UCLA UCLA Electronic Theses and Dissertations

Title

The Molecular Basis of Substrate Recognition, Catalysis, and Regulation in Sortase Enzymes

Permalink https://escholarship.org/uc/item/625123dg

Author Jacobitz, Alex William

Publication Date 2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

The Molecular Basis of Substrate Recognition, Catalysis, and Regulation in

Sortase Enzymes

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Biochemistry and Molecular Biology

by

Alex William Jacobitz

2015

ABSTRACT OF THE DISSERTATION

The Molecular Basis of Substrate Recognition, Catalysis, and Regulation in

Sortase Enzymes

By

Alex William Jacobitz

Doctor of Philosophy in Biochemistry and Molecular Biology University of California, Los Angeles, 2015 Professor Robert T. Clubb, Chair

Gram-positive bacteria utilize sortase enzymes to catalyze a transpeptidation reaction that covalently links protein substrates to the cell wall or polymerizes proteinaceous pili. Sortase enzymes are membrane anchored cysteine transpeptidases that recognize a cell wall sorting signal motif at the C-terminus of their primary protein substrate and covalently attach it to an amino-nucleophile located in their secondary substrate. For cell wall anchoring sortases, this secondary substrate is the cell wall precursor lipid II, while for pilin polymerizing sortases it is a lysine sidechain within their pilin protein substrate. A number of major questions have remained unanswered concerning the binding interactions that govern the substrate specificity of these enzymes, their catalytic mechanism, and the mechanism through which appendages that contact the active site regulate their activity. This dissertation advances the understanding of the sortase mechanism through the elucidation of new sortase structures, and the characterization of their dynamic behavior.

Chapter three of this thesis describes the solution structure of the class D sortase from Bacillus anthracis by NMR. Class D enzymes anchor proteins involved in bacterial sporulation to the cell wall, and this is the first structure of a class D sortase to ever be determined. NMR analysis of the enzyme uncovered a rigid substrate binding pocket and a novel dimerization interface. Chapter four describes the crystal structure of the class B sortase from Staphylococcus aureus bound to a substrate analog. This work provided new insight into the biophysical basis of substrate recognition. The structure combined with computational modeling and molecular dynamics simulations led to the discovery of a substrate-stabilized oxyanion hole that is used to stabilize tetrahedral reaction intermediates. Molecular dynamics simulations and the high degree of sequence conservation inherent in these enzymes suggest that all members of the sortase superfamily will stabilize high energy reaction intermediates in a similar manner. Chapter five investigates the additional structural features some sortases use to regulate access to their active site, focusing on sortase C-1 from Streptococcus pneumoniae as a model. Using NMR, mutagenesis, and biochemical experiments this work demonstrated that the enzyme utilizes a rigid N-terminal appendage, termed the "lid," to maintain the enzyme in an inactive state. The results of *in vitro* biochemical assays indicate that the lid prevents cleavage of the primary substrate by preventing access to the catalytic cysteine residue. Both *in vitro* activity and access to the active site cysteine could be increased by the incorporation of mutations which were shown by NMR to destabilize the lid and increase its flexibility. Based on these results we propose that on the cell surface, lid containing class C enzymes exist in a dormant state and are only activated during pilus biogenesis by interactions with either their substrate or other factors on the cell surface. The results of these experiments have provided new insight into substrate binding, catalysis, and regulation in sortase enzymes.

The dissertation of Alex Jacobitz is approved.

James U. Bowie

Beth A. Lazazzera

Robert T. Clubb, Committee Chair

University of California, Los Angeles

2015

DEDICATION

To my father, who taught me the value of hard work and perseverance, and always faced adversity with humor and grace. This would not have been possible without his strength and guidance.

- Thought is the wind, knowledge the sail, and mankind the vessel -

TABLE OF CONTENTS

Chapter 1. Introduction to Sortase Enzymes		
1.1	Overview	2
1.2	Introduction to Sortase Biology	3
1.3	References	7
Chap	pter 2. Structures and mechanism of the sortase superfamily of transpeptidases	13
2.1	Overview	14
2.2	Introduction	15
2.3	The Staphylococcus aureus SrtA Paradigm	17
2.4	Classification of Sortases	19
2.5	Structural Biology of Sortase Enzymes	20
2.6	Molecular Basis of Substrate Recognition	29
2.7	The Molecular Mechanism of Sortase Enzymes	
2.8	Sortases as Powerful In Vitro Biochemical Reagents	40
2.9	Development of Structure-Based Therapeutics	42
2.10	Conclusion	46
2.11	Figures	49
2.12	References	60

	Bacill	<i>us anthracis</i> spores	76
3.1	Overview		78
3.2	Introd	uction	79
3.3	Result	s and Discussion	82
	3.3.1	Design and Validation of Soluble BaSrtC	82
	3.3.2	NMR Structure of $BaSrtC_{\Delta 55}$	83
	3.3.3	Disordered Loops Positioned near the Active Site Might Mediate in vitro	
		Dimerization	89
	3.3.4	Class A and D Sortases Are Structurally Related	89
	3.3.5	Model of the BaSrtC $_{\Delta 55}$ -Sorting Signal Complex	90
3.4	Experi	mental Procedures	93
	3.4.1	Cloning, Protein Expression, and Purification	93
	3.4.2	NMR Spectroscopy and Structure Calculations	95
	3.4.3	¹⁵ N Relaxation Measurements	96
	3.4.4	Analytical Ultracentrifugation and Enzyme Assays	97
3.5	Figures		98
3.6	References		.108
3.7	Supplementary materials		.117

Chapter 3. Solution structure of the sortase required for efficient production of infectious

Chapter 4. Structural and Computational Studies of the *Staphylococcus aureus* Sortase B-Substrate Complex Reveal a Substrate-stabilized Oxyanion Hole......119

4.1	Overview12		21
4.2	Introduction122		
4.3 Result		s1	24
	4.3.1	Crystal Structure of the SrtB-NPQT* Complex and Computational Modeling	
		of the Thioacyl Intermediate1	24
	4.3.2	Identification of Enzyme-Substrate Interactions Required for Catalysis1	27
	4.3.3	Molecular Dynamics Simulations of SrtA and SrtB Reveal a Conserved	
		Mechanism through which the Substrate Stabilizes the Positioning of the	
		Active Site Arginine Residue1	29
4.4	Discus	ssion1	32
4.5	Experimental Procedures		
	4.5.1	Production, Crystallization, and Structure Determination of SrtB-NPQT*	
		Complex1	39
	4.5.2	Transpeptidation Assay1	40
	4.5.3	Computational Modeling and Molecular Dynamics1	41
4.6	Figure	es1	43
4.7	Refere	ences1	54

Introduction	.167
Results	.171
	Introduction

	5.3.1	Initial NMR analysis indicates SpnSrtC1 forms dimers in solution	171
	5.3.2	NMR dynamics data indicate that the SpnSrtC1 lid is not dynamic in vitro	172
	5.3.3	Mutations to the "lid" of SpnSrtC1 make the region highly dynamic in	
		solution	175
	5.3.4	Mutation of the lid is required to induce in vitro hydrolysis	176
	5.3.5	A closed lid is important for protecting the active site cysteine from	
		oxidation	178
5.4	Discussion		178
5.5 Experimental Procedures		imental Procedures	184
	5.5.1	Cloning, protein expression and purification	184
	5.5.2	NMR spectroscopy	185
	5.5.3	In vitro cleavage assay	186
	5.5.4	<i>In vitro</i> oxidation assay	187
5.6	Figure	es	188
5.7	Refere	ences	196

LIST OF FIGURES

Figure 2.1	Biological mechanism of sortase	19
Figure 2.2	Structure of sortase	50
Figure 2.3	Structural variation by class of sortase	53
Figure 2.4	Sorting signal recognition	54
Figure 2.5	The substrate stabilized oxyanion hole	55
Figure 2.6	Molecular mechanism of sortase enzymes	56
Figure 3.1	NMR spectra and amino acid sequence of BaSrtC) 8
Figure 3.2	MALDI-TOF demonstrating $BaSrtC_{\Delta 55}$ can cleave the LPNTA-	
	containing peptide10)0
Figure 3.3	Structures of <i>B. anthracis</i> class D sortase)1
Figure 3.4	Dynamics and ultracentrifugation analysis10)4
Figure 3.5	Comparison of representative structures from class A–D sortases10)6
Figure 3.6	Substrate binding and active site architecture)7
Figure S3.1	Residue by residue T_2 , T_1 and ${}^{15}N{}^{1}H$ Heteronuclear NOE data for BaSrtC _{$\Delta 55$} 11	17
Figure 4.1	Structure of the SrtB-NPQT* complex14	13
Figure 4.2	Structure and interactions of the NPQT* modifier in complex with SrtB14	15
Figure 4.3	Expanded view of the active site in the energy-minimized model of the SrtB-	
	NPQT thioacyl intermediate	17
Figure 4.4	Transpeptidation activity of wild-type and mutant SrtB14	18
Figure 4.5	Alignment of SrtB-NPQT* and SrtA-LPAT*15	50
Figure 4.6	Structures and free energy profiles of the SrtA and SrtB thioacyl complexes15	51

Figure 4.7	The SrtB transpeptidation mechanism showing how the sorting signal may	
	stabilize the oxyanion hole	153
Figure 5.1	S. pneumoniae SrtC1 structure and lid motif	188
Figure 5.2	NMR analysis shows SpnSrtC1 is dimeric in vitro	189
Figure 5.3	NMR dynamics of SpnSrtC1	190
Figure 5.4	NMR shows SpnSrtC1 mutants have more dynamic lids	192
Figure 5.5	SpnSrtC1 lid mutants are more active in vitro	194
Figure 5.6	The lid of SpnSrtC1 protects the active site cysteine from modification by a	
	sulfhydryl modifier	195

LIST OF TABLES

Table 2.1	Structurally characterized sortase enzymes	51
Table 2.2	Activity of sortase enzymes in vitro	
Table 3.1	Structural Statistics for the Solution Structure of $BaSrtC_{\Delta 55}$	
Table 4.1	Crystallographic data collection and refinement statistics	144

ACKNOWLEDGEMENTS

First and foremost I would like to thank my advisor, Dr. Robert Clubb for his guidance, support, and for the unlimited supply of coffee he provided: without which my PhD would not have been possible. My collaborators, Dr. Jeff Wereszczynski, Dr. J. Andrew McCammon, Dr. Sung Wook Yi, and Dr. Michael Jung for helping in the design and execution of numerous experiments, and I would like to additionally thank Jeff specifically for teaching me to run molecular dynamics simulations. Dr. Michael Sawaya for patiently teaching me X-ray crystallography. Dr. Robert Peterson for helping me to better understand NMR and for writing new pulse programs when necessary. My rotation mentors, Dr. Todd Yeates and Dr. William Gelbart for their advice on topics in and out of the lab. My committee members: Dr. James Bowie, Dr. Pascal Egea, Dr. Beth Lazazzera, and Dr. Joseph Loo for their thoughts and guidance. The undergraduates, John Cho, Allen Tsiyer, and Scott McConnell, and rotation students, Brendan Amer, Grace Huang, Angelyn Nguyen, Reid O'Brien Johnson, Kevin Eden, Mimi Ho, and Jessica Davis, whose efforts in the lab proved to be invaluable. Members of the Clubb lab, Dr. Scott Robson, Dr. Valerie Villareal, Dr. Thomas Spirig, Dr. Ethan Weiner, Dr. Reza Malmirchegini, Dr. Albert Chan, Megan Sjodt, Brendan Amer, Grace Huang, Michele Kattke, and Ramsay Macdonald, whose friendship, scientific discourse, and moral support helped to get me through graduate school. My friends and family for putting up with me for the last 6 years. The Cellular and Molecular Biology Training Grant, Audree Fowler Fellowship in Protein Science, John M. Jordan Memorial Award for Research in Biochemistry, and the UCLA Graduate Division Dissertation Year Fellowship for supporting me throughout my graduate

career, and for the additional opportunities for enriching my scientific development that they have provided.

Chapter two of this dissertation is a version of a manuscript in preparation titled: Structures and Mechanism of the Sortase Superfamily of Transpeptidases. Jacobitz, A. W., Kattke, M. D., Wereszczynski, J., and Clubb, R. T. 2015.

Chapter three of this dissertation is a version of the manuscript: Robson S. A., Jacobitz A. W., Clubb R.T., Phillips M.L., Clubb R.T. Solution Structure of the Sortase Required for Efficient Production of Infectious Bacillus anthracis Spores. *Biochemistry*. 2012; 51(40):7953-7963. doi:10.1021/bi300867t.<u>http://pubs.acs.org/doi/abs/10.1021/bi300867t</u>. Reproduced with permission. Copyright 2011, American Chemical Society. We express great appreciation to Dr. Joseph Loo and members of the Clubb laboratory for their assistance.

Chapter four of this dissertation is a version of the manuscript: Jacobitz, A. W.,
Wereszczynski, J., Yi, S. W., Amer, B. R., Huang, G. L., Nguyen, A. V., Sawaya, M. R., Jung,
M. E., McCammon, J. A., Clubb, R. T. Structural and computational studies of the
Staphylococcus aureus Sortase B-substrate complex reveal a substrate-stabilized oxyanion hole. *Journal of Biological Chemistry*. 2014; jbc.M113.509273.

http://doi.org/10.1074/jbc.M113.509273 Reproduced with permission. Copyright 2014, *Journal of Biological Chemistry*. This work was supported, in whole or in part, by National Institutes of Health Grants AI52217 (to R. T. C. and M. E. J.), GM31749 (to J. A. M.), and T32 GM007185 (to A. W. J. and G. L. H.). This work was also supported by the National Science Foundation, the Howard Hughes Medical Institute, the Center for Theoretical Biological Physics, the National Biomedical Computation Resource, and the National Science Foundation Supercomputer Centers.

Chapter Five of this dissertation is a version of a manuscript in preparation titled: The "lid" of *S. pneumoniae* SrtC1 is not dynamic in solution and regulates substrate access to the active site. Jacobitz A. W., Yi, S. W., McConnell, S., Peterson, R., Jung, M. E., and Clubb, R. T. 2015.

