100
L104
D
P168
R196
C187
D169
A
H126
K170
L
P
T*
V173
L103
I185
A124
W171
S98
V110
V174

Figure 4.3. Model of the thioacyl intermediate. Expanded view of the active site in the energy minimized model of the Ba-SrtA-LPAT thioacyl intermediate. The peptide substrate LPAT (orange) and active site residues Cys^{187} and Arg^{196} (green) are shown as sticks. Hydrogen bonds are indicated by yellow dotted lines. The distance between Thr carbonyl oxygen and Arg^{196} guanidino Cζ is 3.6 Å.

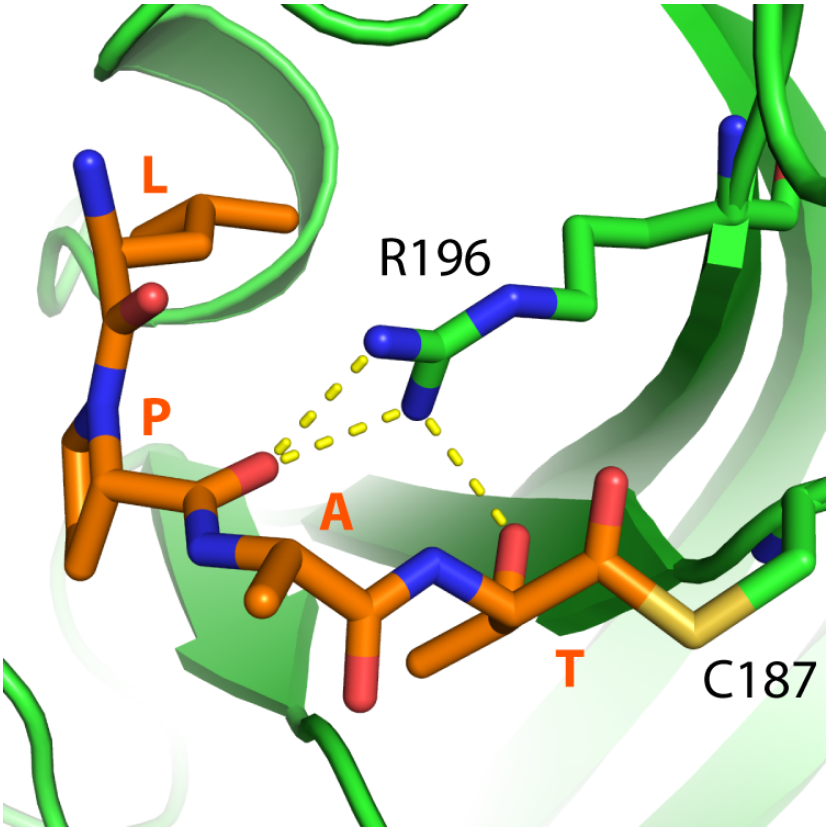
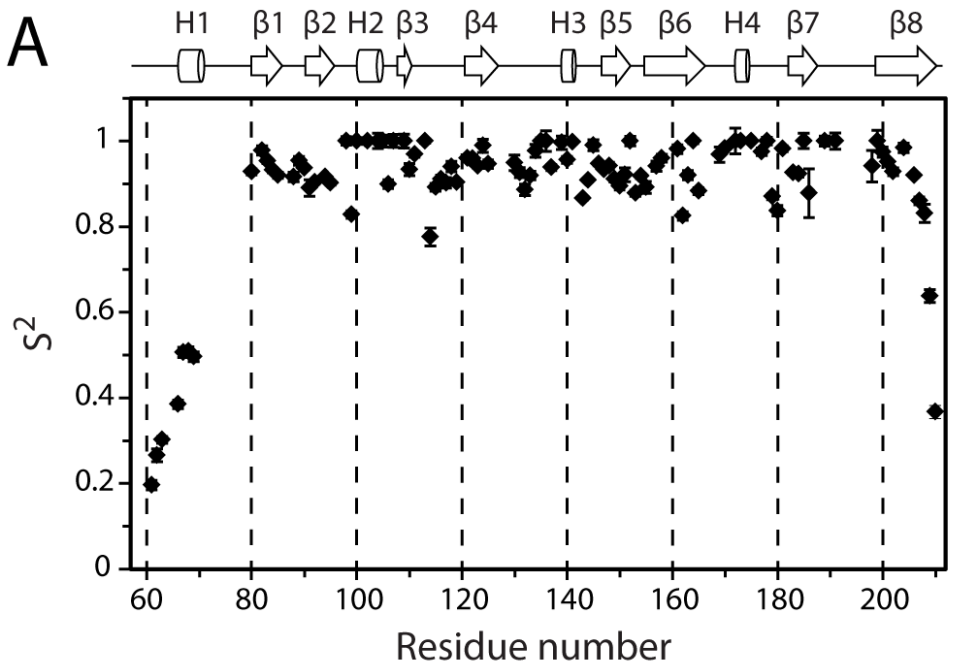


Figure 4.4. Mobility of Ba-SrtA-LPAT* as defined by NMR relaxation data. *A*, scatter plot showing the general order parameter S^2 of the backbone 15N atoms as a function of residue number. The value of S^2 ranges from 0 to 1, with a value of 1 indicating that the amide bond is completely immobilized. The secondary structures of the protein are shown on top of the graph. *B*, ribbon drawing of Ba-SrtA-LPAT* showing the location of the residues in which the backbone amide resonances could not be assigned (black spheres) or have R_{ex} values greater than 10 Hz (gray spheres).


A
H1
β1
β2
H2
β3
β4
H3
β5
β6
H4
β7
β8
1
0.8
0.6
0.4
0.2
0
S2
60
80
100
120
140
160
180
200
Residue number

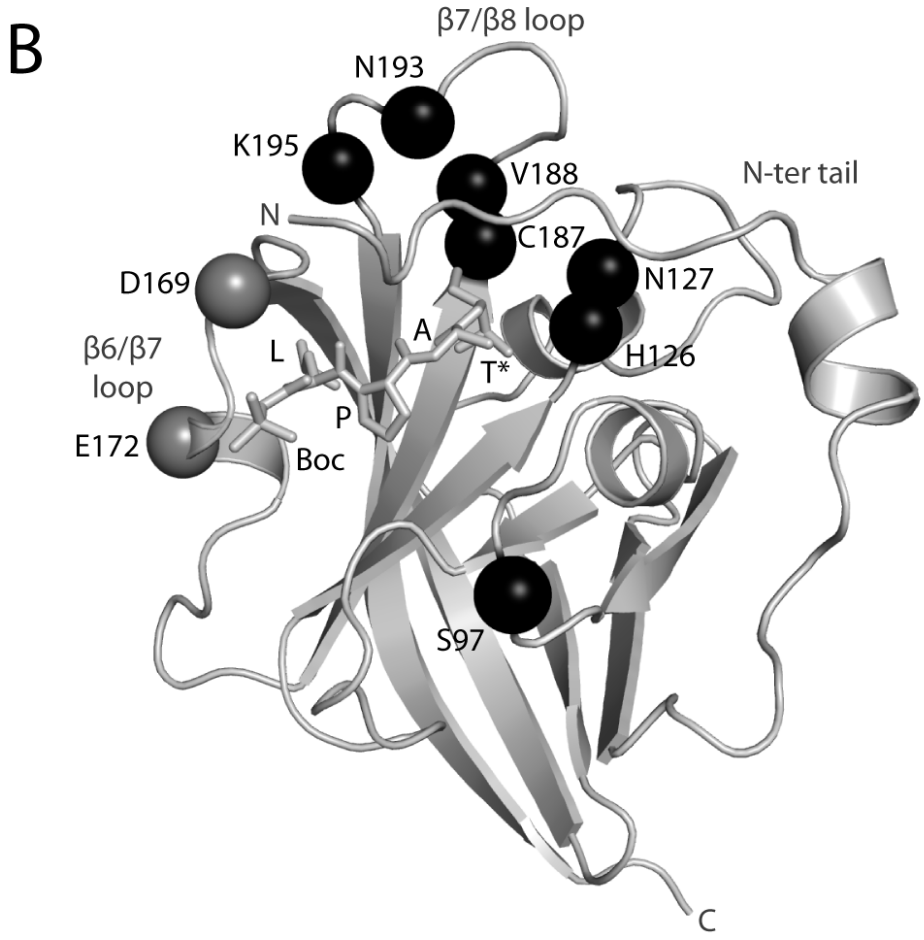

B
β7/β8 loop
N193
K195
V188
N-ter tail
N
C187
D169
N127
A
β6/β7
loop
L
T*
H126
P
E172
Boc
S97
C

Figure 4.5. Kinetic data for N-terminal tail mutants of Ba-SrtA. Representative curves showing the hydrolysis kinetics of S59G (open circles), I61A (filled circles), C187A (open squares) mutants of Ba-SrtA$_{\Delta 56}$, as well as the hydrolysis kinetics of unmutated Ba-SrtA$_{\Delta 56}$ (filled squares) and tail-truncated Ba-SrtA$_{\Delta 64}$ (open diamonds).

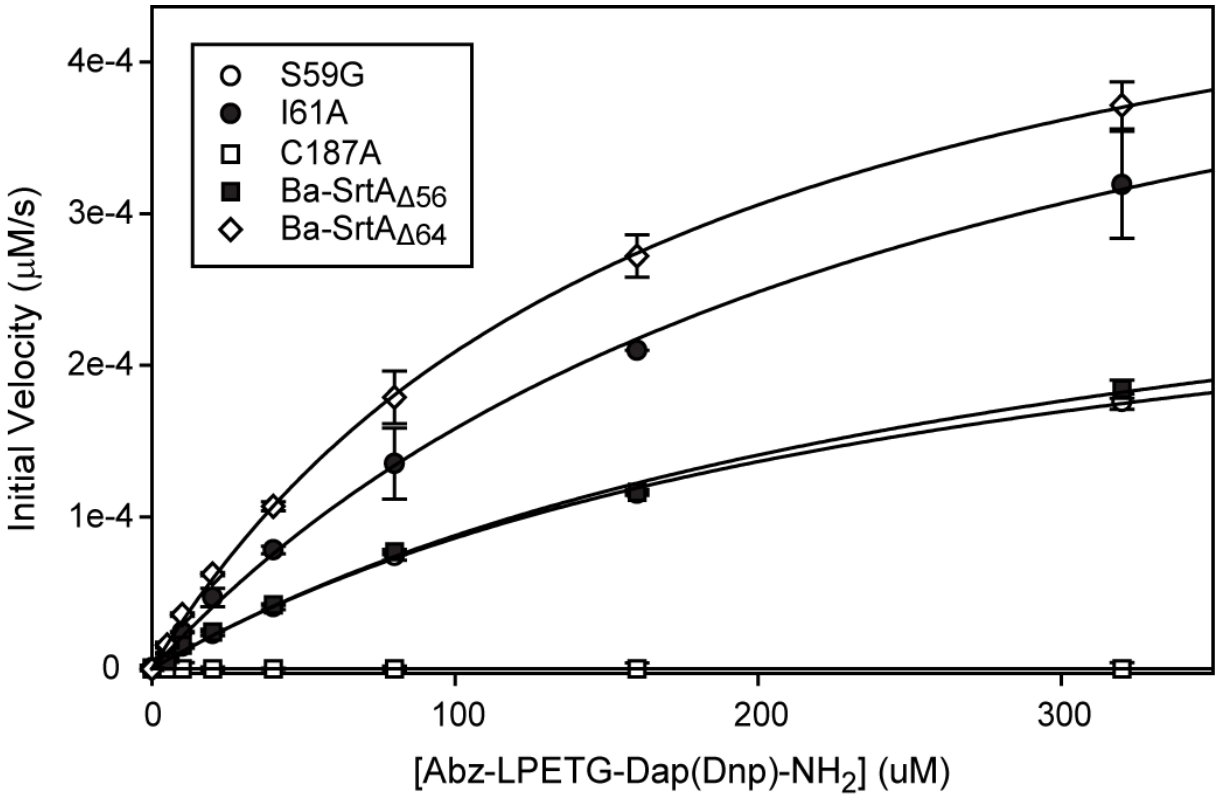
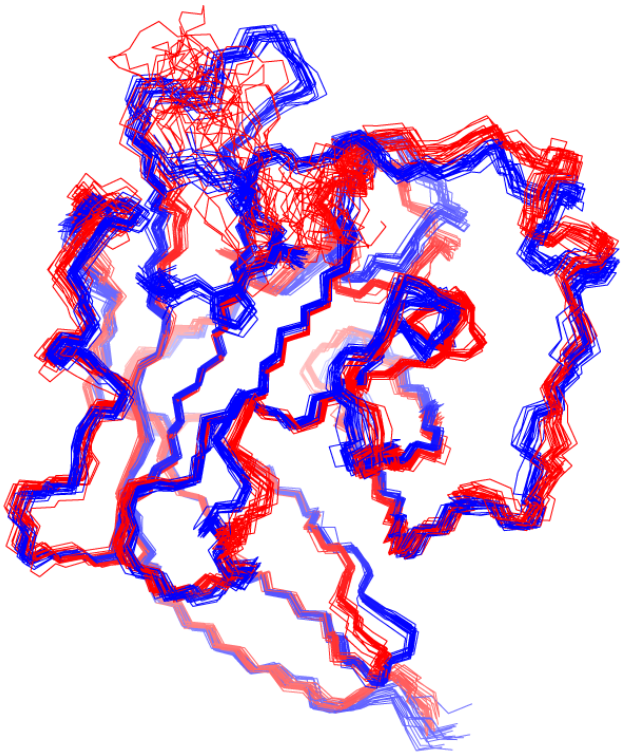
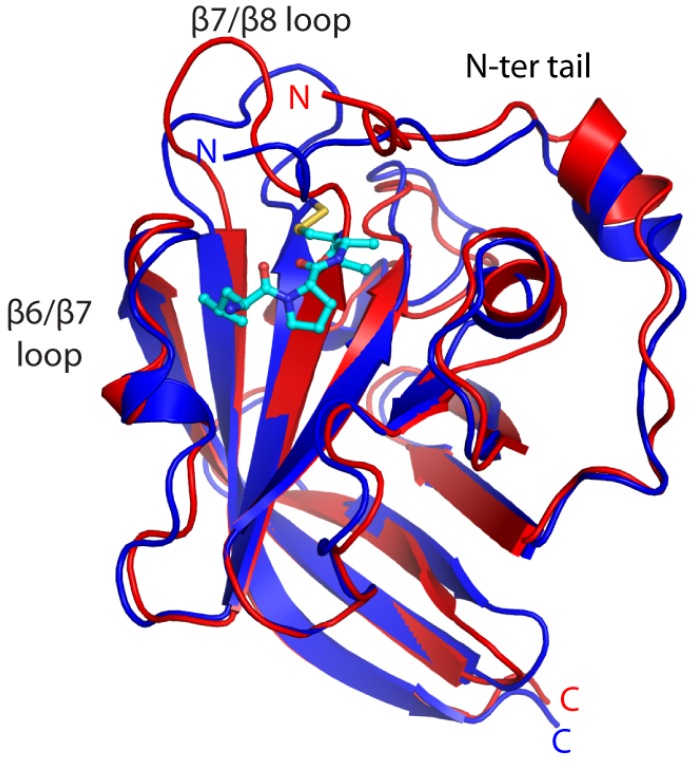