Vita

2007 - 2009	Bachelors of Arts, Chemistry
	California State University, Channel Islands
2010 - 2011	Teaching Assistant, Department of Chemistry and Biochemistry
	University of California, Los Angeles
2011	Candidate in Philosophy, Biochemistry and Molecular Biology
	University of California, Los Angeles
2011 – 2014	Cellular and Molecular Biology Training Grant
2013 - 2014	Audree Fowler Fellowship in Protein Science
2014	John M. Jordan Memorial Award for Research in Biochemistry
2014 - 2015	UCLA Graduate Division Dissertation Year Fellowship

PUBLICATIONS

Jacobitz, A. W., Wereszczynski, J., Yi, S. W., Amer, B. R., Huang, G. L., Nguyen, A. V., Sawaya, M. R., Jung, M. E., McCammon, J. A. & Clubb, R. T. Structural and computational studies of the Staphylococcus aureus Sortase B-substrate complex reveal a substrate-stabilized oxyanion hole. *J. Biol. Chem.* (2014). doi:10.1074/jbc.M113.509273

Robson, S. A., **Jacobitz, A. W.**, Phillips, M. L. & Clubb, R. T. Solution Structure of the Sortase Required for Efficient Production of Infectious Bacillus anthracis Spores. *Biochemistry*. 51, 7953–7963 (2012). doi:10.1021/bi300867t

King, N. P., Jacobitz, A. W., Sawaya, M. R., Goldschmidt, L. & Yeates, T. O. Structure and folding of a designed knotted protein. *Proc. Natl. Acad. Sci.* (2010). doi:10.1073/pnas.1007602107

Chapter 1

Introduction to the Biology of Sortase Transpeptidase Enzymes

1.1 Overview

Sortase enzymes are ubiquitous in Gram-positive bacteria. These enzymes perform a transpeptidation reaction that covalently attaches proteins to the cell wall, or polymerizes protein substrates into long filamentous pili. These surface anchored proteins and pili serve to augment the pathogenic potential of their bacterial proprietor by allowing the organism to more effectively interact with host cells and other bacteria. Sortase anchored surface proteins perform a variety of functions including adhesion, immune evasion, nutrient acquisition, and spore formation. Numerous studies have shown that the removal of sortase enzymes is generally not lethal to the bacterium, but often renders the bacterium avirulent. This observation has led to a wealth of research into the biology, structures, mechanism, regulation, and inhibition of these enzymes.

1.2 Introduction to Sortase Biology

Gram-positive organisms are recognized by the presence of a thick cell wall. While its function in protecting the cell against mechanical and osmotic stresses is well known, the cell wall also provides an additional function as a scaffold for the display of a large number of surface proteins by sortase enzymes¹⁻³. All sortase enzymes catalyze the formation of a peptide bond that covalently joins two substrates together. For most sortases, this transpeptidation reaction links a substrate protein bearing an appropriate cell wall sorting signal (CWSS) containing a C-terminal 5 residue recognition motif and transmembrane helix to the cell wall, but a select subgroup of the enzymes instead catalyze the formation of an isopeptide bond between a CWSS containing protein and a lysine sidechain of a pilin motif containing protein. Based on sequence similarity and structural homology, sortases can be categorized into 6 distinct families. These families each typically operate to anchor a specific type of substrate, with the class A enzymes being constitutively expressed and anchoring a number of diverse proteins to the cell surface, and each of the other types being relegated to a separate operon and only expressed along with their substrates under specific conditions. Most Gram-positives have at least one sortase gene, with certain bacteria harboring genes for multiple sortases which typically function non-redundantly to attach a specific subset of surface proteins to the cell wall. For bacteria with a single sortase, that sortase is typically class A and will attach multiple proteins to the cell wall, with the most prolific of these enzymes being sortase A from Listeria monocytogenes, which is predicted to display 43 distinct proteins^{4,5}. Other enzymes, like *Staphylococcus aureus* sortase B, are instead dedicated to mounting a single substrate⁶.

The majority of sortase enzymes studied to date catalyze the covalent attachment of their substrates to the cell wall, and many of these enzymes have been shown to be particularly important in virulence. Genetic knock-outs of sortase A and B have both been shown to reduce virulence in a mouse model of infection for the human pathogens *S. aureus*^{7,8}, and *Bacillus anthracis*⁹, without affecting the ability of these organisms to grow in rich media. Sortase A KOs have also been shown to reduce virulence of several other human pathogens including *Listeria monocytogenes*^{10,11}, and *Streptococcus pneumoniae*^{12–14}. These studies have highlighted the importance of this class of enzymes and led to the development of a number of small-molecule sortase inhibitors which may eventually lead to the production of a novel anti-virulence therapeutic¹⁵.

In addition to their importance as virulence factors, additional work has gone into understanding the need for multiple classes of sortase and these studies have shown that the different classes function non-redundantly to attach specific substrates to specific regions of the cell wall. While all sortase enzymes are known to recognize a CWSS, the individual classes generally recognize different 5 residue sorting signal motifs, and the structural basis of this recognition is only beginning to be understood^{2,16,17}. Beyond their ability to recognize different sorting signals, these different classes of enzymes have also been shown to anchor their substrates to distinct locations on the cell wall. While most sortases mount their proteins to the outermost surface of the cell wall, certain class B enzymes which mount proteins involved in delivering heme-iron to the cell have been shown to anchor proteins to a more buried location within the cell wall, which is likely necessary to develop a heme transport conduit to deliver heme from the surface of the cell wall to the cell membrane^{18,19}. Sortase enzymes and their substrates have also been shown to colocalize to distinct foci on the bacterial surface. Class A sortases in cocci have been shown to specifically anchor substrates at division septa, where subsequent growth and expansion will lead to a ring-like distribution of surface proteins across

the cell wall^{20–22}. For rod shaped bacteria, cell surface anchoring has been localized to helical patterns along the cell surface^{1,23}. Certain class C sortases which are typically involved in pilus polymerization have also been implicated in surface protein localization, and class A and C sortases have been shown to anchor pili to different locations on the cell surface²⁴. While it is now well established that certain sortases are capable of localizing their substrates to discrete locations on the cell wall, further study is needed to understand the mechanisms that cause this localization to occur.

Instead of attaching proteins to the cell wall, a subset of sortase enzymes catalyze the polymerization of pili. These sortases are typically found clustered in genomic islands along with their substrates, and some bacteria can harbor genomic islands coding for the production of multiple distinct of pili²⁵. These pilus islands typically harbor at least 2 sortases and at least 2 substrates, and the molecular mechanism by which they coordinate to assemble pili is only beginning to be understood⁴. Typically these operons code for one primary pilus protein that is polymerized to make up the shaft of the pilus, and then one or two accessory proteins that function as either the base of the pilus for attachment to the cell wall, or as an additional adhesin to be attached as the tip, or elaborate the length of the shaft²⁶. Many of the class C sortases have high sequence identity, and some studies have suggested that some of these proteins are multifaceted, polymerizing the pilus, attaching accessory pilus proteins, and attaching the pilus to the wall, and that other sortases in the operon may function semi-redundantly^{24,27-29}. Interestingly, several more recent studies have indicated that individual sortases may each be dedicated to specific steps in the pilin polymerization process, recognizing different pilin motifs at from separate domains within a single pilin protein substrate to build more complex branched, or "knobbed" structures^{30,31}.

While much is known about the basic biology of sortase enzymes, structural studies in this field have answered, and will continue to answer questions about their substrate recognition, molecular mechanism, regulation, and the interactions they have with other proteins on the cell surface that are all necessary to build a complete understanding of the sorting reaction that these enzymes perform on the cell surface.

1.3 References

- Liew PX, Wang CLC, Wong S-L. Functional characterization and localization of a Bacillus subtilis sortase and its substrate and use of this sortase system to covalently anchor a heterologous protein to the B. subtilis cell wall for surface display. *J Bacteriol*. 2012;194(1):161-175. doi:10.1128/JB.05711-11.
- Paterson GK, Mitchell TJ. The biology of Gram-positive sortase enzymes. *Trends Microbiol.* 2004;12(2):89-95. doi:10.1016/j.tim.2003.12.007.
- Desvaux M, Dumas E, Chafsey I, Hébraud M. Protein cell surface display in Grampositive bacteria: from single protein to macromolecular protein structure. *FEMS Microbiol Lett.* 2006;256(1):1-15. doi:10.1111/j.1574-6968.2006.00122.x.
- Spirig T, Weiner EM, Clubb RT. Sortase enzymes in Gram-positive bacteria. *Mol Microbiol.* 2011;82(5):1044-1059. doi:10.1111/j.1365-2958.2011.07887.x.
- Boekhorst J, de Been MWHJ, Kleerebezem M, Siezen RJ. Genome-wide detection and analysis of cell wall-bound proteins with LPxTG-like sorting motifs. *J Bacteriol*. 2005;187(14):4928-4934. doi:10.1128/JB.187.14.4928-4934.2005.
- Mazmanian SK, Ton-That H, Su K, Schneewind O. An iron-regulated sortase anchors a class of surface protein during Staphylococcus aureus pathogenesis. *Proc Natl Acad Sci U* SA. 2002;99(4):2293-2298. doi:10.2307/3057958.
- 7. Mazmanian SK, Liu G, Jensen ER, Lenoy E, Schneewind O. Staphylococcus aureus sortase mutants defective in the display of surface proteins and in the pathogenesis of

animal infections. *Proc Natl Acad Sci U S A*. 2000;97(10):5510-5515. doi:10.1073/pnas.080520697.

- Weiss WJ, Lenoy E, Murphy T, et al. Effect of srtA and srtB gene expression on the virulence of Staphylococcus aureus in animal models of infection. *J Antimicrob Chemother*. 2004;53(3):480-486. doi:10.1093/jac/dkh078.
- 9. Zink SD, Burns DL. Importance of srtA and srtB for growth of Bacillus anthracis in macrophages. *Infect Immun.* 2005;73(8):5222-5228. doi:10.1128/IAI.73.8.5222.
- Bierne H, Mazmanian SK, Trost M, et al. Inactivation of the srtA gene in Listeria monocytogenes inhibits anchoring of surface proteins and affects virulence. *Mol Microbiol*. 2002;43(4):869-881. http://www.ncbi.nlm.nih.gov/pubmed/11929538. Accessed August 20, 2015.
- Garandeau C, Réglier-Poupet H, Dubail I, Beretti J-L, Berche P, Charbit A. The sortase SrtA of Listeria monocytogenes is involved in processing of internalin and in virulence. *Infect Immun.* 2002;70(3):1382-1390. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=127754&tool=pmcentrez&ren dertype=abstract. Accessed August 21, 2015.
- Chen S, Paterson GK, Tong HH, Mitchell TJ, DeMaria TF. Sortase A contributes to pneumococcal nasopharyngeal colonization in the chinchilla model. *FEMS Microbiol Lett*. 2005;253(1):151-154. doi:10.1016/j.femsle.2005.09.052.

- Paterson GK, Mitchell TJ. The role of Streptococcus pneumoniae sortase A in colonisation and pathogenesis. *Microbes Infect*. 2006;8(1):145-153. doi:10.1016/j.micinf.2005.06.009.
- Kharat AS, Tomasz A. Inactivation of the srtA Gene Affects Localization of Surface Proteins and Decreases Adhesion of Streptococcus pneumoniae to Human Pharyngeal Cells *In vitro*. *Infect Immun*. 2003;71(5):2758-2765. doi:10.1128/IAI.71.5.2758-2765.2003.
- Suree N, Jung ME, Clubb RT. Recent advances towards new anti-infective agents that inhibit cell surface protein anchoring in Staphylococcus aureus and other gram-positive pathogens. *Mini Rev Med Chem*. 2007;7(10):991-1000. doi:10.2174/138955707782110097.
- Jacobitz AW, Wereszczynski J, Yi SW, et al. Structural and computational studies of the Staphylococcus aureus Sortase B-substrate complex reveal a substrate-stabilized oxyanion hole. *J Biol Chem.* February 2014:jbc.M113.509273. doi:10.1074/jbc.M113.509273.
- Suree N, Liew CK, Villareal VA, et al. The Structure of the Staphylococcus aureus Sortase-Substrate Complex Reveals How the Universally Conserved LPXTG Sorting Signal Is Recognized. *J Biol Chem.* 2009;284(36):24465-24477. doi:10.1074/jbc.M109.022624.
- Marraffini LA, Schneewind O. Anchor Structure of Staphylococcal Surface Proteins V. ANCHOR STRUCTURE OF THE SORTASE B SUBSTRATE IsdC. *J Biol Chem*. 2005;280(16):16263-16271. doi:10.1074/jbc.M500071200.

- Mazmanian SK, Skaar EP, Gaspar AH, et al. Passage of Heme-Iron Across the Envelope of Staphylococcus aureus. *Science (80-)*. 2003;299(5608):906-909. doi:10.1126/science.1081147.
- Raz A, Fischetti VA. Sortase A localizes to distinct foci on the Streptococcus pyogenes membrane. *Proc Natl Acad Sci U S A*. 2008;105(47):18549-18554. doi:10.1073/pnas.0808301105.
- DeDent A, Bae T, Missiakas DM, Schneewind O. Signal peptides direct surface proteins to two distinct envelope locations of Staphylococcus aureus. *EMBO J.* 2008;27(20):2656-2668. doi:10.1038/emboj.2008.185.
- DeDent AC, McAdow M, Schneewind O. Distribution of protein A on the surface of Staphylococcus aureus. *J Bacteriol*. 2007;189(12):4473-4484. doi:10.1128/JB.00227-07.
- Bruck S, Personnic N, Prevost M-C, Cossart P, Bierne H. Regulated Shift from Helical to Polar Localization of Listeria monocytogenesCell Wall-Anchored Proteins. *J Bacteriol*. 2011;193(17):4425-4437. doi:10.1128/JB.01154-10.
- Fälker S, Nelson AL, Morfeldt E, et al. Sortase-mediated assembly and surface topology of adhesive pneumococcal pili. *Mol Microbiol*. 2008;70(3):595-607. doi:10.1111/j.1365-2958.2008.06396.x.
- Mandlik A, Swierczynski A, Das A, Ton-That H. Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. *Trends Microbiol*. 2008;16(1):33-40. doi:10.1016/j.tim.2007.10.010.

- Hendrickx APA, Budzik JM, Oh S-Y, Schneewind O. Architects at the bacterial surface sortases and the assembly of pili with isopeptide bonds. *Nat Rev Microbiol*. 2011;9(3):166-176. doi:10.1038/nrmicro2520.
- Mora M, Bensi G, Capo S, et al. Group A Streptococcus produce pilus-like structures containing protective antigens and Lancefield T antigens. *Proc Natl Acad Sci U S A*. 2005;102(43):15641-15646. doi:10.1073/pnas.0507808102.
- Ton-That H, Marraffini LA, Schneewind O. Sortases and pilin elements involved in pilus assembly of Corynebacterium diphtheriae. *Mol Microbiol.* 2004;53(1):251-261. doi:10.1111/j.1365-2958.2004.04117.x.
- Neiers F, Madhurantakam C, Fälker S, et al. Two crystal structures of pneumococcal pilus sortase C provide novel insights into catalysis and substrate specificity. *J Mol Biol.* 2009;393(3):704-716. doi:10.1016/j.jmb.2009.08.058.
- El Mortaji L, Terrasse R, Dessen A, Vernet T, Di Guilmi AM. Stability and Assembly of Pilus Subunits of Streptococcus pneumoniae. *J Biol Chem*. 2010;285(16):12405-12415. doi:10.1074/jbc.M109.082776.
- El Mortaji L, Fenel D, Vernet T, Di Guilmi AM. Association of RrgA and RrgC into the Streptococcus pneumoniae pilus by sortases C-2 and C-3. *Biochemistry*. 2012;51(1):342-352. doi:10.1021/bi201591n.

Chapter 2

Structures and Mechanism of the Sortase Superfamily of

Transpeptidases

The work described in this chapter is a version of a manuscript to be submitted for publication.

Jacobitz, A. W., Kattke, M. D., Wereszczynski, J., and Clubb, R. T., 2015.

2.1 Overview

Sortase enzymes are ubiquitous in Gram-positives and catalyze a transpeptidation reaction that can be used to anchor surface proteins to the peptidoglycan or polymerize them into long filamentous pili. Gram-positive organisms require these surface proteins to effectively interact with their environment and to mount infections. Because many of these surface proteins are necessary for pathogenesis, sortase enzymes are considered virulence factors, and this designation has led to extensive characterization of their structures and mechanism. To date, over 40 structures of sortase enzymes have been deposited in the PDB and this wealth of structural information combined with numerous *in vitro* and *in vivo* analyses of their catalytic activity has led to a thorough understanding of the molecular mechanism of substrate binding and catalysis. These structural and mechanistic observations have prompted a number of computational studies to understand the dynamic behavior of these enzymes and to develop inhibitors. Here we review what is known about the structures and mechanism of these enzymes.

2.2 Introduction

Gram-positive organisms utilize a thick cell wall composed of cross-linked peptidoglycan to protect themselves against environmental stressors and as a scaffold for the display of a large number of surface proteins by sortase enzymes. Sortase enzymes are cysteine transpeptidases that link their two substrates together via the formation of a new covalent bond^{1–3}. Many of these enzymes catalyze the covalent attachment of surface proteins to peptidoglycan, while a subset of the family instead catalyzes the polymerization of filamentous pili on the cell surface. Since their discovery over 15 years ago⁴, genome sequencing efforts have revealed that Gram-positive bacteria harboring a conventional cell wall typically encode multiple sortase enzymes⁵⁻⁷. The surface proteins which these enzymes attach to the cell wall are important for adhesion, nutrient acquisition, spore formation, and immune evasion^{2,6,8}, and it has been shown that sortase knockouts are defective in establishing persistent infections^{9,10}. Since many surface proteins are necessary for efficiently interacting with a host organism, and their deletion results in a defect in pathogenesis, sortase enzymes have been classified as virulence factors^{9,10}. This designation has led to the development of several small-molecule inhibitors of sortase, and the search for molecules with increased potency and specificity is a subject of ongoing research^{8,11–15}. Beyond their importance as a potential drug target, the ability of these enzymes to covalently join two distinct substrates through the formation of a new peptide bond has led to their development as a biochemical reagent to covalently attach proteins to each other, attach fluorophores or drugs to proteins or antibodies, to cyclize proteins, and to immobilize peptides on solid surfaces 16-18.

The sortase superfamily is large. At present over 800 sortase enzymes in 300 species of bacteria have been identified⁶. Based on their primary sequences these enzymes have been classified into six distinct families, whose members share related atomic structures and

frequently have similar functions. On the cell surface all sortases characterized to date function as cysteine transpeptidase enzymes that attach surface proteins bearing a C-terminal cell wall sorting signal (CWSS) consisting of a C-terminal hydrophobic domain¹⁹ with a positively charged cytoplasmic anchor and a 5 residue recognition motif to an amino nucleophile located on the cell surface. While the basic chemistry of this reaction appears to be highly conserved, sortases can function to either attach proteins to the cell wall, or assemble pili (Fig. 1). All sortases characterized to date are believed to be associated with the extracellular membrane via an N-terminal trans-membrane helix and a short positively charged cytoplasmic anchor^{7,20,21}. Additionally, sequence conservation and a number of structural and biochemical studies have determined that the active site of all sortase enzymes studied to date consists of a His-Cys-Arg catalytic triad^{22,23}. In atomic structures of sortases, these residues are invariantly clustered in the active site such that the cysteine residue is bracketed by the histidine and arginine residues. Sortase enzymes are also known to harbor a highly conserved TLXTC motif leading to, and encompassing the catalytic cysteine²⁴ which could be important for stabilizing the position of the active site cysteine residue and maintaining it in a catalytically competent orientation. While the basic architecture is conserved, unique class specific variations enable distinct CWWS and nucleophiles to be recognized, and in some instances the variations appear to play critical regulatory roles.

A number of excellent reviews have been written regarding the overall function of sortases in biology, their specific role in pilin assembly, their development as biochemical reagents and efforts to discover inhibitors of this important family^{2,7,8,25–29}. Here we review what is known about the structural basis of catalysis and sortase function on the cell surface, highlighting recent structural and computational studies that have revealed unique family

specific structural variations, as well as the mechanism of sorting signal recognition and enzyme regulation.

2.3 The Staphylococcus aureus SrtA Paradigm

The majority of early work to understand the structure and mechanism of sortase enzymes was conducted on the Staphylococcus aureus SrtA (SaSrtA) enzyme, and this enzyme is typically considered to be the archetypal member of the family. Here we briefly summarize what has been learned about the molecular basis of its function from biochemical and structural studies. The first structure of the enzyme was determined by the Clubb group in 2001^{30} (Fig. 2). Initial analysis of the protein sequence revealed the presence of an N-terminal signal sequence, and transmembrane helix preceding a 181 residue soluble domain containing the absolutely conserved catalytic cysteine residue⁴. The structure of this active soluble domain, SaSrtA Δ 59 was determined by NMR and showcases the now canonical closed 8 stranded beta barrel architecture. This represented the first example of a new protein fold, now defined by the Structural Classification of Proteins (SCOP) project³¹ as a "sortase fold." This core domain architecture is highly conserved across all families of sortase with small alterations present in the various other classes (expanded upon below). The structure is well resolved with the exception of a single loop between strands $\beta 6$ and $\beta 7$. Subsequently, a crystal structure of the same construct was solved at 2.0 Å resolution³². The two structures are very similar with an RMSD of 1.97 over all C α , but there are several slight differences: first, the crystal structure showcases elongated β 7 and β 8 strands, and a short helix in the $\beta 6/\beta 7$ loop in two of the 3 proteins in the asymmetric unit. This loop was unstructured in the solution NMR structure, and the authors cite the lack of a single structure for these residues in the asymmetric unit as additional evidence for its flexibility. The
active site architecture is also similar in both structures with the catalytic histidine, cysteine, and arginine residues positioned together within a deep groove. The conserved TLXTC motif which ends with the catalytic cysteine residue in all sortases largely forms stabilizing interactions with other residues in the protein's core, with the exception of the "X" residue (an Ile for SaSrtA) which extends below the active site cysteine to form a portion of the hydrophobic pocket that surrounds the active site. SaSrtA has additionally been shown to require Ca²⁺ for efficient catalysis, and removal of the ion results in 5 fold lower activity³⁰. NMR chemical shift mapping revealed that the Ca²⁺ ion binds in a pocket between the $\beta 3/\beta 4$ loop and the $\beta 6/\beta 7$ loop, where it is coordinated by the sidechains of E105, E108, D112, E171 and a backbone carbonyl from N114³³. NMR dynamics experiments also indicated that binding of calcium altered the mobility of the $\beta 6/\beta 7$ loop which contacts the active site, replacing fast, ps-ns timescale motions with slower, μ s-ms timescale motions which are typically associated with ligand binding and catalysis.

Substrates for the enzyme are recognized by the presence of a (CWSS) harboring the canonical LP*X*TG sorting signal. Catalysis begins with a nucleophilic attack from the active site cysteine residue on the carbonyl carbon in the peptide bond between the Thr and Gly residues in the sorting signal. This forms a tetrahedral intermediate that quickly collapses to form a semi-stable thioacyl intermediate covalently attaching the substrate to the sortase's active site cysteine^{7,24,34}. The secondary substrate, a pentaglycine cross-bridge peptide on the cell wall precursor lipid II, then performs a nucleophilic attack on the carbonyl carbon in the newly formed thioacyl bond between the sortase and substrate, forming a second tetrahedral intermediate that collapses to form a new peptide bond between the primary substrate and lipid II^{34–36}. This covalent complex is then incorporated into the cell wall via the standard cell wall

synthesis machinery. If the secondary substrate is not available, the reaction can proceed via a secondary, terminal pathway wherein water performs the second nucleophilic attack and cleaves the thioacyl bond between sortase and substrate, hydrolyzing the peptide and releasing it to solvent (**Fig. 2D**). In this latter reaction, the enzyme acts as a protease. On the cell surface this would have the presumably negative effect of cleaving the protein from its membrane anchor where it would be then free to be released from the microbe.

2.4 Classification of Sortases

Members of the sortase-superfamily have been classified into distinct sub-families based on their primary sequences⁷. Many organisms utilize multiple sortases simultaneously to elaborate their surfaces with a diverse array of protein substrates. These enzymes frequently operate non-redundantly with each sortase displaying proteins that contain specific sorting signal motifs. This 5 residue sorting signal is generally similar to the canonical LPXTG motif seen for SaSrtA with variations to the Leu and Gly positions of the recognition motif occurring most frequently among signals for the different classes, while the Pro and Thr residues are largely invariant^{6,37}. Class A and B enzymes are prevalent in bacteria in the Firmicutes phylum. Many class A enzymes are thought to perform a "housekeeping" role in that they are constitutively expressed and generally anchor multiple protein substrates to the cell wall with a variety of functions dependent on the organism. Typically in organisms with only a single sortase, the sortase will be of the class A variety. Most class B enzymes also anchor proteins to the cell wall. In many instances these enzymes anchor proteins involved in iron acquisition and typically exist on a separate, iron regulated operon with their substrate(s). Not all class B enzymes are involved in heme acquisition, and some members of this group are instead responsible for assembling

pili^{38,39}. Class C enzymes are the most unique members of the sortase superfamily. Most members of this class function to construct pili, requiring them to function as pilus polymerases by joining the CWSS of one pilin subunit to a lysine residue in a second pilin subunit. The polymerization of these subunits results in the construction of long (up to 1 μ m) filamentous fibers which are used for adhesion to host cells and have also been implicated in biofilm formation^{7,40,41}. The remaining types of sortases are less well studied but appear to attach proteins to the cell wall like their class A and B counterparts. Many species of *bacilli* encode class D enzymes that attach proteins involved in spore formation to the cell wall. Members of the class E enzymes are predominantly found in Actinobacteria species. They are thought perform a housekeeping role similar to class A enzymes in Firmicutes and have also been shown to anchor proteins involved in the formation of aerial hyphae^{6,42}. Class F enzymes are also present in Actinobacteria, but their function is currently unknown. Here we focus on the structural variations inherent in each class of sortase, as the biological implications of these enzymes and their substrates has been thoroughly reviewed elsewhere^{7,24,25,27,28}.

2.5 Structural Biology of Sortase Enzymes

At present, over 40 structures of sortase enzyme have been reported (**Table 1**). All sortases studied to date are constructed from the same conserved 8 β -stranded sortase fold shown in Fig. 2. However, members from each class exhibit unique variations that may be reflective of their distinct functionalities in the cell. Generally the most significant discrepancies can be localized to 4 distinct structural foci: (*i*) the N-terminal segment of the enzyme between the transmembrane helix and the first β -strand of the barrel, (*ii*) the loop between strands $\beta \delta$ and $\beta 7$ (the $\beta 6/\beta 7$ loop), (*iii*) the loop between strands $\beta 7$ and $\beta 8$ (the $\beta 7/\beta 8$ loop), and (*iv*) the C-

terminus of the enzyme which for some class C sortases harbors an additional transmembrane helix, although this has not been visualized in any reported structure thus far^{20,21} (**Fig. 3**). In many instances these unique structural differences have been shown to influence the enzyme's ability to recognize its specific sorting signal motif, and in others they are postulated to perform a regulatory role by controlling substrate access to the active site. In particular, N-terminal appendages in some enzymes have been shown to modulate sorting signal access to the active site, while large structural variation in the $\beta 6/\beta 7$ and $\beta 7/\beta 8$ loops has been implicated in substrate binding. Below we compare the known structures of sortases, highlighting class specific differences and the functional implications of these differences where known.

Class A enzymes. Beyond the archetypal SaSrtA enzyme from *S. aureus* described in detail above, class A enzyme structures have been reported for *Streptococcus pyogenes*⁴³, *Bacillus anthracis*⁴⁴, *Streptococcus agalactiae*⁴⁵, and *Streptococcus mutans*⁴⁶. All of these enzymes are believed to perform a housekeeping role and recognize the LPXTG sorting signal. The $\beta 6/\beta 7$ loop of these enzymes contains a short helix that in structures of SaSrtA and BaSrtA bound to their cognate sorting signal plays a critical role in signal recognition^{47,48}. Interestingly, members of the class A family differ in the mobility of this key active site loop. In SaSrtA the loop is disordered in the absence of substrate, whereas all other class A enzymes display an ordered $\beta 6/\beta 7$ loop and thus maintain a preformed binding site for the sorting signal. Many of the structures also differ substantially from one another in the surface near the active site histidine residue. This difference was first highlighted in the crystal structure of *S. pyogenes* SrtA (SpySrtA)⁴³. The catalytic core of this enzyme assumes the same canonical 8 stranded β -barrel sortase fold, but unlike SaSrtA, the active site Cys-S is oriented towards the active site His- δ N, at a distance of 5.4 Å (in SaSrtA these groups are pointed away from another and separated by 6.5

Å). Most notably, a unique groove runs between these two residues whose walls are defined by residues in helix H1 and the $\beta 7/\beta 8$ loop, and whose base is defined by residues in the $\beta 4/\beta 5$ loop. This groove provides a potential exit tunnel for the substrate that could accommodate binding of a full length protein bearing amino acids C-terminal of the LPXTG motif: a feature which was not obvious in SaSrtA structures wherein the active site histidine residue is actually occluded from solvent and this channel closed by interactions between residues in the $\beta 7/\beta 8$ loop and helix H1. The structure of the S. agalactiae SrtA (SagSrtA) has also been determined and is most similar to the SpySrtA structure, maintaining the same exit channel between His118 and Cys184 active site residues with its most distinguishing feature being an altered conformation for its $\beta 6/\beta 7$ loop which is likely a byproduct of its crystallization conditions as this loop is involved in coordinating a Zn^{2+} ion in conjunction with a Glu158 from an adjacent protein in the asymmetric unit⁴⁵. The active site His118 residue is also involved in coordinating a Zn^{2+} ion along with His136 from a neighboring SagSrtA in the asymmetric unit, but this interaction does not seem to move the residue into a non-physiological position, as it aligns well with the histidine residues from active sites of other SrtA enzymes. The B. anthracis SrtA (BaSrtA) structure is noteworthy in that it possesses an N-terminal appendage which wraps behind helix H1 and contacts the active site, potentially acting as a cap to prevent access by the secondary substrate, lipid II (Fig. 3A)⁴⁴. The most recently determined sortase structure, that of S. mutans SrtA (SmSrtA), was crystallized with a significant portion of its N-terminus intact⁴⁶. Interestingly, this enzyme crystallized in an apparent dimer in which an extended N-terminal helix interacts with residues in the active site of a symmetry related molecule. The authors suggest that this interaction is likely an artifact of crystallization, but it is interesting to consider that in many recent sortase structures solved with extended N-termini (especially when considering other classes of sortase, discussed

further below), these regions frequently maintain some form of interaction with the active site. It should also be noted that an additional structure of a class A enzyme, SrtA from *Streptococcus pneumoniae* (SpnSrtA), exists in the PDB as a β -strand-swapped dimer, in which the $\beta 6/\beta 7$ loop extends from one monomer into an adjacent monomer in the asymmetric unit placing the $\beta 7$ and $\beta 8$ strands in the β -barrel core of an adjacent molecule in the asymmetric unit. No publications associated with this PDB entry could be found in PubMed, and it is unclear whether this uncharacteristic result is biologically relevant.