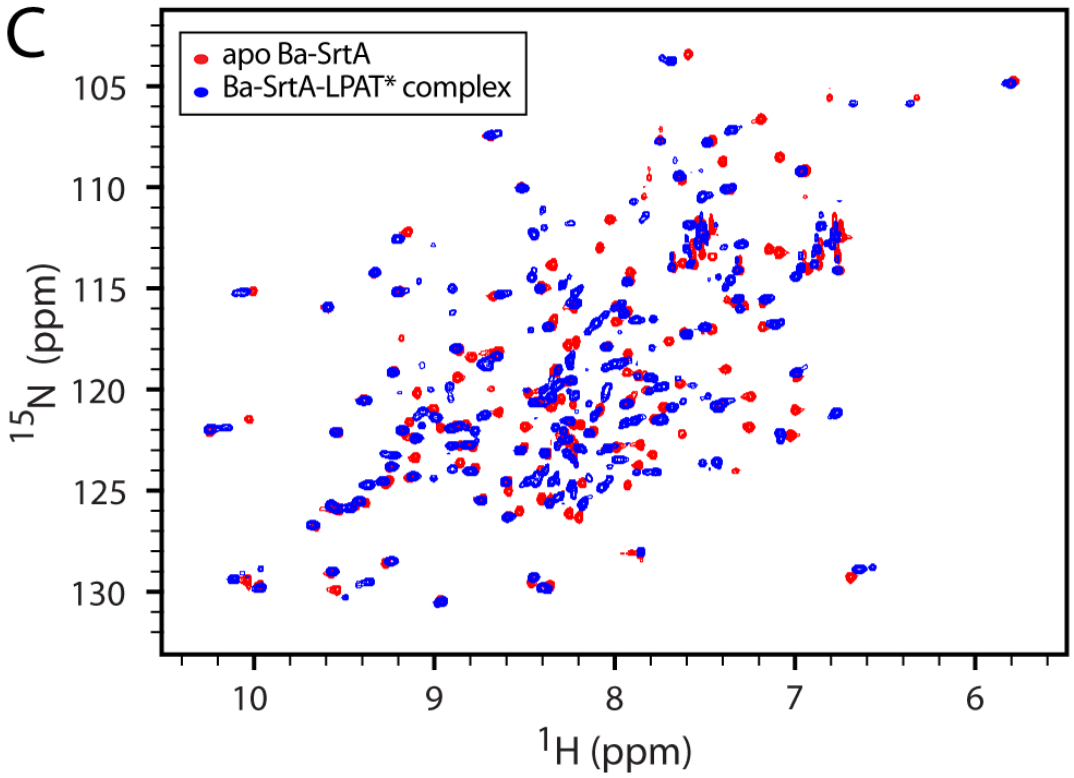
Figure 4.6. Comparison of apo Ba-SrtA and Ba-SrtA-LPAT* complex. *A*, overlay of the ensemble of NMR structures of apo Ba-SrtA (Protein Data Bank code 2KW8; red) and the Ba-SrtA-LPAT* complex (blue). The comparison shows that the structurally disordered β7/β8 loop becomes more ordered upon binding the sorting signal peptide. *B*, superposition of apo Ba-SrtA (red) and the Ba-SrtA-LPAT* complex (blue) in ribbon representation. The backbone atoms of Ile^{61}-Thr^{186}, Arg^{196}-Lys^{210} of the two structures align to a root mean square deviation (r.m.s. deviation) of 0.73 Å. *C*, overlay of the ^{1}H-^{15}N HSQC spectra of apo Ba-SrtA (red) and the Ba-$SrtA_{\Delta 56}$-LPAT* complex (blue). *D*, histogram plot of the compound chemical shift changes (Δδ) for the backbone amide hydrogen and nitrogen atoms of apo Ba-SrtA after the addition of Boc-LPAT*. Chemical shift changes are calculated by the equation $\Delta\delta=((\Delta\delta_H)^2+(\Delta\delta_N/6.49)^2)^{1/2}$, where $\Delta\delta_H$ and $\Delta\delta_N$ are, respectively, the amide proton and nitrogen chemical shift difference for a given residue in the presence and absence of Boc-LPAT*. The dashed line represents one standard deviation above the average Δδ of all amino acids. Amino acids experiencing significant chemical shift changes are labeled, and a schematic of secondary structures of the enzyme is shown above the plot. *E*, scatter plot showing the change in the general order parameter S^2 after the addition of Boc-LPAT*. A positive number means the backbone amide becomes more rigid, while a negative number means it becomes more disordered. A schematic of secondary structures of the enzyme is shown above the plot. *F*, histogram plot of the difference in RCI between Boc-LPAT* bound Ba-SrtA and apo Ba-SrtA. The dashed line represents one standard deviation above the average difference in RCI of all amino acids. A schematic of secondary structures of the enzyme is shown above the plot.