Class B enzymes. The class B sortases which are typically involved in iron acquisition are largely similar in structure to the class A sortases discussed previously with the exception of a longer, structured $\beta 6/\beta 7$ loop. The first structure of a class B enzyme was of the S. aureus SrtB (SaSrtB) enzyme, solved by X-ray crystallography in 2004⁴⁹ in the presence of various sulfhydryl modifiers. When compared to the canonical SaSrtA structure, the most obvious differences are in the presence of two additional helices N-terminal of the first β -strand and a much longer loop inserted between strands $\beta 6$ and $\beta 7$ which contains an additional α -helix (Fig. 3B). These structural differences could potentially be important for dictating the class B enzyme's preference for anchoring substrates to buried, uncrosslinked portions of the cell wall^{50,51}. The unmodified structures of SaSrtB and *B. anthracis* SrtB (BaSrtB) were solved later the same year and exhibit remarkably similar structures (C_{α} RMSD = 3.23Å). The main differences between these two enzymes are localized to a slight conformational change in the $\beta 6/\beta 7$ loop, a lack of density for a section of BaSrtB's $\beta 7/\beta 8$ loop, and an altered conformation of the short loop between helices H1 and H2. An active site cysteine mutant of the SrtB enzyme from *Clostridium difficile* (CdSrtB) was also solved earlier this year by X-ray crystallography⁵² and displays an almost identical structure to SaSrtB (RMSD = 1.93 Å for all C_a). Interestingly,

this class B enzyme may bear a more generalized function as a genetic analysis has predicted 7 possible substrates, none of which are iron associated as is typical of other class B enzymes⁵². Beyond these fairly similar class B sortases, the structure of two additional pilin polymerizing class B sortases from S. pyogenes (SpySrtB) and S. pneumoniae (SpnSrtB) (also referred to as SrtG-1) have also been solved^{38,53}. These proteins are extremely unique in that their genetic localization and biochemical analysis would have grouped them with class C pilin polymerizing sortases, but their structures indicate that they are in fact most similar to class B types (Fig. 3). The SpySrtB and SpnSrtB structures have a similar arrangement of secondary structure elements, including the helix containing $\beta 6/\beta 7$ loop, but also bear an additional short β -strand after this conserved helix. This alteration adds an uncommon 9th β -strand to these proteins, but this strand is not inserted into the conserved β -barrel core and instead pairs with a portion of strand $\beta 6$ that extends above the core of the beta barrel. Understanding how these class B sortases fulfill the job of the structurally distinct class C sortases to polymerize pili instead of attaching proteins to the cell wall will likely be an extremely important step towards understanding the structural characteristics that are necessary for these disparate functions.

Class C enzymes. Class C sortase structures are most notable for the presence of a conserved, elongated N-terminal region that occludes the active site, termed "the lid" (**Fig.** 3C)^{7,28,54,55}. This class has also been the recipient of a significant number of structural studies, with a total of 15 structures currently deposited in the PDB. Interestingly, every class C, pilinpolymerizing sortase studied to date contains this lid region, which invariably contains a DP(F/W/Y) motif. In all but 1 of the class C sortases studied to date, the conserved DP(F/W/Y) motif contacts the active site through a hydrogen bond between the Asp residue in the lid and the arginine residue in the active site as well as through a sulfur-aromatic interaction between the

active site Cvs-S and the aromatic sidechain in the conserved lid motif. Because of these key interactions, these two residues have been referred to as "anchors" of the lid^{54,55}. The first class C enzymes to be structurally characterized were SrtC-1 and SrtC-3 from S. pneumoniae (SpnSrtC1 and SpnSrtC3) by Manzano and coworkers in 2008⁵⁴. Upon analyzing the structures the authors noted that the B-factors for the lid region covering the active site were the highest in the structure and suggested that the lid was likely mobile in solution and that this mobility would unmask the active site allowing for recognition of the substrate. Subsequent structures of class C sortases from S. agalactiae^{21,56,57}, A. oris⁵⁸, and S. pneumoniae⁵⁹ furthered this idea with additional structures showing similar patterns of elevated B-factors or stretches of missing electron density in regions flanking the anchor residues in the lid. Although many organisms utilize multiple class C sortases to polymerize pili, systems for which multiple SrtC enzymes have been solved for an individual organism have shown minimal structural variation. These systems typically show only small shifts in the position or amount of density found for the N-terminal extension harboring the lid, with perhaps the most significant deviation from the norm being the addition of a short Cterminal α -helix opposite the active site in the structure of SpnSrtC3 which has not been functionally characterized⁵⁴. Although the structures of most class C sortases show the lid blocking access to the active site, a structure of S. agalactiae SrtC-1 (SagSrtC1) bound to MTSET, a nonspecific sulfhydryl modifier, also maintains the lid in a closed state indicating that the active site cysteine for this enzyme is at least partially accessible even while the lid is closed⁵⁶. Mutations to the lid region have shown no effect on transpeptidation *in vivo*^{20,21}, while similar mutations have instead been shown to decrease enzyme stability⁵⁵ and increase the rate of hydrolysis in $vitro^{21}$ leading to the theory that the lid performs a regulatory function. Interestingly SagSrtC1 has been solved in multiple crystal forms, and one of these structures,

solved in space group C2, showcases the lid in an "open" conformation providing a model for what is presumed to be a necessary lid-opening event expected to precede catalysis (**Fig. 3C**, compare left to right)⁵⁶. The enzyme maintains the typical sortase fold and, excluding the N-terminal extension preceding the β -barrel core, is extremely similar to SagSrtC1 structures previously solved in space groups P2₁2₁2₁ and P3₁2 with an average backbone RMSD of 0.72 Å. However, in space group C2, residues A38-E71 which typically form the flexible lid structure, instead form an extended helical structure with the aromatic lid anchor residue of the conserved DP(F/W/Y) motif, Y51, displaced from the active site by over 30 Å to a position where it stacks against the backbone of helix H2. An additional crystal structure of *S. suis* SrtC-1 (SsSrtC1) also maintains a similar open form of the lid, with the same extended helix replacing what was expected to be a closed lid⁶⁰.

Class D enzymes. Class D sortases are predicted to perform a specialized function in spore formation and have received less attention than classes A-C, with only two structures currently available in the PDB. The first class D structure was determined by NMR for the *B. anthracis* SrtD enzyme (BaSrtD, and also previously referred to as SrtC) in 2012⁶¹. The BaSrtD solution structure is most structurally similar to class A enzymes, as it lacks the lid seen in class C enzymes and the elongated $\beta 6/\beta 7$ loop that is characteristic of class B enzymes; however, unlike SaSrtA, BaSrtD reveals an ordered 3₁₀ helix within the $\beta 6/\beta 7$ loop, indicating a rigid binding pocket for recognition of substrates containing an LPNTA motif (**Fig. 3D**). This solution structure reveals a catalytic core formed by residues 56-198, which adopts the conserved eight, β -stranded sortase fold with two short helices positioned on opposite faces of the protein. The N-terminus of BaSrtD is predicted to contain a single transmembrane helix formed by residues 1-19, followed by an experimentally-determined, unstructured linker region formed by

approximately 30 amino acids preceding the catalytic core. Interestingly, NMR studies of BaSrtD also indicate that the $\beta 2/\beta 3$ and $\beta 4/H1$ surface loops adjacent to the active site histidine are structurally disordered, as residues for these regions are completely unassigned, and the $\beta 7/\beta 8$ loop following and preceding the active site cysteine (Cys173) and arginine (Arg185), respectively, is also poorly ordered. BaSrtD was also shown to form dimers in vitro, and these disordered loops could possibly contribute to a dimerization interface, or this surface surrounding the active site may be responsible for associating BaSrtD with its lipid II substrate. Recently, a 1.99Å resolution crystal structure was solved for *Clostridium perfringens* SrtD (CpSrtD), representing only the second characterized enzyme from this class⁶². The CpSrtD structure contains eight β -strands, adopting the characteristic β -barrel sortase fold with the catalytic triad residues (Cys171, His109, and Arg178) positioned within the enzyme's active site. Interestingly, based on sequence conservation, this enzyme was originally assigned to the class E family of sortases, and CpSrtD contains two alpha helices at its N-terminus, a marked difference from the BaSrtD structure. CpSrtD also exhibits catalytic activity *in vitro*, which is enhanced in a magnesium-dependent manner, making it one of only two sortases (along with SaSrtA) known to be modulated by the presence of metal ions. In further contrast to the BaSrtD enzyme, both the $\beta 2/\beta 3$ and $\beta 4/H1$ loops are well ordered in the CpSrtD crystal structure, with the $\beta 2/\beta 3$ loop adopting a two-turn alpha helix. Furthermore, CpSrtD is thought to exist as a monomer, supported through crystallographic and dynamic light scattering studies. Together, these observations suggest that BaSrtD and CpSrtD could represent distinct subclasses of the sortase D family.

Class E enzymes. In Actinobacteria, class E sortases have been shown to act as housekeeping enzymes, similar to class A enzymes in Firmicutes, and to be involved in the

production of aerial hyphae. These enzymes have been minimally characterized, with the first and only structure for a class E enzyme determined recently by the Clubb group⁶³. Two sortase enzymes from *Streptomyces coelicolor*, ScSrtE1 and ScSrtE2, have been confirmed as bonafide class E enzymes in that they selectively process the unique LAXTG substrate motif *in vitro*⁶⁴. These enzymes are each predicted to contain a single transmembrane domain, formed by residues 139-161 in ScSrtE1 and residues 31-53 in ScSrtE2, followed by an extracellular domain containing the characteristic sortase fold with the conserved catalytic triad (Cys320, His251, Arg329) positioned within the active site. Although most seemingly related to class C enzymes, the ScSrtE1 structure reveals a largely disordered N-terminus. ScSrtE1 also maintains an elongated $\beta 6/\beta 7$ loop with a short, preformed helix, which suggests a rigid substrate-binding pocket.

Class F enzymes. Lastly, class F enzymes are very poorly understood, with unknown function, unknown substrate motif specificity, and no structural characterization to date. Here, we attempt to gain insight to class F enzymes by analyzing five remaining sortase genes from *S. coelicolor* that belong to the class F sortase family⁷. Two of these five enzymes are predicted to contain an N-terminal transmembrane helix based on analysis with TMHMM⁶⁵. Extracellular domains of the remaining three enzymes were identified using sequence alignment with ClustalW2, and homology models of the extracellular domains of these five class F enzymes generated with the Phyre 2 server have low sequence coverage, ranging from 69-83%, indicating their divergence from other sortase classes. These SrtF enzymes are apparently most similar to class D sortases, as the BaSrtD structure was utilized as the top template in all five models. Notable features of these SrtF models include a preformed helix within the $\beta 6/\beta 7$ loop for four out of five enzymes, as well as a disordered N-terminus that is predicted with high confidence.

In addition to structural variations within enzyme classes, some sortases have also been shown to localize to distinct foci on the membrane $^{66-68}$, leading to the hypothesis that sortases may associate with one another on the cell surface⁶⁹. Several sortases have already been shown to form dimeric assemblies *in vitro*^{61,70}, and recently, the functional role of SaSrtA dimerization was investigated *in vivo*⁷¹. Zhu et al. generated a triple alanine mutant of three residues of SaSrtA, which was shown to disrupt dimerization in vivo. S. aureus strains expressing this monomeric SaSrtA showed improved surface display of adhesive proteins, as well as more efficient invasion of host mammalian cells, suggesting that the monomeric form has increased activity in vivo. This is in stark contrast to an earlier study which claimed that SaSrtA was more active when chemically cross-linked to maintain its dimeric form *in vitro*⁷⁰. Interestingly, MD simulations of WT SrtA $\Delta 59$ and the monomeric mutant form of the protein showed that the triple alanine mutation induces a "fixed" $\beta 6/\beta 7$ loop, and the authors proposed that monomer-dimer equilibrium regulates the enzymatic activity of SaSrtA most likely by influencing the conformation of the $\beta 6/\beta 7 \log^{71}$. Determining the prevalence of *in vivo* complex formation for other sortase homodimers and exploring the potential of heterodimeric and higher order assemblies between different classes of sortases and/or other cell surface translocation or cell wall synthesis machinery present on the cell surface is an active area of research.

2.6 Molecular Basis of Substrate Recognition

Sorting signal recognition has become one of the most intriguing areas of recent study for sortase enzymes. Many sortase enzymes have been shown to recognize their substrates with high specificity and have the ability to differentiate between the appropriate sorting motif or their substrates and that of other sortase enzymes *in vitro* and *in vivo*^{47,59,72–76}. This strict recognition

allows the sortases to function non-redundantly to catalyze reactions with specific substrates even when coexpressed in the same cell. In addition to this strict specificity for primary substrates, different species of bacteria utilize distinct cross-bridge peptides within their lipid II cell wall precursor molecule which are typically only recognized as a secondary substrate by the cell-wall associated sortases for that species. The pilin associated sortases on the other hand are required to recognize a 4 residue lysine-containing pilin motif located within their protein substrate. This motif is typically different for different sortases, and is thought to regulate specificity in systems with multiple class C pilin associated sortases^{75,76}. Understanding the principles that guide substrate recognition for sortase enzymes continues to be an extremely difficult area of study. Since sortase enzymes will perform an off target hydrolysis reaction in the absence of an appropriate nucleophile, slowly cleaving their substrate, structural studies of these complexes have required the implementation of modified sortases or substrates to form complexes stable enough for structural studies. To date, 4 structures have been solved of complexes between sortases and their respective substrates or substrate analogs^{32,37,47,48}. To facilitate discussion of the binding interactions that govern recognition of the sorting signal substrate, we henceforth refer to residues in each sorting signal in relation to their distance from the scissile bond, where residues in an L-P-X-T-G sorting signal are referred to as P4-P3-P2-P1-P1', and the binding sites for those residues on the enzyme as S4-S3-S2-S1-S1', respectively.

The first example of a sortase-substrate complex was a crystal structure of an SaSrtA-C193A mutant with an LPETG peptide soaked into the crystal³². While the complex appeared to show the LPETG peptide lying across the surface of the enzyme near the active site, the validity of this complex was quickly called into question as the highly conserved Leu residue at P4 of the substrate and the Thr residue at P1 were both seen pointing into solvent. Given that these

residues are highly conserved across all substrates for this enzyme⁶, and that an *in vitro* analysis with peptide libraries had shown that both of these residues are extremely important for substrate recognition⁷³, it was expected that these residues would be recognized by numerous contacts to the enzyme. Additionally, there was only density for the bound peptide in 1 of the 3 proteins in the asymmetric unit, indicating extremely weak binding. A breakthrough came when the Jung group at UCLA synthesized an LPAT* substrate mimic, where T* is a modified threonine containing a sulfhydryl moiety in place of its carboxyl oxygens. This molecule mimics the native LPXTG substrate, but can form a disulfide bond with the active site cysteine leading to the production of a stable enzyme-substrate complex suitable for structural studies⁷⁷. Subsequent NMR analysis of the SaSrtA-LPAT* complex led to the localization of the substrate binding site⁷⁸, and eventually the structure of a covalent SaSrtA-LPAT* complex⁴⁷ (Fig. 4A). This structure revealed a reorganization of both the $\beta 6/\beta 7$ loop to include the appearance of a short 3_{10} helix, and an opening of the $\beta 7/\beta 8$ loop upon binding the substrate, potentially indicating a regulatory mechanism which would open the loop to expose the active site histidine residue to solvent and allow for binding of the secondary substrate only after the primary substrate had bound. Additional ¹⁵N relaxation analysis revealed that the $\beta 6/\beta 7$ loop, which was highly dynamic in the unbound state, becomes more rigid upon binding of the substrate⁴⁷. In addition to significant changes in the active site of the enzyme, the structure also revealed the major residues that make up the binding pocket for the substrate and showed that the previous peptide-soaked crystal structure had most likely displayed the peptide non-specifically bound. This non-specific binding could have either been due to the low affinity of the interaction $(K_m = 7.33 \text{ mM})^{79}$, or due to crystal packing interactions preventing the reorganization of the $\beta 6/\beta 7$ loop which appears to be necessary to appropriately construct the binding pocket for recognition of the peptide. One

of the most notable features of SaSrtA-LPAT* complex is the large hydrophobic S4 pocket made up of residues V161, V166, V168, and L169 in the reordered $\beta 6/\beta 7$ loop, and I199 on strand $\beta 8$ which binds the P4 Leu residue. The complex also explains the requirement for a Pro residue at position P3, as this is seen forming a kink in the peptide backbone which directs the rest of the peptide towards the active site cysteine. Interestingly, the residue in the P2 position which, based on sequences of known substrates⁶ and biochemical data⁷³ can be any amino acid, is seen pointing into the active site where it makes contact with A118 and I182. This space does not appear to be large enough to accommodate the presence of larger amino acid side chains which was curious given that these are known to be recognized by the enzyme. Additionally, the P1 threonine residue which is conserved in over 95% of all sortase substrates is seen here pointing out of the active site towards solvent, only making a single hydrogen bond to the active site histidine side chain. The positions of the active site residues in this structure are compatible with biochemical data and the placement of the active site arginine residue hydrogen bonding exclusively with the backbone carbonyl of the P2 Ala residue seemed to indicate that this absolutely conserved residue is likely involved in orienting the substrate for catalysis as opposed to activating the active site Cys or acting as a general base in catalysis as had been proposed previously.

A subsequent crystal structure of SaSrtB covalently bound to an NPQT* substrate mimic of its native NPQTN substrate served to confirm several observations from the SaSrtA-LPAT* structure and clarify some of the more ambiguous ones (**Fig. 4B**)³⁷. The P4 residue, an asparagine in this sorting signal, is primarily recognized by contacts to the $\beta 6/\beta 7$ loop, as in the SaSrtA-LPAT* structure, although in a more solvent exposed position, concordant with the more hydrophilic nature of the asparagine sidechain. This is also compatible with biochemical data from the McCafferty group which has shown that replacement of the $\beta 6/\beta 7$ loop of SaSrtA with that of SaSrtB is sufficient for altering the specificity of the resultant chimeric protein to the SaSrtB sorting signal's NPQTN motif⁸⁰. The P3 proline residue also appears to provide the same function as seen in the SaSrtA-LPAT* structure, kinking the peptide to point it towards the catalytic cysteine. There are also some notable differences between this and the SaSrtA-LPAT* structure, one of which being that the $\beta 6/\beta 7$ loop for this enzyme appears to be maintained in a binding-competent "lock and key" conformation since the structure of the complex was almost identical to that of the unbound crystal structure (RMSD = 0.44 Å for all C_{α}). The P2 and P1 residues also appear to be bound in a unique conformation in the SaSrtB-NPQT* complex. The P2 Gln residue rests along the wall of the pocket, pointing out towards solvent, as opposed to pointing into the base of the active site as the P2 Ala residue does in the SaSrtA-LPAT* structure. Interestingly, the P1 Thr residue in the SaSrtB-NPQT* structure is buried in the active site (Thr-in position) where it forms two hydrogen bonds with the active site arginine residue, instead of pointing towards solvent and hydrogen bonding with the active site histidine residue as seen in the SaSrtA-LPAT* structure (Thr-out position).

The third and most recent structure utilizing the same type of disulfide bond forming substrate mimic, that of BaSrtA-LPAT*, bears similarities to both of the previously described structures (**Fig. 4C**)⁴⁸. The substrate mimic peptide in this structure maintains the same general orientation as seen in both previous structures providing further evidence that the $\beta 6/\beta 7$ loop is instrumental in building the S4 site and dictating specificity for the P4 residue, while the P3 Pro residue is necessary for providing a kink in the peptide backbone to orient the P1 and P2 residues toward the active site cysteine. Interestingly, this enzyme substrate complex maintains a Thr-in conformation, reminiscent of that seen in the SaSrtB-NPQT* substrate complex, where the P1

Thr is buried within the binding pocket leading to the active site, but here it does not interact with the active site arginine. The active site arginine instead forms a hydrogen bond with a backbone carbonyl in the peptide P3 Pro residue, similar to what was seen in the SaSrtA-LPAT* structure, albeit in this structure the H-bond was to the adjacent P2 Ala carbonyl. Unlike the flexible $\beta 6/\beta 7$ loop found in the unmodified SaSrtA, the $\beta 6/\beta 7$ loop of BaSrtA was already shown to be structured in the absence of substrate⁴⁴. The structure of the modified complex confirmed the presence of a preformed binding pocket as the backbone atoms for these residues exhibit an RMSD of 0.76 Å. Conversely, the $\beta 7/\beta 8$ loop, which was disordered in the unmodified BaSrtA structure, undergoes a disordered to ordered transition when bound to the substrate mimic which is correlated with a \sim 7 Å displacement of the active site cysteine residue upon binding the peptide. Taken together, these three substrate-mimic bound structures show conclusively that the S4 site is built by residues in the $\beta 6/\beta 7$ loop which is essential in determining the specificity of these enzymes for their P4 residue, and that the highly conserved Pro residue and P3 is necessary to properly orient the residues at positions P2 and P1 for interaction with the active site (Fig. 4D). While both the SaSrtB-NPQT* and SaSrtA-LPAT* structures share an overall similar conformation for the P1 and P2 residues, the interactions seen between the P1 Thr and the catalytic triad are not consistent across the three structures and likely indicate internal flexibility inherent to this portion of the active site for some members of this enzyme family. These inconsistencies could also be caused, or compounded, by potential slight inaccuracies in complexes with these substrate mimics caused by the use of an unnatural disulfide bond in place of the native thioester.

An energy minimized model of the thioacyl intermediate that exists in the SaSrtB-NPQT complex provided additional information about the nature of the catalytic mechanism. This

model (Fig. 5) shows that the highly conserved P1-Thr residue acts through a hydrogen bonding network to stabilize the position of the active site arginine residue so that it is maintained in close proximity to the thioacyl bond. In this position it can act as part of an oxyanion hole, in concert with the backbone amide of E224, immediately C-terminal to the active site cysteine, to stabilize the two high-energy tetrahedral intermediates which occur during catalysis³⁷. This novel "substrate-stabilized oxyanion hole" could serve as a regulatory mechanism to help increase specificity for the substrate and thereby prevent unwanted catalysis, as biochemical analysis shows that the enzyme is intolerant to even conservative mutations (Ser or Val) to the Thr at the P1 position of its substrate³⁷. A similar energy minimized model was produced for the BaSrtA-LPAT complex which displayed an analogous hydrogen bonding interaction between the P1 Thr residue and the active site arginine, along with a similar orientation for the P1 Thr carbonyl in the thioacyl bond that places it in proximity to the arginine guanidino group for stabilization of the negatively charged tetrahedral intermediate⁴⁸. It should be noted however, that the geometry of these two models is not identical, and that the active site arginine in the BaSrtA structure is in a more extended structure that allows it to make additional hydrogen bonding interactions with the backbone carbonyl of the P3 Pro residue, in place of a direct hydrogen bonding interaction with the thioacyl carbonyl oxygen as seen in the SaSrtB-NPQT model.

To rectify the discrepancies between the SaSrtA-LPAT* Thr-out and SaSrtB-NPQT* Thr-in conformations, Jacobitz et. al. performed molecular dynamics (MD) simulations for three thioacyl intermediate systems: SaSrtA-LPAT, SaSrtB-NPQT, and SaSrtB-NPAT³⁷. Through the use of umbrella sampling calculations, the free energy landscape of transitions between the Thr-out and Thr-in states was mapped for each complex, and it was found that SaSrtA-LPAT could adopt both the Thr-in and Thr-out states with equal probability, while for both NPQT and NPAT

substrates, SaSrtB only samples the Thr-in state. Based on these results, it appears as if the Thrin conformation is likely the more evolutionarily conserved state, and that the inherent flexibility of SaSrtA allows for the Thr-out conformation with an LPAT substrate that was captured by Suree et al⁴⁷.

Additional MD studies with SaSrtA have provided additional insight into the recognition process of sortase enzymes. By performing conventional and accelerated MD simulations of both the sorting signal free and bound states, Kappel et al. proposed that sorting signal binding is a mixture of conformational selection and induced fit mechanisms⁸¹. For example, the $\beta 6/\beta 7$ loop appears to follow the conformational selection paradigm: it sampled a range of stable conformations in the apo state, some of which were relatively close to the bound configurations. In contrast, the $\beta 7/\beta 8$ "open" state from the NMR structure was only stable in the presence of a bound sorting signal, suggesting an induced fit mechanism. In addition, analysis of the sorting signal-bound conformations showed that an allosteric network runs throughout the protein, linking the calcium, sorting signal, and proposed lipid-II binding regions to one another. In a complementary study, Moritsugu et al. used the multiscale enhanced sampling method to probe the allosteric effects of the calcium ion and sorting signal⁸². Simulations of each combination of bound states showed that binding of both molecules is required to stabilize the $\beta 6/\beta 7$ and $\beta 7/\beta 8$ loops in conformations observed in the NMR-LPAT* structure. Overall, these simulations point towards a mechanism in which calcium, sorting signal, and potentially lipid II binding are modulated by a dynamic network that includes the $\beta 6/\beta 7$ loop region in SaSrtA.

Other aspects of the SaSrtA recognition and catalytic processes have also been explored by computational studies. Biswas et al. used a hybrid MD and biochemical experimental approach to probe the roles of the conserved sorting signal Leu and Pro residues in substrate binding⁸³. Comparative simulations with LPAT, APAT, and LAAT substrates demonstrated that contacts between the leucine sidechain and SaSrtA help to stabilize the $\beta 6/\beta 7$ loop, whereas the kink that is induced by the proline appears to be essential for recognition. In another study, Tian and Eriksson performed simulations in which His120 and Cys184 were in their zwitterionic and neutral forms⁸⁴. Their results showed that Arg197 adopts distinct conformations based upon the charged state of the protein, which helps to stabilize the catalytically active form. It should be noted that each of these studies was performed with the sorting signal in the Thr-out state. Although the global effects of the Thr-in and Thr-out states on the induced fit/conformational selection process, allosteric networks, and recognition processes are likely similar, subtle differences may exist that influence some of the fine details that resulted from these simulations.

When compared to the recent accumulation of information regarding the binding of the primary sorting signal substrate, there is relatively little known about the location of the binding pocket or interactions involved in recognizing the secondary substrate of sortase enzymes. Progress on this front has been slowed most significantly by the poor solubility and difficulty inherent in obtaining large quantities of the intact lipid II substrate. To date, the most convincing line of evidence stems from NMR chemical shift mapping of the SaSrtA-LPAT* complex with a triglycine peptide titrated into the sample to mimic the pentaglycine cross-bridge present in the *S. aureus* lipid II molecule by Suree et al⁴⁷. Analysis of the chemical shift perturbations resulting from this addition revealed a continuous surface spanning portions of the $\beta 7/\beta 8$ loop, $\beta 4/H2$ loop, and an N-terminal segment of helix H1. While this surface lies near the active site and bound peptide, it is too large to definitively assign a single binding pocket to the secondary substrate. It is also interesting to note that these chemical shift changes were not observed when titrating the triglycine peptide against an unbound ¹⁵N labeled SaSrtA but only occur after the

SaSrtA-LPAT* complex has formed, presumably due to a rearrangement of the $\beta 7/\beta 8$ loop which creates a more open conformation necessary for substrate binding in this enzyme. An X-ray structure of SaSrtB bound to MTSET, a sulfhydryl modifier, with a triglycine peptide soaked into the crystal has also been solved and shows the triglycine peptide bound exclusively to the $\beta 7/\beta 8$ loop⁸⁵. While this is compatible with NMR chemical shift data from the SaSrtA-LPAT* complex, it does not fit the accepted view of the mechanism which would have the incoming nucleophile deprotonated by the active site histidine residue. In this complex the N-terminal amine of the triglycine peptide is 6.4 Å from the active site histidine which is occluded from solvent by a closed $\beta 7/\beta 8$ loop. This is also incompatible with the clear groove extending past the active site histidine and cysteine residues visible in SpySrtA, SagSrtA, and SmSrtA structures that has been hypothesized to be the binding groove for the secondary substrate in these enzymes^{43,45,46}. Confidently discerning the true location of the sortase mechanism.

2.7 The Molecular Mechanism of Sortase Enzymes

The culmination of the past 16 years of research since the discovery of the sortase enzymes has been a relatively complete understanding of the molecular mechanism, compiled from a large number of studies encompassing *in vitro* biochemical data, mutagenesis, X-ray and NMR structures, NMR dynamics, bioinformatics, computational modeling, and MD simulations. While there will still undoubtedly be variations, improvements, and exceptions added to the current understanding of the mechanism, the existing model is backed by a large and ever growing body of evidence. The current model of the molecular mechanism is presented in Figure 6. Kinetic studies suggest that catalysis occurs through a ping-pong mechanism that begins when

the sortase recognizes the CWSS of a membrane anchored protein as it binds in a groove made by residues in the $\beta 6/\beta 7$ loop, strands $\beta 4$, $\beta 7$, $\beta 8$, and closed by the $\beta 2/\beta 3$ and $\beta 3/\beta 4$ loops (Fig. **4D**). Here the sorting signal's L-shaped structure dictated by the highly conserved proline residue at P3 (>90% conserved) will orient residue P4 for recognition in subsite S4 at the $\beta 6/\beta 7$ loop, as the remainder of the peptide is directed deeper into the active site. In order for catalysis to proceed, the enzyme must contain a properly charged active site with cysteine and histidine residues in their thiolate and imidazolium ion forms, respectively. Based on pKa measurements of the SaSrtA^{79,86} and BaSrtA⁴⁴ enzymes, and the similarly slow *in vitro* catalytic rates of all known sortases studied to date, this is expected to comprise <1% of the available population of enzymes at physiological pH. This is also confirmed by structural analysis of SaSrtA wherein none of the residues in the active site are close enough to support the presence of any ion pairing, with a Cys-S-His- δ N distance of 6.5 and 7.6 Å for NMR and crystal structures, respectively. If the cysteine residue is appropriately charged it can perform a nucleophilic attack on the P1 Thr's carbonyl carbon, leading to the formation of the first tetrahedral intermediate, which is likely coordinated by the substrate-stabilized oxyanion hole postulated to exist between the active site arginine residue and a backbone amide in the $\beta 7/\beta 8 \log^{37}$. This intermediate quickly collapses to form a semi-stable thioacyl intermediate between the substrate's P1-Thr and the active site cysteine, while the leaving peptide is protonated by the active site histidine residue. The active site arginine residue may also participate in the reaction at this stage and others by hydrogen bonding with the backbone carbonyls of the bound substrate to maintain its orientation in the active site^{47,48,84}. This interaction is potentially dependent on the position (Thr-in or Thr-out) of the bound P1 Thr residue and also on the particular sortase in question as existing structures of sortase-substrate complexes have depicted multiple binding positions for this residue and

molecular dynamics (MD) studies have indicated that it can sample multiple conformations in some enzymes which could potentially influence the position of the active site arginine residue³⁷. In the next step of the reaction, the secondary substrate enters the active site where its amine group is thought to be deprotonated by the active site histidine residue before performing a nucleophilic attack on the thioacyl bond. This results in a second tetrahedral intermediate which is again thought to be stabilized by the substrate-stabilized oxyanion hole. This second tetrahedral intermediate quickly collapses to form a new peptide bond between the incoming nucleophile and the P1-Thr residue which then exits the active site to be incorporated into the cell wall via the standard cell wall synthesis machinery, or added to additional protein subunits in the case of pilin polymerizing sortases.