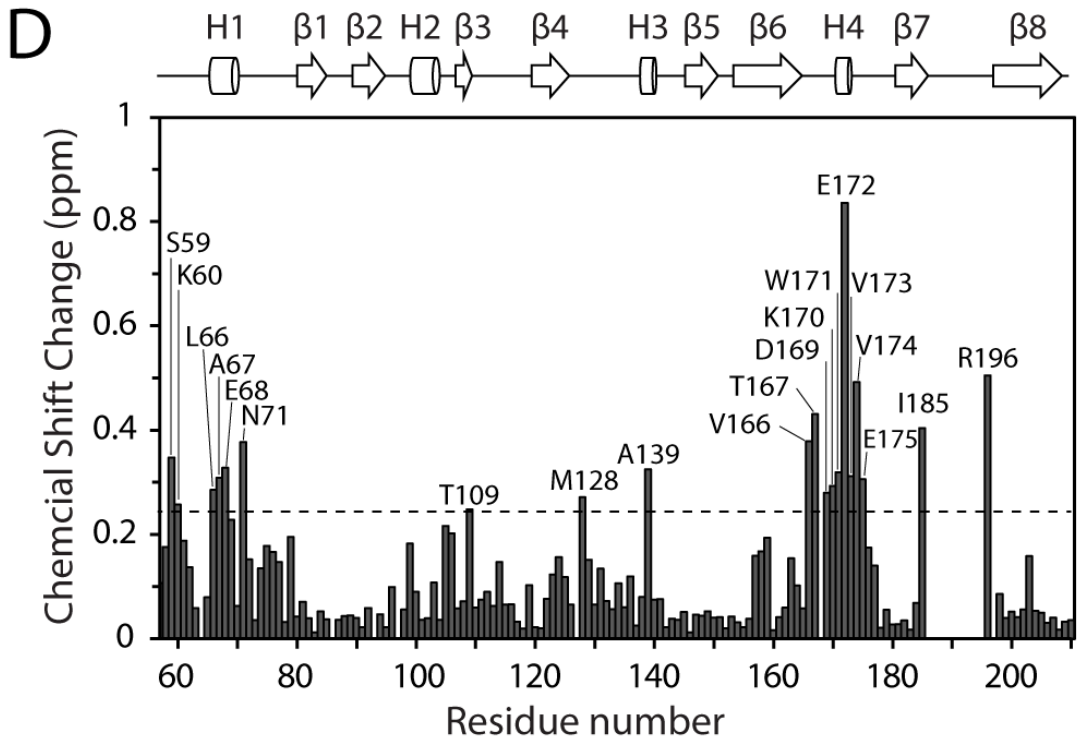
A

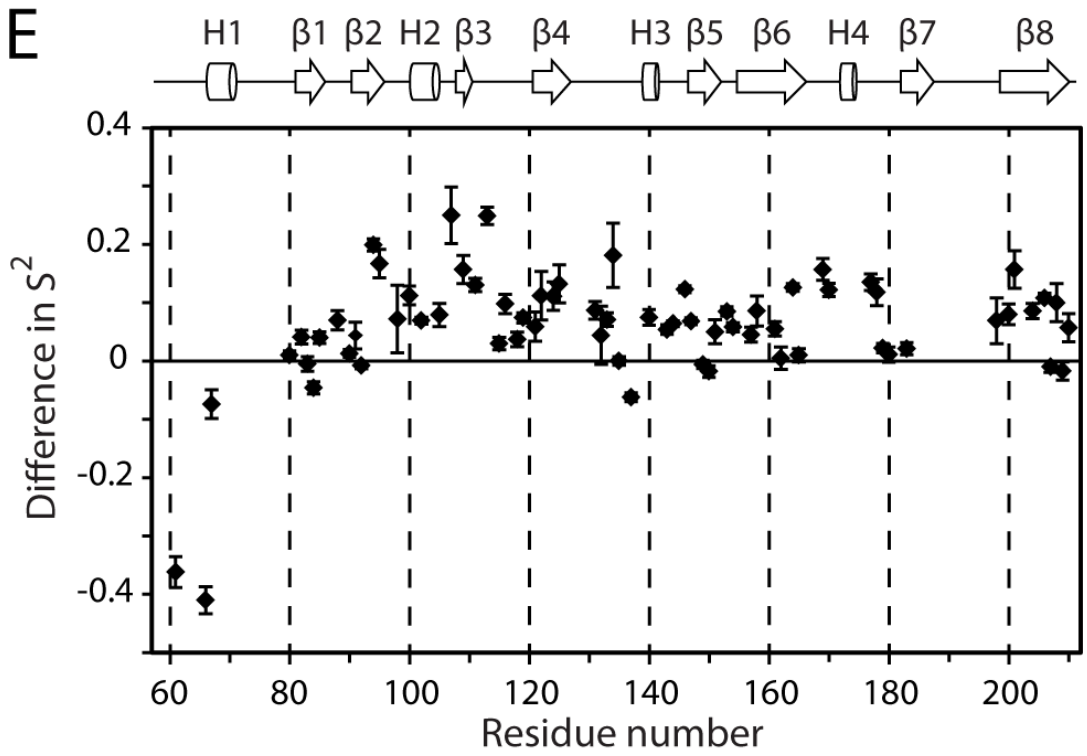


B




C
apo Ba-SrtA
Ba-SrtA-LPAT* complex
105
110
115
120
125
130
^{15}N (ppm)
10
9
8
7
6
^{1}H (ppm)


D
H1
β1
β2
H2
β3
β4
H3
β5
β6
H4
β7
β8
Chemcial Shift Change (ppm)
0
0.2
0.4
0.6
0.8
1
S59
K60
L66
A67
E68
N71
T109
M128
A139
E172
W171
V173
K170
D169
V174
T167
V166
E175
I185
R196
60
80
100
120
140
160
180
200
Residue number


E
H1
β1
β2
H2
β3
β4
H3
β5
β6
H4
β7
β8
0.4
0.2
0
-0.2
-0.4
Difference in S2
60
80
100
120
140
160
180
200
Residue number

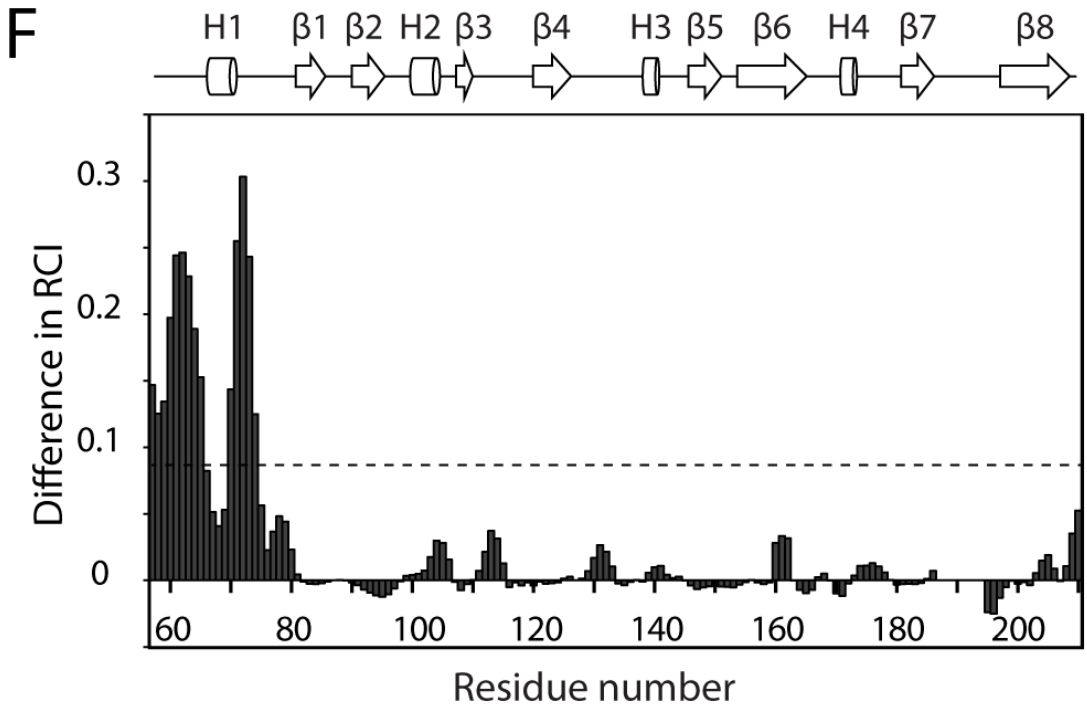

F
H1
β1
β2
H2
β3
β4
H3
β5
β6
H4
β7
β8
0.3
0.2
0.1
0
Difference in RCI
60
80
100
120
140
160
180
200
Residue number

Figure 4.7. Comparison of peptide binding poses in different substrate analog bound sortase complexes. Expanded view of the active sites in *A*, Ba-SrtA-LPAT*; *B*, Sa-SrtA-LPAT* (Protein Data Bank code 2KID); and *C*, Sa-SrtB-NPQT* (Protein Data Bank code 4LFD). Proteins are represented by their solvent accessible surfaces, while the substrate analogs are labeled and shown as sticks. The active site Arg, Cys and His are colored blue, red, and green, respectively. Note that in Sa-SrtB-NPQT*, the active site His^{130} is obstructed from view by the side chain of Leu^{96}. *D*, superposition of the three substrate analogs from Ba-SrtA-LPAT* (cyan), Sa-SrtA-LPAT* (orange) and Sa-SrtB-NPQT* (green).

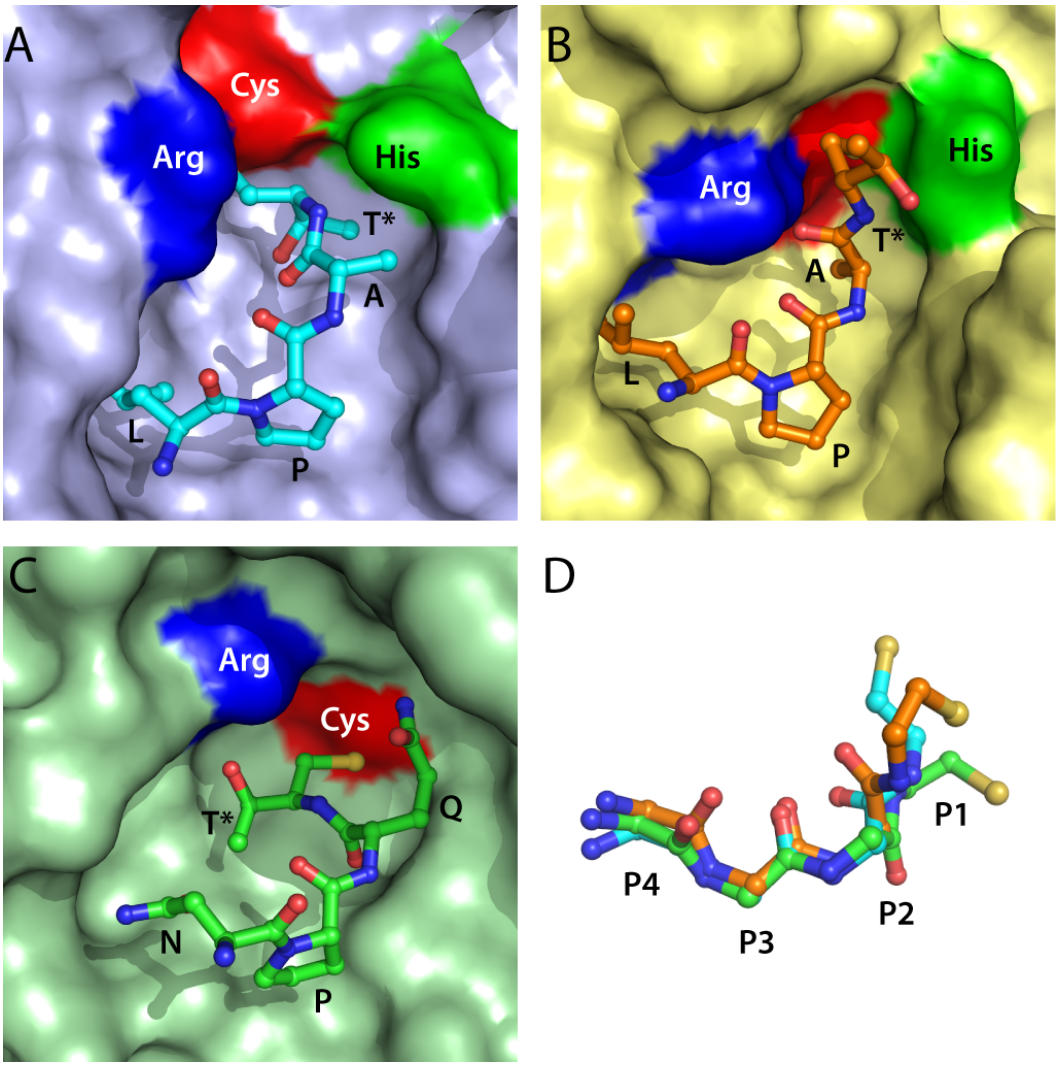
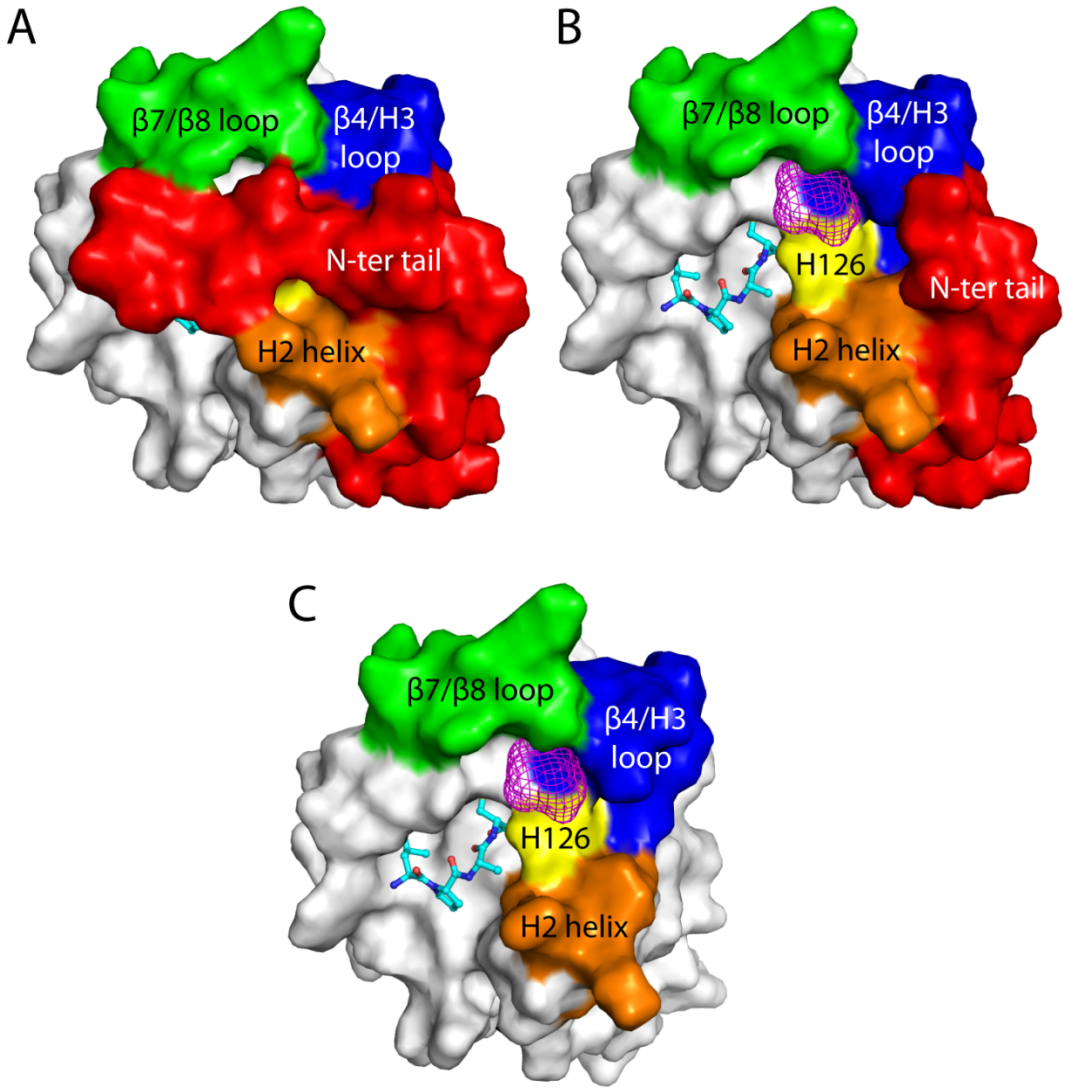


Figure 4.8. Potential lipid II binding site in Ba-SrtA. Solvent accessible surface representations are shown for Ba-SrtA-LPAT* with *A*, the entire N-terminal extension; *B*, Asp^{57}-Pro^{64} of the N-terminal extension removed; and *C*, the entire N-terminal extension removed. The N-terminal extension, H2 helix, β4/H3 loop, β7/β8 loop and active site His^{126} are colored red, orange, blue, green and yellow, respectively. Substrate analog LPAT* is shown as stick. Ligand binding pocket, potentially for lipid II binding, predicted by FTSite, is represented by magenta mesh.