2.8 Sortases as Powerful In Vitro Biochemical Reagents

Many of the sortase enzymes have been studied using a variety of different methods *in vitro*. The first demonstration of *in vitro* activity was performed by Ton-That and coworkers using the prototypical SaSrtA enzyme⁸⁷. Their assay was the first of many to utilize FRET based reporters of activity by incorporating donor and quencher fluorophores onto both ends of an LPXTG sorting signal peptide. Cleavage of this peptide, either through hydrolysis, hydroxylaminolysis, or the native transpeptidation reaction, liberates the donor from the quencher and this activity can be detected by an increase in fluorescence. While the initial study demonstrated only cleavage and hydroxylaminolysis, subsequent studies also demonstrated the ability of SaSrtA to catalyze the native transpeptidation reaction *in vitro* using only short peptide substrates where one contains a sorting signal motif, and the other a series of 1 to 5 glycines^{88,89}. While this assay was useful for demonstrating a preference for secondary substrates containing at

least 2 glycines at the N-terminus⁸⁹, and for initial characterization of the active site histidine cysteine and arginine residues^{22,23}, early kinetics analysis was found to be plagued by the inner filter effect in which very high concentrations of substrate will lead to a donor quenching effect even after cleavage of the FRET pair, leading to inaccuracies in kinetic measurements⁹⁰. Development of an HPLC assay and subsequent revision of the kinetic parameters based on fitting the parameters to a ping-pong hydrolytic shunt model to more accurately account for loss of activity due to hydrolysis of the peptide revealed relatively weak binding affinity for the LPETG sorting signal ($K_m = 7.33 \pm 1.01$ mM) and secondary Gly₅ substrate ($K_m = 196 \pm 64 \mu$ M) and low k_{cat} of 0.28 ± 0.02 s^{-1 79}. Although this is a fairly low turnover number, the measured pKas of the active site residues indicate that the majority of the enzyme is actually reverseprotonated, with only a small percentage ($\sim 0.06\%$) of the enzyme actually in the appropriately charged, active form in vitro. Accounting for this low population of appropriately charged enzyme indicates that this subpopulation of the enzyme must actually be extremely active, with a k_{cat} of over 10⁵ M^{-1 79}. It has also been noted that the reaction occurs extremely rapidly (<3 min from expression to surface attachment) in vivo¹⁹ indicating that there are likely other factors involved in the reaction that are not adequately replicated in the *in vitro* systems described to date. This effect could simply be due to the reduction of the search problem from 3 dimensions to 2 by the incorporation of membrane anchors to both sortase and substrate, or the existence of as yet uncharacterized interactions between other cell surface factors that either activate the sortase or modulate the positions of protein and substrate for more efficient catalysis. Kinetic analysis has also been conducted on several other sortase enzymes, but these have all reported even slower rates of *in vitro* catalysis indicating a similar mechanism which is likely plagued by similar shortcomings in all simplified in vitro systems^{43,44,80}. A number of additional studies have

measured the endpoints of reactions at a defined incubation time and lack the comprehensive analysis necessary to extract kinetic parameters. A summary of the results of *in vitro* reactions with sortases performed to date is available in Table 2. The fact that many of these analyses are highly simplified, along with numerous differences in substrate and buffer requirements for the various members of the sortase family have made a comparison of the various kinetic rates of these enzymes difficult.

2.9 Development of Structure-Based Therapeutics

As a validated virulence factor, the sortase superfamily represents an attractive antibiotic target for which inhibitor development is being actively pursued⁹¹. Many classes of sortase inhibitors have been identified with high-throughput screening (HTS) via monitoring FRET-based substrate cleavage *in vitro* or virtual docking of small molecule libraries against sortase structures *in silico*⁹¹. Several sortase inhibitor classes function through covalent modification of the active site cysteine, resulting in irreversible inactivation of the enzyme, and several Cyslinked, inhibitor:sortase complexes have been structurally characterized. Many of these structures display sortase enzymes covalently linked with general thiol modifiers; however, several structures exhibit complexes with novel, sortase-specific compounds that provide insight into active site interactions and potential derivatization strategies. Furthermore, structure-activity relationship (SAR) and molecular docking studies for several lead compounds have facilitated rational design of inhibitors with improved biochemical and pharmacological properties, bringing a novel class of therapeutics for Gram-positive infections closer within reach.

Maresso *et al.* identified AAEK inhibitors using an *in vitro* FRET-based HTS, and subsequently, solved the crystal structures of BaSrtB linked to AAEK1 and AAEK2 inhibitors¹³.

Sortases specifically activate AAEK inhibitors via a β -elimination reaction that generates an olefin intermediate which covalently modifies the active site cysteine. AAEK compounds currently inhibit different class sortases from *staphylococci* and *bacilli* species with IC₅₀s in the low micromolar range (~5-50 μ M). A comparison of the structures of native and inhibitor-bound BaSrtB reveals several key differences, including rotation of the active site cysteine by ~180° to accommodate the ligand, displacement of the conserved arginine away from active site, and higher mobility and different conformations of residues from the H5/ β 6 loop. Interestingly, the aryl group of the thienylpropanone adduct engages in stacking interactions with a critical tyrosine (Tyr138) within the BaSrtB active site. Furthermore, the inhibitor-bound form reveals two available binding sites, a cationic site above and anionic site below the aryl group that could be exploited to improve potency and selectivity of the AAEK inhibitor class.

To identify more potent SrtA inhibitors, Suree *et al.* performed an *in vitro* HTS of a 30,000 compound library that resulted in three promising leads, the rhodanine, pyridazinone, and pyrazolethione classes¹¹. All three lead classes inhibited SaSrtA and BaSrtA reversibly with IC₅₀s in the low to sub-micromolar range. SAR studies of the most promising class, the pyridazinone lead, identified critical regions of the chemical scaffold that are required for sortase inhibition. The binding modes of these inhibitors were modeled using induced-fit docking protocol and the SaSrtA-LPAT* structure. In this model, the pyridazinone phenyl ring is buried in a hydrophobic pocket consisting of side chains from Ile199 on the β 8 strand and Val166 and Val168 from the $\beta 6/\beta 7$ loop of SaSrtA, while the carbonyl oxygen on the pyridazinone ring is positioned towards the active site Arg197 side chain. The inhibitor's thiol group points towards the conserved His120, and mutation with a chloro group significantly reduces activity.

Interestingly, mutating the R1 substituent to an ethoxy moiety is well-tolerated, presumably due to additional hydrophobic contacts with side chains of H1 residues, Pro94 and Ala92.

To improve the success rate of virtual screening against the substrate-bound form of SaSrtA, Chan et al. virtually screened over 55,000 ligands, the top-ranked 500 of which were docked onto the NMR structure of SaSrtA with a relaxed complex scheme to account for protein receptor flexibility¹². This approach identified 24 unique leads, and *in vitro* validation of eight out of 19 compounds yielded a 42.2% success rate with IC₅₀s ranging from 47-368 μ M. Interestingly, the 2-phenyl-2,3-dihydro-1H-perimidine scaffold, which was the most active compound experimentally, did not rank highly during the first round of virtual screening, but ranked substantially higher when docked to the ensemble of molecular dynamics (MD) simulated sortase conformers. This compound appears to exploit regions of the enzyme that have evolved for substrate binding, as computational docking studies suggest that the DHP group within the 2phenyl-2,3-dihydro-1H-perimidine lead scaffold, will bind in the S4 site adjacent to the $\beta 6/\beta 7$ loop. Interestingly, the naphthalene ring moves within the hydrophobic pocket during MD simulations, suggesting that adding non-polar substituents here could improve binding. Finally, the phenyl group of the lead scaffold is positioned towards the conserved catalytic triad residues and the attached carboxyl moiety is wedged between and the active site histidine and arginine residues where it forms hydrogen bonds with both.

Zhulenkovs *et al.* identified benzisothiazolinone-based inhibitors via *in vitro* screening, followed by secondary screening by HSQC and subsequent structural determination of the inhibitor bound to SaSrtA using NMR⁹². Secondary screening of inhibitor binding by HSQC provided unique benefits, including identifying interacting sortase residues, estimating potency based on adduct accumulation, and determining reversibility of binding, and also highlighted the

susceptibility of in vitro screening to yield false positives (11 of 41 compounds showed no binding by HSQC). The experimentally determined NMR structure of the enzyme-inhibitor complex was then used to design analog molecules, which irreversibly inhibited SaSrtA with IC₅₀s in the low micromolar range (\sim 3-7 μ M). The NMR structure indicates various inhibitor conformations, although interacting groups are nearly identical, which is possibly due to the dynamic nature of the $\beta 6/\beta 7$ and $\beta 7/\beta 8$ loops. The high affinity interaction is facilitated by covalent binding of a benzo[d]isothiazol-3(2H)-one heterocycle moiety to the catalytic cysteine side chain; however, positions 4, 5, and 6 of this moiety did not show contacts with SaSrtA, suggesting that derivatization to larger substituents could facilitate additional contacts. Within the structure ensemble, the phenyl group of the inhibitor adduct can interact with L97 and A118 or participate in pi-pi stacking interactions with W194. The linker region preceding the adamantyl moiety was varied to contain one or two carbonyl groups, which may hydrogen bond with the active site arginine. In addition, the length of the linker region could be varied to optimally bury the adamantyl moiety within the enzyme's hydrophobic pocket, where it was experimentally determined to interact with V166, V168, L169 in the $\beta 6/\beta 7$ loop and T180, I182, and I199 from the β 7 and β 8 strands.

Recently, the structure of SmSrtA was solved with the natural product flavonoid precursor, trans-chalcone⁴⁶. The structure depicts chalcone bound outside of the SmSrtA active site; however, this compound was verified biochemically to covalently modify the active site cysteine residue through Michael addition. Wallock-Richards *et al.* generated a model of the Cys-linked SmSrtA-trans-chalcone complex utilizing the crystal structures of SaSrtA-LPETG and BaSrtB-AAEK. The fortuitous interaction of the N-terminus from the symmetry-related molecule within the SmSrtA active site guided overlay of the aromatic rings of chalcone at the

phenyl rings of Phe67 and Phe69, which presumably represents the binding mode of this compound. This binding mode seems to be shared among other sortase inhibitor classes, including methyl(2E)-2,3-bus(4-methoxyphenyl)prop-2-enoate in which correct positioning of two phenyl groups is essential for SrtA inhibition⁹³. In fact, the criteria for two aromatic regions seems to be a common feature of SrtA inhibitors and has been implemented to filter compounds using a pharmacophore model generated from all SrtA inhibitors available in the literature before completing virtual screening and MD simulations⁹⁴.

Structure-based HTS is an attractive approach to identify and optimize sortase inhibitors. The wealth of structural information available for sortase enzymes is evolving virtual screening strategies to account for the dynamic nature of SrtA by implementing docking against multiple structures and models generated through MD simulations^{12,95}. These structure-based screening strategies, in combination with SAR of lead compounds, have produced many promising inhibitor classes with IC₅₀s reaching the low- and sub-micromolar range^{11,15}, as well as informed our understanding of important active site contacts that mediate binding affinity and exploited the knowledge of these contacts for further optimization.

2.10 Conclusion

Sortase enzymes are ubiquitous in Gram-positive organisms and are commonly associated with increased pathogenicity. Beyond their medical relevance, these enzymes have also been shown to be extremely useful tools for site specifically linking proteins and small molecules *in vitro*. Because of this, a considerable effort has been put forth to understand the structural features that dictate substrate and small molecule inhibitor binding, as well as the molecular mechanism of catalysis. Recent studies have uncovered a conserved mechanism of substrate binding, shed light on the importance of variable regions of sortase structure, elucidated the molecular mechanism, and produced a number of small molecule sortase inhibitors. In spite of the many discoveries that have been made in the last decade, many questions remain to be answered. No conclusive evidence has emerged to pinpoint the secondary substrate binding site and very little is known about this latter half of the mechanism. While there are a number of hypotheses, each with varying degrees of experimental backing, there has been no consensus in the field. It is conceivable that this lack of consensus is the result of individual enzymes or specific classes recognizing their secondary substrates in very different ways. This could potentially stem from the needs of certain classes to attach their primary protein substrates to different locations on the cell surface or recognize a secondary protein substrate instead of lipid II. Exploration of this interaction in a number of different enzymes will be necessary to draw useful conclusions about the nature of secondary substrate recognition. A number of enzymes have also been shown to harbor N-terminal extensions which in some cases contact or completely occlude the active site. For class C enzymes it appears that lid removal must be a necessary step in uncovering the active site for catalysis, and while structures of class C enzymes with their lids displaced have emerged recently, it remains to be seen whether these structures are physiologically relevant, what would be responsible for removing the lid *in vivo*, and how this lid removal would take place. Understanding the various ways sortase enzymes utilize these additional features in substrate recognition and catalysis is an area of active research. Finally, although many small molecule sortase inhibitors have been identified, to date none has been successful enough to move forward towards becoming a therapeutic for treating Gram-positive infections in humans. Given the rising prevalence of antibiotic resistant bacteria in recent years,

pressure to develop viable sortase inhibitors as therapeutics is rising and will undoubtedly lead to the discovery and characterization of more potent and specific compounds in the future.

2.11 Figures



Figure 2.1 Biological mechanism of sortase.

Sortase and substrate are both membrane bound. **1** Sortase recognizes a sorting signal motif (here the LPXTG sorting signal for SrtA types is shown) within CWSS and performs a nucleophilic attack. **2** LPXTG sorting signal is cleaved to produce sortase-substrate thioacyl intermediate. Mechanism for cell wall anchoring proceeds through "**a**" steps (top) and mechanism for pilus assembly proceeds through "**b**" steps (bottom). **3a** Cross-bridge peptide from lipid II molecule performs a nucleophilic attack on the thioacyl intermediate. **4a** new peptide bond is formed between the lipid II molecule and surface protein which is then incorporated into mature cell wall. **3b** a lysine residue within the pilin motif performs nucleophilic attack on thioacyl intermediate. **4b** covalently linked pilin proteins can either be attached to cell wall as in **3a** and **4a** or polymerized further by additional rounds of **3b**.



Figure 2.2 Structure of sortase.