4.8 Tables

Table 4.1. Statistics for the NMR structure of Ba-SrtA-LPAT* complex

	$\langle SA \rangle^a$	$\langle \overline{SA} \rangle^a$
R.m.s. deviations from NOE interproton distance restraints (Å)		
All (1861)	0.031 ± 0.002	0.025
Intermolecular (59)	0.054 ± 0.006	0.045
R.m.s. deviations from dihedral angle restraints (degrees)[b] (318)	0.501 ± 0.055	0.970
R.m.s. deviations from $^3J_{HN}{}^{\alpha}$ coupling constant (Hz) (77)	0.701 ± 0.021	0.648
R.m.s. deviations from secondary ^{13}C shifts		
$^{13}C_{\alpha}$ (ppm) (141)	1.166 ± 0.021	1.188
$^{13}C_{\beta}$ (ppm) (141)	1.187 ± 0.021	1.196
Residual dipolar coupling *R*-factors (%)[c]		
D_{NH} (84)	4.3 ± 0.2	6.4
$D_{NC'}$ (67)	23.1 ± 0.8	27.9
Deviations from idealized covalent geometry		
Bonds (Å)	0.0084 ± 0.0001	0.0082
Angles (degrees)	0.748 ± 0.013	0.769
Impropers (degrees)	0.814 ± 0.013	0.872
PROCHECK-NMR[d]		
Most favorable region (%)	88.5 ± 1.2	87.8
Additionally allowed region (%)	10.6 ± 1.3	11.5
Generously allowed region (%)	1.0 ± 0.5	0.7
Disallowed region (%)	0.0 ± 0.0	0.0
Coordinate precision[e]		
Protein backbone (Å)	0.42 ± 0.06	
Protein heavy atoms (Å)	0.89 ± 0.06	

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

[b] Experimental dihedral angle restraints comprised 150 ϕ, 144 ψ and 24 χ_1 angles

[c] The dipolar coupling R-factor ranges between 0 and 100% and is defined as the ratio of the r.m.s. deviation between observed and calculated values to the expected r.m.s. deviation if the vectors were randomly distributed, given by $[2D_a^2(4 + 3\eta^2)/5]^{1/2}$, where D_a is the magnitude of the principle component of the alignment tensor, and η is the rhombicity, calculated to be 10.6 Hz and 0.19, respectively.

[d] PROCHECK-NMR data includes residues Asp^{57} to Lys^{210} of Ba-$SrtA_{\Delta 56}$, and the substrate analog Boc-LPAT*.

[e] The coordinate precision is defined as the average atomic r.m.s. deviation of the 20 individual SA structures and their mean coordinates. The reported values are for residues Asp^{57} to Lys^{210} of Ba-$SrtA_{\Delta 56}$, and the LPAT* part of the substrate analog. The backbone value refers to the N, C^{α}, and C' atoms.

Table 4.2. Enzyme hydrolysis kinetics of the Ba-SrtA enzyme

	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($\mu M^{-1}s^{-1}$)
S59G	$3.3 \times 10^{-5} \pm 1 \times 10^{-6}$	280 ± 16	$1.2 \times 10^{-7} \pm 8 \times 10^{-9}$
I61A	$5.8 \times 10^{-5} \pm 5 \times 10^{-6}$	265 ± 38	$2.2 \times 10^{-7} \pm 4 \times 10^{-8}$
C187A	NA[a]	NA[a]	NA[a]
Ba-SrtA$_{\Delta 56}$	$3.6 \times 10^{-5} \pm 2 \times 10^{-6}$	306 ± 23	$1.2 \times 10^{-7} \pm 1 \times 10^{-8}$
Ba-SrtA$_{\Delta 64}$	$5.7 \times 10^{-5} \pm 2 \times 10^{-6}$	173 ± 11	$3.3 \times 10^{-7} \pm 2 \times 10^{-8}$

[a] NA, not active

Table 4.3. Dihedral angles of the substrate analogs

		Ba-SrtA-LPAT*	Sa-SrtA-LPAT*	Sa-SrtB-NPQT*
P4	ψ (°)	102 ± 12	135 ± 18	133
	χ_1 (°)	-125 ± 13	-1 ± 10	-138
	χ_2 (°)	-168 ± 74	67 ± 5	N/A
P3	ϕ (°)	-67 ± 4	-62 ± 13	-68
	ψ (°)	-171 ± 6	-176 ± 13	175
P2	ϕ (°)	-40 ± 5	49 ± 7	-72
	ψ (°)	-178 ± 8	150 ± 44	124
	χ_1 (°)	N/A	N/A	179
	χ_2 (°)	N/A	N/A	60
P1	ϕ [a] (°)	67 ± 5	-118 ± 39	-112
	ψ[b] (°)	155 ± 24	-2 ± 34	62
	χ_1 (°)	81 ± 6	40 ± 93	81

[a] Pseudo-phi angle defined by the dihedral angle between C(P2), N(P1), CA(P1) and CX(methylene carbon of P1).

[b] Pseudo-psi angle defined by the dihedral angle between N(P1), CA(P1), CX(methylene carbon of P1) and SG(sulfur of P1).

[c] NA, not applicable.

Table 4.4. Model-free parameters calculated for Ba-SrtA-LPAT*

Residue	Model	S^2	S^2_f	S^2_s	t_e (ps)	R_{ex} (Hz)
61	5	0.196 ± 0.011	0.944 ± 0.023	0.207 ± 0.013	927.1 ± 30.3	
62	5	0.266 ± 0.015	0.932 ± 0.017	0.285 ± 0.016	1046.3 ± 15.7	
63	5	0.302 ± 0.008	0.881 ± 0.008	0.343 ± 0.009	1087.5 ± 17.1	
66	5	0.385 ± 0.010	0.887 ± 0.010	0.434 ± 0.011	1141.1 ± 14.7	
67	5	0.506 ± 0.012	0.972 ± 0.010	0.521 ± 0.011	991.9 ± 23.8	
68	5	0.509 ± 0.010	0.950 ± 0.009	0.536 ± 0.010	1119.4 ± 20.7	
69	5	0.496 ± 0.011	0.917 ± 0.011	0.541 ± 0.011	1075.2 ± 45.1	
80	1	0.928 ± 0.007		0.928 ± 0.007		
82	1	0.978 ± 0.011		0.978 ± 0.011		
83	3	0.953 ± 0.006		0.953 ± 0.006		0.546 ± 0.139
84	1	0.932 ± 0.006		0.932 ± 0.006		
85	1	0.920 ± 0.007		0.920 ± 0.007		
88	1	0.916 ± 0.011		0.916 ± 0.011		
89	3	0.954 ± 0.010		0.954 ± 0.010		2.038 ± 0.219
90	1	0.937 ± 0.005		0.937 ± 0.005		
91	4	0.890 ± 0.019		0.890 ± 0.019	47.3 ± 12.7	1.048 ± 0.325
92	1	0.904 ± 0.005		0.904 ± 0.005		
94	1	0.916 ± 0.002		0.916 ± 0.002		
95	3	0.902 ± 0.005		0.902 ± 0.005		2.833 ± 0.212
98	1	1.000 ± 0.010		1.000 ± 0.010		
99	5	0.828 ± 0.009	0.900 ± 0.006	0.920 ± 0.008	1025.0 ± 119.9	
102	1	1.000 ± 0.003		1.000 ± 0.003		
104	3	1.000 ± 0.018		1.000 ± 0.018		1.971 ± 0.290
105	1	1.000 ± 0.014		1.000 ± 0.014		
106	3	0.899 ± 0.010		0.899 ± 0.010		3.734 ± 0.156
107	1	1.000 ± 0.015		1.000 ± 0.015		

109	1	1.000 ± 0.016		1.000 ± 0.016		
110	3	0.933 ± 0.014		0.933 ± 0.014		1.537 ± 0.442
111	1	0.969 ± 0.007		0.969 ± 0.007		
113	1	1.000 ± 0.005		1.000 ± 0.005		
114	5	0.776 ± 0.021	0.962 ± 0.018	0.807 ± 0.019	1019.6 ± 119.9	
115	1	0.892 ± 0.009		0.892 ± 0.009		
116	2	0.911 ± 0.009		0.911 ± 0.009	14.6 ± 4.2	
117	3	0.902 ± 0.012		0.902 ± 0.012		1.807 ± 0.595
118	4	0.940 ± 0.012		0.940 ± 0.012	18.7 ± 8.4	1.378 ± 0.244
119	1	0.903 ± 0.004		0.903 ± 0.004		
121	3	0.960 ± 0.007		0.960 ± 0.007		3.051 ± 0.269
122	3	0.959 ± 0.008		0.959 ± 0.008		2.119 ± 0.244
123	1	0.941 ± 0.005		0.941 ± 0.005		
124	1	0.989 ± 0.015		0.989 ± 0.015		
125	3	0.946 ± 0.010		0.946 ± 0.010		2.069 ± 0.374
130	3	0.949 ± 0.018		0.949 ± 0.018		2.247 ± 0.280
131	3	0.926 ± 0.012		0.926 ± 0.012		1.859 ± 0.320
132	3	0.886 ± 0.013		0.886 ± 0.013		2.389 ± 0.210
133	4	0.919 ± 0.010		0.919 ± 0.010	6.6 ± 1.9	1.286 ± 0.184
134	3	0.977 ± 0.015		0.977 ± 0.015		1.890 ± 0.379
135	1	1.000 ± 0.007		1.000 ± 0.007		
136	1	1.000 ± 0.024		1.000 ± 0.024		
137	3	0.938 ± 0.007		0.938 ± 0.007		3.901 ± 0.248
139	1	0.996 ± 0.016		0.996 ± 0.016		
140	1	0.956 ± 0.008		0.956 ± 0.008		
141	1	0.998 ± 0.004		0.998 ± 0.004		
143	5	0.866 ± 0.004	0.904 ± 0.006	0.957 ± 0.008	753.8 ± 198.5	
144	1	0.908 ± 0.004		0.908 ± 0.004		
145	1	0.990 ± 0.013		0.990 ± 0.013		