(A) SaSrtA NMR structure shown as cartoon, showcasing 8 Stranded β -barrel, with active site residues shown as sticks. (B) SaSrtA NMR structure shown as a surface representation in *green* with active residues Arg in *blue* and Cys in *yellow*. The active site His is occluded by a closed $\beta 7/\beta 8$ loop, and there is no obvious groove for a full length peptide to exit the active site. (C) SpySrtA structure shown as a surface representation, colored as in (B), with active site His residue also shown in *cyan*. An open $\beta 7/\beta 8$ loop creates a clear channel that can be seen running between active Cys and His residues indicating the likely path of binding for a full length peptide substrate. (D) Sortase can catalyze both a reversible transpeptidation reaction, and an irreversible hydrolysis reaction (indicated by dashed arrow).

Organism and Sortase	PDB	Bound Ligands or Substrates	Method
		Class A	
S. aureus SrtA	1IJA	N/A	NMR
S. aureus SrtA-C184A	1T2O	N/A	Xray
S. aureus SrtA	1T2P	N/A	Xray
S. aureus SrtA	1T2W	LPETG	Xray
S. aureus SrtA	2KID	Cbz-LPAT*, Ca ²⁺	NMR
S. aureus SrtA	2MLM	benzo[d]isothiazol-3-one based inhibitor	NMR
B. anthracis SrtA	2KW8	N/A	NMR
B. anthracis SrtA	unpub	Cbz-LPAT*	NMR
S. pyogenes SrtA	3FN5	N/A	Xray
S. pyogenes SrtA	3FN6	Cys in suflphenic acid form	Xray
S. pyogenes SrtA	3FN7	N/A	Xray
S. agalactiae SrtA	3RCC	Zn^{2+}	Xray
S. pneumoniae SrtA	408L	N/A	Xray
S. pneumoniae SrtA-C207A	408T	N/A	Xray
S. mutans SrtA	4TQX	chalcone	Xray
		Class B	
S. aureus SrtB	1NG5	N/A	Xray
S. aureus SrtB	1QWZ	MTSET	Xray
S. aureus SrtB	1QX6	E-64	Xray
S. aureus SrtB	1QXA	Gly ₃	Xray
S. aureus SrtB	4FLD	Cbz-NPQT*	Xray
B. anthracis SrtB	1RZ2	N/A	Xray
B. anthracis SrtB	20QW	AAEK1	Xray
B. anthracis SrtB	20QZ	AAEK2	Xray
S. pyogenes SrtB	3PSQ	Zn^{2+} , Cl^{-}	Xray
C. difficile SrtB	4UX7	N/A	Xray

Table 2.1 Structurally characterized sortase enzymes.

	C	Class C	
A. oris SrtC-1	2XWG	Ca ²⁺	Xray
S. pneumoniae SrtC-1	2W1J	glycerol	X-ray
S. pneumoniae SrtC-3	2W1K	N/A	Xray
S. pneumoniae SrtC-1-H131D	2WTS	alanine	Xray
S. pneumoniae SrtC-2	3G66	N/A	X-ray
S. pneumoniae SrtC-2	3G69	SO4 ²⁻	Xray
S. agalactiae SrtC-1Pilus Island-2a	300P	N/A	Xray
<i>S. agalactiae</i> SrtC-1 Pilus Island-1 "Type III"	3RBI	N/A	Xray
<i>S. agalactiae</i> SrtC1- Pilus Island-1 C184A; KDPYS to IPNTG	3RBJ	N/A	Xray
S. agalactiae SrtC-1 Pilus Island-1 "Type II"	3RBK	N/A	Xray
S. agalactiae SrtC1 Pilus Island-1 "Type I" - open lid	3TB7	N/A	Xray
S. agalactiae SrtC-1 Pilus Island-1	3TBE	MTSET	Xray
S. agalactiae SrtC-2 Pilus Island-1	4G1H	Ca^{2+}	Xray
S. agalactiae SrtC-1 Pilus Island-1	4G1J	N/A	Xray
S. suis SrtC-1	3RE9	N/A	Xray
	C	Class D	
B. anthracis SrtD	2LN7	N/A	NMR
C. perfringens SrtD	4D70	none	Xray
	C	Class E	
S. coelicolor SrtE-1	unpub	N/A	Xray



Figure 2.3 Structural variation by class of sortase.

Sortases representative of the major themes seen for each class are displayed as a cartoon representation, with active site residues shown as sticks. The hallmark sortase beta barrel is highlighted in *blue*. Major sources of structural variability are highlighted: N-terminus, *red*; $\beta 6/\beta 7$ loop, *orange*; $\beta 7/\beta 8$ loop, *green*.


Figure 2.4 Sorting signal recognition.

(A) The SaSrtA-LPAT* complex. (B) The SaSrtB-NPQT* complex. (C) The BaSrtA-LPAT* complex, shown with N-terminal appendage removed from view for clarity. Enzymes are shown as surface representations with SrtA types in *light green* and SaSrtB in *light blue*, Substrate mimics are shown as grey sticks. Active site Cys and Arg residues are shown as *gold* and *blue* surfaces, respectively. (D) Conserved recognition sites for sortase enzymes. Left, SaSrtA shown as a transparent surface representation with recognition subsites determined from the combination of sortase structures color coded as follows: S4 is shown in *red*, S3 in *orange*, S2 in *green*, and S1 in *magenta*, and active site Arg in *blue*, Cys in *gold*, and His in *cyan*. Right, Cartoon diagram of SaSrtA with secondary structure elements that contribute to substrate binding labeled for clarity.



Figure 2.5 The substrate stabilized oxyanion hole.

The energy minimized model of the SaSrtB-NPQT thioacyl intermediate displayed with SaSrtB as *light blue* cartoon and residues in the active site and oxyanion hole shown as sticks. NPQT substrate shown as grey sticks. The substrate's P1 Thr residue's sidechain hydroxyl, and backbone carbonyl participate in a hydrogen bonding network with the active site Arg, and the backbone amide of Glu224 that together build an oxyanion hole to stabilize the high energy tetrahedral reaction intermediates. Reproduced from Jacobitz et al 2012.



Figure 2.6 Molecular mechanism of sortase enzymes.

The active site of sortase consists of a His-Cys-Arg triad, and in its active form, the His and Cys residues will form a thiolate-imidazolium ion pair (a). The first step in the reaction is the recognition of an appropriate sorting signal (here the LPXTG sorting signal for SrtA types is shown), and the active site Cys residue will perform a nucleophilic attack on the carbonyl carbon at the substrate's P1 position (b). This leads to a tetrahedral intermediate whose oxyanion is stabilized by the nearby Arg residue which is likely held in proximity by interactions with the side chain of the substrate's P1 residue which is a threonine in over 95% of all substrates (c). The active His residue concomitantly donates a proton to the leaving group and the tetrahedral transition state then collapses to form a semi-stable thioacyl intermediate between the substrate's P1 residue and the active site Cys (d). Next, the secondary substrate, (Here shown as the lipid II molecule used by cell wall anchoring sortases) enters the active site where its terminal amine is

deprotonated by the active His residue before it performs a nucleophilic attack on the carbonyl carbon in the thioacyl bond (e) forming a second tetrahedral intermediate (f) before collapsing to form a new peptide bond between the two substrates, which is then released to leave the regenerated active site (a).

Sortase	Primary Substrate (Km mM)	Secondary Substrate (K _m μM)	Cleavage k _{cat} (s ⁻¹)	Transpeptidation <i>k</i> _{cat} (s ⁻¹)				
Class A								
SaSrtA _{A24} ⁷⁹	Abz-LPETG-Dap(Dnp) (7.33 ± 1.01)	$\begin{array}{c} Gly_5\\ (196\pm 64) \end{array}$	0.086 ± 0.015	0.28 ± 0.02				
SaSrtA _{A24} 90	Abz-LPETG-Dap(Dnp) (5.5 mM)	Gly5 (140)	NR	0.27				
SaSrtA _{A24} 96	Abz-LPETGG-Dap(Dnp) $(8.76 \pm 0.78)^{a}$	Gly5 (NR)	NR	1.10 ± 0.06 ^a				
SaSrtA _{A59} 97	Abz-LPETGK-(Dnp) (7.6 ± 0.5)	$\begin{array}{c} Gly_3\\ (140\pm 30) \end{array}$	NR	1.5 ± 0.2				
SaSrtA A59 Evolved tetramutant ⁹⁷	Abz-LPETGK-(Dnp) (0.17 ± 0.03)	$\begin{array}{c} Gly_3 \\ (4800\pm700) \end{array}$	NR	4.8 ± 0.8				
SpySrtA _{A81} ⁴³	Abz-LPETGG-Dap(Dnp) (0.83 ± 0.11)	Ala ₂ (NR)	NR	0.0136 ± 0.0011				
BaSrtA ₄₅₆ 44	Abz-LPETG-Dap(Dnp) $(0.038 \pm 4)^{\dagger}$	m-DAP*	$0.0004 \pm 0.0001^{\dagger}$	*				
BaSrtA ₄₅₆ 48	Abz-LPETG-Dap(Dnp) (0.306 ± 0.023)	m-DAP*	$3.6 \pm 0.2 \text{ x} 10^{-5}$	*				
BaSrtA∆64 ⁴⁸	Abz-LPETG-Dap(Dnp) (0.173 ± 0.011)	m-DAP*	$5.7 \pm 0.2 \text{ x } 10^{-5}$	*				
SmutSrtA _{A40} ⁴⁶	Dabcyl-QALPETGEE-Edans $(0.0904 \pm 0.0047)^{\dagger}$	NR	Yes	NR				
	С	lass B						
$SaSrtB_{\Delta 21}^{80}$	Abz-KVENPQTNAGT- Dap(DNP) (7.8 ± 2)	Gly ₅ (NR)	NR	$5.4 \pm 0.5 \text{ x } 10^{-4}$				
SaSrtB _{A31} ³⁷	SNKDKVEN PQTN AGT (1.8)	Gly₅ (NR)	NR	1.010-4				
BaSrtB∆37 ⁹⁸	Abz -KTD NPKTG DEA- Dap(DNP)	NR	Yes	NR				
CdSrtB _{A26} 99	KIVK SPKTG DETQLMK KPPV PPKTG DSTTIGK	Gly _{4/5} or Ala- _D -Glu- DAP	NR	NR				

Table 2.2Activity of sortase enzymes in vitro.

Class C

SpnSrtC1 ₁₇₋₂₂₈ 54	IPQTG in RrgB ₃₀₋₆₃₃	YPKN in RrgB ₃₀₋₆₃₃	No	Yes
SagSrtC1 ₄₃₋₂₅₄ from PI-2a ²¹	Dabcyl-KKVT IPQTG GIGT- Edans (0.0138)	NR	Yes	NR
SagSrtC1 ₄₂₋₃₀₅ from PI-1 ⁵⁷	Dabcyl-RPPGV FPKTG GIG- Edans (0.01358 ± 0.00063)	NR	$1.16 \pm 0.044 \text{ x} 10^{-3}$	NR
	Dabcyl-RPSIPNTGGIG-Edans (0.03100 ± 0.00462)	NR	$1.77 \pm 0.101 \text{ x}10^{-3}$	NR
	Dabcyl-RGGL IPKTG EQQ- Edans (0.01639 ± 0.00250)	NR	$0.77 \pm 0.038 \text{ x}10^{-3}$	NR
SagSrtC2 ₄₂₋₂₈₃ from PI-1 ⁵⁷	Dabcyl-RPPGV FPKTG GIG- Edans (0.006385 ± 0.00142)	NR	$1.04 \pm 0.058 \text{ x} 10^{-3}$	NR
	Dabcyl-RGGL IPKTG EQQ- Edans (0.02733 ± 0.00435)	NR	$4.36 \pm 0.256 \text{ x}10^{-4}$	NR
	Dabcyl-RPSIPNTGGIG-Edans (0.05715 ± 0.00354)	NR	$5.56 \pm 0.174 \text{ x}10^{-3}$	NR
	CI	ass D		
BaSrtD.see ⁶¹	VOGEKLPNTASNN	m-DAP (NR)	Ves	NR

BaSrtD _{∆55} ⁶¹	VQGEK LPNTA SNN	m-DAP (NR)	Yes	NR
BaSrtD ₄₅₅ ¹⁰⁰	Abz-GEKLPNTASNN-Dnp	m-DAP (NR)	Yes	NR
CpSrtD _{A23-187} ⁶²	$A\beta_{1-16}$ -LPQTGS	NR	Yes	NR

Sorting signals for all substrates are highlighted in bold

Errors are reported where published

"Yes" indicates the reaction was performed in vitro but kinetics parameters were not reported

NR - Not Reported

m-DAP - meso-diaminopimelic acid

^a These values calculated assuming a hydrolytic shunt mechanism

[†] Values reported from fluorescence assay and subject to inner filter effect and are likely underestimates of true parameters

* The enzyme reportedly does not perform this reaction in vitro

59

2.12 References

- Liew PX, Wang CLC, Wong S-L. Functional characterization and localization of a Bacillus subtilis sortase and its substrate and use of this sortase system to covalently anchor a heterologous protein to the B. subtilis cell wall for surface display. *J Bacteriol*. 2012;194(1):161-175. doi:10.1128/JB.05711-11.
- Paterson GK, Mitchell TJ. The biology of Gram-positive sortase enzymes. *Trends Microbiol.* 2004;12(2):89-95. doi:10.1016/j.tim.2003.12.007.
- Desvaux M, Dumas E, Chafsey I, Hébraud M. Protein cell surface display in Grampositive bacteria: from single protein to macromolecular protein structure. *FEMS Microbiol Lett.* 2006;256(1):1-15. doi:10.1111/j.1574-6968.2006.00122.x.
- Mazmanian SK, Liu G, Ton-That H, Schneewind O. Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall. *Science*. 1999;285(5428):760-763. http://www.ncbi.nlm.nih.gov/pubmed/10427003.
- Pallen MJ, Lam AC, Antonio M, Dunbar K. An embarrassment of sortases a richness of substrates? *Trends Microbiol*. 2001;9(3):97-101. doi:10.1016/S0966-842X(01)01956-4.
- Comfort D, Clubb RT. A Comparative Genome Analysis Identifies Distinct Sorting Pathways in Gram-Positive Bacteria. *Infect Immun*. 2004;72(5):2710-2722. doi:10.1128/IAI.72.5.2710-2722.2004.
- Spirig T, Weiner EM, Clubb RT. Sortase enzymes in Gram-positive bacteria. *Mol Microbiol.* 2011;82(5):1044-1059. doi:10.1111/j.1365-2958.2011.07887.x.

- Suree N, Jung ME, Clubb RT. Recent advances towards new anti-infective agents that inhibit cell surface protein anchoring in Staphylococcus aureus and other gram-positive pathogens. *Mini Rev Med Chem*. 2007;7(10):991-1000. doi:10.2174/138955707782110097.
- Mazmanian SK, Liu G, Jensen ER, Lenoy E, Schneewind O. Staphylococcus aureus sortase mutants defective in the display of surface proteins and in the pathogenesis of animal infections. *Proc Natl Acad Sci U S A*. 2000;97(10):5510-5515. doi:10.1073/pnas.080520697.
- Weiss WJ, Lenoy E, Murphy T, et al. Effect of srtA and srtB gene expression on the virulence of Staphylococcus aureus in animal models of infection. *J Antimicrob Chemother*. 2004;53(3):480-486. doi:10.1093/jac/dkh078.
- Suree N, Yi SW, Thieu W, et al. Discovery and structure-activity relationship analysis of Staphylococcus aureus sortase A inhibitors. *Bioorg Med Chem*. 2009;17(20):7174-7185. doi:10.1016/j.bmc.2009.08.067.
- Chan AH, Wereszczynski J, Amer BR, et al. Discovery of Staphylococcus aureus sortase A inhibitors using virtual screening and the relaxed complex scheme. *Chem Biol Drug Des*. 2013;82(4):418-428. doi:10.1111/cbdd.12167.
- Maresso AW, Wu R, Kern JW, et al. Activation of inhibitors by sortase triggers irreversible modification of the active site. *J Biol Chem*. 2007;282(32):23129-23139. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3366505&tool=pmcentrez&re ndertype=abstract. Accessed July 27, 2015.