146	1	0.947 ± 0.005		0.947 ± 0.005		
147	1	0.933 ± 0.003		0.933 ± 0.003		
148	1	0.942 ± 0.008		0.942 ± 0.008		
149	3	0.911 ± 0.007		0.911 ± 0.007		0.945 ± 0.170
150	3	0.895 ± 0.009		0.895 ± 0.009		1.344 ± 0.223
151	3	0.921 ± 0.016		0.921 ± 0.016		0.630 ± 0.260
152	3	1.000 ± 0.011		1.000 ± 0.011		0.000 ± 0.252
153	2	0.878 ± 0.007		0.878 ± 0.007	40.3 ± 5.0	
154	1	0.918 ± 0.006		0.918 ± 0.006		
155	3	0.892 ± 0.014		0.892 ± 0.014		1.330 ± 0.206
157	1	0.941 ± 0.012		0.941 ± 0.012		
158	1	0.960 ± 0.009		0.960 ± 0.009		
161	1	0.981 ± 0.010		0.981 ± 0.010		
162	5	0.825 ± 0.010	0.864 ± 0.008	0.954 ± 0.011	1160.4 ± 321.6	
164	1	1.000 ± 0.006		1.000 ± 0.006		
165	4	0.883 ± 0.008		0.883 ± 0.008	16.0 ± 4.8	2.060 ± 0.221
169	3	0.968 ± 0.018		0.968 ± 0.018		6.770 ± 0.360
170	1	0.983 ± 0.003		0.983 ± 0.003		
172	3	1.000 ± 0.030		1.000 ± 0.030		13.073 ± 1.225
173	1	1.000 ± 0.004		1.000 ± 0.004		
175	1	1.000 ± 0.004		1.000 ± 0.004		
177	3	0.975 ± 0.009		0.975 ± 0.009		1.074 ± 0.367
178	1	1.000 ± 0.009		1.000 ± 0.009		
179	4	0.870 ± 0.007		0.870 ± 0.007	14.6 ± 2.9	2.106 ± 0.146
180	5	0.837 ± 0.012	0.883 ± 0.009	0.948 ± 0.009	1123.4 ± 300.9	
181	3	0.982 ± 0.007		0.982 ± 0.007		0.760 ± 0.200
183	3	0.926 ± 0.005		0.926 ± 0.005		0.900 ± 0.082

184	3	0.923 ± 0.007		0.923 ± 0.007		1.783 ± 0.356
185	1	1.000 ± 0.018		1.000 ± 0.018		
186	3	0.878 ± 0.057		0.878 ± 0.057		3.791 ± 1.127
189	1	1.000 ± 0.012		1.000 ± 0.012		
191	1	1.000 ± 0.019		1.000 ± 0.019		
198	3	0.941 ± 0.037		0.941 ± 0.037		2.408 ± 0.641
200	3	0.974 ± 0.012		0.974 ± 0.012		2.150 ± 0.250
201	1	0.951 ± 0.003		0.951 ± 0.003		
202	1	0.928 ± 0.009		0.928 ± 0.009		
204	1	0.984 ± 0.011		0.984 ± 0.011		
206	1	0.920 ± 0.005		0.920 ± 0.005		
207	5	0.860 ± 0.009	0.899 ± 0.007	0.956 ± 0.007	1444.0 ± 792.7	
208	5	0.831 ± 0.021	0.898 ± 0.012	0.926 ± 0.013	1104.0 ± 203.2	
209	5	0.638 ± 0.015	0.848 ± 0.009	0.752 ± 0.012	1309.9 ± 56.5	
210	5	0.367 ± 0.015	0.876 ± 0.013	0.419 ± 0.015	1042.7 ± 17.9	

4.9 References

1. Mock, M., and Fouet, A. (2001) *Annu Rev Microbiol* **55**, 647-671
2. Holty, J. E., Bravata, D. M., Liu, H., Olshen, R. A., McDonald, K. M., and Owens, D. K. (2006) *Ann Intern Med* **144**, 270-280
3. Guidi-Rontani, C., Weber-Levy, M., Labruyere, E., and Mock, M. (1999) *Mol Microbiol* **31**, 9-17
4. Navarre, W. W., and Schneewind, O. (1999) *Microbiol Mol Biol Rev* **63**, 174-229
5. Mazmanian, S. K., Liu, G., Ton-That, H., and Schneewind, O. (1999) *Science* **285**, 760-763
6. Spirig, T., Weiner, E. M., and Clubb, R. T. (2011) *Mol Microbiol* **82**, 1044-1059
7. Popp, M. W., and Ploegh, H. L. (2011) *Angew Chem Int Ed Engl* **50**, 5024-5032
8. Maresso, A. W., and Schneewind, O. (2008) *Pharmacol Rev* **60**, 128-141
9. Suree, N., Jung, M. E., and Clubb, R. T. (2007) *Mini Rev Med Chem* **7**, 991-1000
10. Gaspar, A. H., Marraffini, L. A., Glass, E. M., Debord, K. L., Ton-That, H., and Schneewind, O. (2005) *J Bacteriol* **187**, 4646-4655
11. Maresso, A. W., Chapa, T. J., and Schneewind, O. (2006) *J Bacteriol* **188**, 8145-8152
12. Marraffini, L. A., and Schneewind, O. (2007) *J Bacteriol* **189**, 6425-6436
13. Zink, S. D., and Burns, D. L. (2005) *Infect Immun* **73**, 5222-5228
14. Ton-That, H., Liu, G., Mazmanian, S. K., Faull, K. F., and Schneewind, O. (1999) *Proc Natl Acad Sci U S A* **96**, 12424-12429
15. Ton-That, H., Mazmanian, S. K., Alksne, L., and Schneewind, O. (2002) *J Biol Chem* **277**, 7447-7452

16. Marraffini, L. A., Ton-That, H., Zong, Y., Narayana, S. V., and Schneewind, O. (2004) *J Biol Chem* **279**, 37763-37770

17. Frankel, B. A., Tong, Y., Bentley, M. L., Fitzgerald, M. C., and McCafferty, D. G. (2007) *Biochemistry* **46**, 7269-7278

18. Huang, X., Aulabaugh, A., Ding, W., Kapoor, B., Alksne, L., Tabei, K., and Ellestad, G. (2003) *Biochemistry* **42**, 11307-11315

19. Frankel, B. A., Kruger, R. G., Robinson, D. E., Kelleher, N. L., and McCafferty, D. G. (2005) *Biochemistry* **44**, 11188-11200

20. Suree, N., Liew, C. K., Villareal, V. A., Thieu, W., Fadeev, E. A., Clemens, J. J., Jung, M. E., and Clubb, R. T. (2009) *J Biol Chem* **284**, 24465-24477

21. Bentley, M. L., Lamb, E. C., and McCafferty, D. G. (2008) *J Biol Chem* **283**, 14762-14771

22. Tian, B. X., and Eriksson, L. A. (2011) *J Phys Chem B* **115**, 13003-13011

23. Zong, Y., Mazmanian, S. K., Schneewind, O., and Narayana, S. V. (2004) *Structure* **12**, 105-112

24. Jacobitz, A. W., Wereszczynski, J., Yi, S. W., Amer, B. R., Huang, G. L., Nguyen, A. V., Sawaya, M. R., Jung, M. E., McCammon, J. A., and Clubb, R. T. (2014) *J Biol Chem* **289**, 8891-8902

25. Weiner, E. M., Robson, S., Marohn, M., and Clubb, R. T. (2010) *J Biol Chem* **285**, 23433-23443

26. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J Biomol NMR* **6**, 277-293

27. Garrett, D. S., Powers, R., Gronenborn, A. M., and Clore, G. M. (1991) *J Magn Reson* **95**, 214-220

28. Keller, R. (2004) *The Computer Aided Resonance Assignment Tutorial*, CATINA Verlag, Goldau, Switzerland

29. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., and Skelton, N. J. (2006) *Protein NMR spectroscopy*, Elsevier Science and Technology, San Diego, CA

30. Teng, Q. (2005) *Structural Biology: Practical NMR Applications*, Springer Verlag, New York

31. Iwahara, J., Wojciak, J. M., and Clubb, R. T. (2001) *J Biomol NMR* **19**, 231-241

32. Zwahlen, C., Legault, P., Vincent, S. J., Greenblatt, J., Konrat, R., and Kay, L. E. (1997) *J Am Chem Soc* **119**, 6711-6721

33. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) *J Biomol NMR* **44**, 213-223

34. Vuister, G. W., and Bax, A. (1993) *J Am Chem Soc* **115**, 7772-7777

35. Gagne, S. M., Tsuda, S., Li, M. X., Chandra, M., Smillie, L. B., and Sykes, B. D. (1994) *Protein Sci* **3**, 1961-1974

36. Clore, G. M., Bax, A., and Gronenborn, A. M. (1991) *J Biomol NMR* **1**, 13-22

37. Powers, R., Garrett, D. S., March, C. J., Frieden, E. A., Gronenborn, A. M., and Clore, G. M. (1993) *Biochemistry* **32**, 6744-6762

38. Herrmann, T., Guntert, P., and Wuthrich, K. (2002) *J Biomol NMR* **24**, 171-189

39. Herrmann, T., Guntert, P., and Wuthrich, K. (2002) *J Mol Biol* **319**, 209-227

40. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) *J Magn Reson* **160**, 65-73

41. Grishaev, A., and Bax, A. (2004) *J Am Chem Soc* **126**, 7281-7292

42. Ruckert, M., and Otting, G. (2000) *J Am Chem Soc* **122**, 7793-7797

43. Ottiger, M., Delaglio, F., and Bax, A. (1998) *J Magn Reson* **131**, 373-378

44. Bax, A., Kontaxis, G., and Tjandra, N. (2001) *Methods Enzymol* **339**, 127-174

45. Dosset, P., Hus, J. C., Marion, D., and Blackledge, M. (2001) *J Biomol NMR* **20**, 223-231

46. Koradi, R., Billeter, M., and Wuthrich, K. (1996) *J Mol Graph* **14**, 51-55, 29-32

47. DeLano, W. L. (2006) The PyMOL Molecular Graphics System. DeLano Scientific, LLC, Palo Alto, CA

48. Goddard, T. D., and Kneller, D. G. (2006) SPARKY 3. University of California, San Francisco

49. Naik, M. T., Suree, N., Ilangovan, U., Liew, C. K., Thieu, W., Campbell, D. O., Clemens, J. J., Jung, M. E., and Clubb, R. T. (2006) *J Biol Chem* **281**, 1817-1826

50. Bruschweiler, R., Liao, X., and Wright, P. E. (1995) *Science* **268**, 886-889

51. Lee, L. K., Rance, M., Chazin, W. J., and Palmer, A. G., 3rd. (1997) *J Biomol NMR* **9**, 287-298

52. Lipari, G., and Szabo, A. (1982) *J Am Chem Soc* **104**, 4546-4559

53. Lipari, G., and Szabo, A. (1982) *J Am Chem Soc* **104**, 4559-4570

54. Cole, R., and Loria, J. P. (2003) *J Biomol NMR* **26**, 203-213

55. Mandel, A. M., Akke, M., and Palmer, A. G., 3rd. (1995) *J Mol Biol* **246**, 144-163

56. Case, D. A., Darden, T. A., Cheatham, T. E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Goetz, A. W., Kolossvary, I., Wong, K. F., Paesani, F., Vanicek, J., Wolf, R. M., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M. J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Salomon-

Ferrer, R., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., and Kollman, P. A. (2012) AMBER 12. University of California, San Francisco

57. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) *J Comput Chem* **26**, 1781-1802

58. Suree, N., Yi, S. W., Thieu, W., Marohn, M., Damoiseaux, R., Chan, A., Jung, M. E., and Clubb, R. T. (2009) *Bioorg Med Chem* **17**, 7174-7185

59. Tollinger, M., Skrynnikov, N. R., Mulder, F. A., Forman-Kay, J. D., and Kay, L. E. (2001) *J Am Chem Soc* **123**, 11341-11352

60. Kleckner, I. R., and Foster, M. P. (2011) *Biochim Biophys Acta* **1814**, 942-968

61. Palmer, A. G., 3rd, Kroenke, C. D., and Loria, J. P. (2001) *Methods Enzymol* **339**, 204-238

62. Loria, J. P., Berlow, R. B., and Watt, E. D. (2008) *Acc Chem Res* **41**, 214-221

63. Schneewind, O., and Missiakas, D. M. (2012) *Philos Trans R Soc Lond B Biol Sci* **367**, 1123-1139

64. Zong, Y., Bice, T. W., Ton-That, H., Schneewind, O., and Narayana, S. V. (2004) *J Biol Chem* **279**, 31383-31389

65. Berjanskii, M. V., and Wishart, D. S. (2005) *J Am Chem Soc* **127**, 14970-14971

66. Berjanskii, M. V., and Wishart, D. S. (2008) *J Biomol NMR* **40**, 31-48

67. Neiers, F., Madhurantakam, C., Falker, S., Manzano, C., Dessen, A., Normark, S., Henriques-Normark, B., and Achour, A. (2009) *J Mol Biol* **393**, 704-716

68. Cozzi, R., Malito, E., Nuccitelli, A., D'Onofrio, M., Martinelli, M., Ferlenghi, I., Grandi, G., Telford, J. L., Maione, D., and Rinaudo, C. D. (2011) *FASEB J* **25**, 1874-1886

69. Cozzi, R., Prigozhin, D., Rosini, R., Abate, F., Bottomley, M. J., Grandi, G., Telford, J. L., Rinaudo, C. D., Maione, D., and Alber, T. (2012) *PLoS One* **7**, e49048

70. Schechter, I., and Berger, A. (1967) *Biochem Biophys Res Commun* **27**, 157-162

71. Comfort, D., and Clubb, R. T. (2004) *Infect Immun* **72**, 2710-2722

72. Kruger, R. G., Dostal, P., and McCafferty, D. G. (2004) *Anal Biochem* **326**, 42-48

73. Ngan, C. H., Hall, D. R., Zerbe, B., Grove, L. E., Kozakov, D., and Vajda, S. (2012) *Bioinformatics* **28**, 286-287

74. Brenke, R., Kozakov, D., Chuang, G. Y., Beglov, D., Hall, D., Landon, M. R., Mattos, C., and Vajda, S. (2009) *Bioinformatics* **25**, 621-627