- Maresso AW, Schneewind O. Sortase as a target of anti-infective therapy. *Pharmacol Rev.* 2008;60(1):128-141. doi:10.1124/pr.107.07110.
- Zhang J, Liu H, Zhu K, et al. Antiinfective therapy with a small molecule inhibitor of Staphylococcus aureus sortase. *Proc Natl Acad Sci U S A*. 2014;111(37):13517-13522. doi:10.1073/pnas.1408601111.
- Proft T. Sortase-mediated protein ligation: an emerging biotechnology tool for protein modification and immobilisation. *Biotechnol Lett.* 2010;32(1):1-10. doi:10.1007/s10529-009-0116-0.
- Tsukiji S, Nagamune T. Sortase-Mediated Ligation: A Gift from Gram-Positive Bacteria to Protein Engineering. *ChemBioChem.* 2009;10(5):787-798. doi:10.1002/cbic.200800724.
- Mao H, Hart SA, Schink A, Pollok BA. Sortase-mediated protein ligation: a new method for protein engineering. *J Am Chem Soc.* 2004;126(9):2670-2671. http://www.ncbi.nlm.nih.gov/pubmed/14995162. Accessed July 1, 2015.
- Schneewind O, Model P, Fischetti VA. Sorting of protein A to the staphylococcal cell wall. *Cell*. 1992;70(2):267-281. http://www.ncbi.nlm.nih.gov/pubmed/1638631.
- Wu C, Mishra A, Reardon ME, et al. Structural determinants of Actinomyces sortase SrtC2 required for membrane localization and assembly of type 2 fimbriae for interbacterial coaggregation and oral biofilm formation. *J Bacteriol*. 2012;194(10):2531-2539. doi:10.1128/JB.00093-12.

- Cozzi R, Malito E, Nuccitelli A, et al. Structure analysis and site-directed mutagenesis of defined key residues and motives for pilus-related sortase C1 in group B Streptococcus. *FASEB J.* 2011;25(6):1874-1886. doi:10.1096/fj.10-174797.
- Ton-That H, Mazmanian SK, Alksne L, Schneewind O. Anchoring of surface proteins to the cell wall of Staphylococcus aureus. Cysteine 184 and histidine 120 of sortase form a thiolate-imidazolium ion pair for catalysis. *J Biol Chem*. 2002;277(9):7447-7452. doi:10.1074/jbc.M109945200.
- Marraffini LA, Ton-That H, Zong Y, Narayana SVL, Schneewind O. Anchoring of surface proteins to the cell wall of Staphylococcus aureus. A conserved arginine residue is required for efficient catalysis of sortase A. *J Biol Chem*. 2004;279(36):37763-37770. doi:10.1074/jbc.M405282200.
- Clancy KW, Melvin JA, McCafferty DG. Sortase transpeptidases: insights into mechanism, substrate specificity, and inhibition. *Biopolymers*. 2010;94(4):385-396. doi:10.1002/bip.21472.
- Navarre WW, Schneewind O. Surface Proteins of Gram-Positive Bacteria and Mechanisms of Their Targeting to the Cell Wall Envelope. *Microbiol Mol Biol Rev.* 1999;63(1):174-229. http://mmbr.asm.org/content/63/1/174. Accessed January 10, 2013.
- Mazmanian SK, Ton-That H, Schneewind O. Sortase-catalysed anchoring of surface proteins to the cell wall of Staphylococcus aureus. *Mol Microbiol*. 2001;40(5):1049-1057. http://www.ncbi.nlm.nih.gov/pubmed/11401711.

- 27. Telford JL, Barocchi MA, Margarit I, Rappuoli R, Grandi G. Pili in Gram-positive pathogens. *Nat Rev Microbiol*. 2006;4(7):509-519. doi:10.1038/nrmicro1443.
- Mandlik A, Swierczynski A, Das A, Ton-That H. Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. *Trends Microbiol*. 2008;16(1):33-40. doi:10.1016/j.tim.2007.10.010.
- Danne C, Dramsi S. Pili of gram-positive bacteria: roles in host colonization. *Res Microbiol.* 163(9-10):645-658. doi:10.1016/j.resmic.2012.10.012.
- Ilangovan U, Ton-That H, Iwahara J, Schneewind O, Clubb RT. Structure of sortase, the transpeptidase that anchors proteins to the cell wall of Staphylococcus aureus. *Proc Natl Acad Sci.* 2001;98(11):6056-6061. doi:10.1073/pnas.101064198.
- Andreeva A, Howorth D, Chothia C, Kulesha E, Murzin AG. SCOP2 prototype: a new approach to protein structure mining. *Nucleic Acids Res*. 2014;42(Database issue):D310-D314. doi:10.1093/nar/gkt1242.
- Zong Y, Bice TW, Ton-That H, Schneewind O, Narayana SVL. Crystal structures of Staphylococcus aureus sortase A and its substrate complex. *J Biol Chem*. 2004;279(30):31383-31389. doi:10.1074/jbc.M401374200.
- Naik MT, Suree N, Ilangovan U, et al. Staphylococcus aureus Sortase A transpeptidase.
 Calcium promotes sorting signal binding by altering the mobility and structure of an active site loop. *J Biol Chem.* 2006;281(3):1817-1826. doi:10.1074/jbc.M506123200.

- Schneewind O, Fowler A, Faull KF. Structure of the cell wall anchor of surface proteins in Staphylococcus aureus. *Science*. 1995;268(5207):103-106. http://www.ncbi.nlm.nih.gov/pubmed/7701329.
- 35. Perry AM, Ton-That H, Mazmanian SK, Schneewind O. Anchoring of surface proteins to the cell wall of Staphylococcus aureus. III. Lipid II is an *in vivo* peptidoglycan substrate for sortase-catalyzed surface protein anchoring. *J Biol Chem*. 2002;277(18):16241-16248. doi:10.1074/jbc.M109194200.
- 36. Ton-That H, Faull KF, Schneewind O. Anchor structure of staphylococcal surface proteins. A branched peptide that links the carboxyl terminus of proteins to the cell wall. J Biol Chem. 1997;272(35):22285-22292. http://www.ncbi.nlm.nih.gov/pubmed/9268378.
- Jacobitz AW, Wereszczynski J, Yi SW, et al. Structural and computational studies of the Staphylococcus aureus Sortase B-substrate complex reveal a substrate-stabilized oxyanion hole. *J Biol Chem.* February 2014:jbc.M113.509273. doi:10.1074/jbc.M113.509273.
- Kang HJ, Coulibaly F, Proft T, Baker EN. Crystal Structure of Spy0129, a Streptococcus pyogenes Class B Sortase Involved in Pilus Assembly. *PLoS One*. 2011;6(1):e15969. doi:10.1371/journal.pone.0015969.
- Lazzarin M, Cozzi R, Malito E, et al. Noncanonical sortase-mediated assembly of pilus type 2b in group B Streptococcus. *FASEB J.* July 2015. doi:10.1096/fj.15-272500.
- Ton-That H, Schneewind O. Assembly of pili in Gram-positive bacteria. *Trends Microbiol.* 2004;12(5):228-234. doi:10.1016/j.tim.2004.03.004.

- 41. Ton-That H, Schneewind O. Assembly of pili on the surface of Corynebacterium diphtheriae. *Mol Microbiol*. 2003;50(4):1429-1438.
 http://www.ncbi.nlm.nih.gov/pubmed/14622427. Accessed June 4, 2015.
- 42. Elliot MA, Karoonuthaisiri N, Huang J, et al. The chaplins: a family of hydrophobic cellsurface proteins involved in aerial mycelium formation in Streptomyces coelicolor. *Genes Dev*. 2003;17(14):1727-1740. doi:10.1101/gad.264403.
- Race PR, Bentley ML, Melvin JA, et al. Crystal Structure of Streptococcus pyogenes Sortase A. *J Biol Chem.* 2009;284(11):6924-6933. doi:10.1074/jbc.M805406200.
- Weiner EM, Robson S, Marohn M, Clubb RT. The Sortase A enzyme that attaches proteins to the cell wall of Bacillus anthracis contains an unusual active site architecture. *J Biol Chem.* 2010;285(30):23433-23443. doi:10.1074/jbc.M110.135434.
- Khare B, Krishnan V, Rajashankar KR, et al. Structural differences between the Streptococcus agalactiae housekeeping and pilus-specific sortases: SrtA and SrtC1. *PLoS One*. 2011;6(8):e22995. doi:10.1371/journal.pone.0022995.
- Wallock-Richards DJ, Marles-Wright J, Clarke D, et al. Molecular Basis of Streptococcus mutans Sortase A Inhibition by the Flavonoid Natural Product trans-Chalcone. *Chem Commun.* June 2015. doi:10.1039/C5CC01816A.
- 47. Suree N, Liew CK, Villareal VA, et al. The Structure of the Staphylococcus aureusSortase-Substrate Complex Reveals How the Universally Conserved LPXTG Sorting

Signal Is Recognized. *J Biol Chem*. 2009;284(36):24465-24477. doi:10.1074/jbc.M109.022624.

- 48. Chan AH, Yi SW, Terwilliger AL, Maresso AW, Jung ME, Clubb RT. Structure of the Bacillus anthracis Sortase A Enzyme Bound to Its Sorting Signal: a Flexible Aminoterminal Appendage Modulates Substrate Access. *unpublished*. 2015.
- Zong Y, Mazmanian SK, Schneewind O, Narayana SV. The Structure of Sortase B, a Cysteine Transpeptidase that Tethers Surface Protein to the Staphylococcus aureus Cell Wall. *Structure*. 2004;12(1):105-112. doi:10.1016/j.str.2003.11.021.
- Mazmanian SK, Skaar EP, Gaspar AH, et al. Passage of Heme-Iron Across the Envelope of Staphylococcus aureus. *Science (80-)*. 2003;299(5608):906-909. doi:10.1126/science.1081147.
- Marraffini LA, Schneewind O. Anchor Structure of Staphylococcal Surface Proteins V. ANCHOR STRUCTURE OF THE SORTASE B SUBSTRATE IsdC. *J Biol Chem*. 2005;280(16):16263-16271. doi:10.1074/jbc.M500071200.
- Chambers CJ, Roberts AK, Shone CC, Acharya KR. Structure and function of a Clostridium difficile sortase enzyme. *Sci Rep.* 2015;5:9449. doi:10.1038/srep09449.
- Shaik MM, Lombardi C, Maragno Trindade D, et al. A structural snapshot of type II pilus formation in Streptococcus pneumoniae. *J Biol Chem.* July 2015:M115.647834 - . doi:10.1074/jbc.M115.647834.

- Manzano C, Contreras-Martel C, El Mortaji L, et al. Sortase-mediated pilus fiber biogenesis in Streptococcus pneumoniae. *Struct (London, Engl 1993)*. 2008;16(12):1838-1848. doi:10.1016/j.str.2008.10.007.
- Manzano C, Izoré T, Job V, Di Guilmi AM, Dessen A. Sortase activity is controlled by a flexible lid in the pilus biogenesis mechanism of gram-positive pathogens. *Biochemistry*. 2009;48(44):10549-10557. doi:10.1021/bi901261y.
- 56. Khare B, Fu Z-Q, Huang I-H, Ton-That H, Narayana SVL. The crystal structure analysis of group B Streptococcus sortase C1: a model for the "lid" movement upon substrate binding. *J Mol Biol*. 2011;414(4):563-577. doi:10.1016/j.jmb.2011.10.017.
- Cozzi R, Prigozhin D, Rosini R, et al. Structural Basis for Group B Streptococcus Pilus 1 Sortases C Regulation and Specificity. *PLoS One*. 2012;7(11):e49048. doi:10.1371/journal.pone.0049048.
- Persson K. Structure of the sortase AcSrtC-1 from Actinomyces oris. *Acta Crystallogr D Biol Crystallogr*. 2011;67(Pt 3):212-217. doi:10.1107/S0907444911004215.
- Neiers F, Madhurantakam C, Fälker S, et al. Two crystal structures of pneumococcal pilus sortase C provide novel insights into catalysis and substrate specificity. *J Mol Biol*. 2009;393(3):704-716. doi:10.1016/j.jmb.2009.08.058.
- 60. Lu G, Qi J, Gao F, Yan J, Tang J, Gao GF. A novel "open-form" structure of sortaseC from Streptococcus suis. *Proteins*. 2011;79(9):2764-2769. doi:10.1002/prot.23093.

- Robson SA, Jacobitz AW, Clubb RT, Phillips ML, Clubb RT. Solution Structure of the Sortase Required for Efficient Production of Infectious Bacillus anthracis Spores. *Biochemistry*. 2012;51(40):7953-7963. doi:10.1021/bi300867t.
- Suryadinata R, Seabrook SA, Adams TE, Nuttall SD, Peat TS. Structural and biochemical analyses of a Clostridium perfringens sortase D transpeptidase. *Acta Crystallogr D Biol Crystallogr.* 2015;71(Pt 7):1505-1513. doi:10.1107/S1399004715009219.
- 63. Kattke M, Jacobitz AW, Clubb RT. The Structure of the SrtE1 Sortase from Streptomyces coelicolor. *unpublished*. 2015.
- 64. Duong A, Capstick DS, Di Berardo C, et al. Aerial development in Streptomyces coelicolor requires sortase activity. *Mol Microbiol*. 2012;83(5):992-1005. doi:10.1111/j.1365-2958.2012.07983.x.
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 2001;305(3):567-580. doi:10.1006/jmbi.2000.4315.
- 66. Raz A, Fischetti VA. Sortase A localizes to distinct foci on the Streptococcus pyogenes membrane. *Proc Natl Acad Sci U S A*. 2008;105(47):18549-18554. doi:10.1073/pnas.0808301105.
- 67. Hu P, Bian Z, Fan M, Huang M, Zhang P. Sec translocase and sortase A are colocalised in a locus in the cytoplasmic membrane of Streptococcus mutans. *Arch Oral Biol.* 2008;53(2):150-154. doi:10.1016/j.archoralbio.2007.08.008.

- Kline KA, Kau AL, Chen SL, et al. Mechanism for sortase localization and the role of sortase localization in efficient pilus assembly in Enterococcus faecalis. *J Bacteriol*. 2009;191(10):3237-3247. doi:10.1128/JB.01837-08.
- Mandlik A, Das A, Ton-That H. The molecular switch that activates the cell wall anchoring step of pilus assembly in gram-positive bacteria. *Proc Natl Acad Sci U S A*. 2008;105(37):14147-14152. doi:10.1073/pnas.0806350105.
- Lu C, Zhu J, Wang Y, et al. Staphylococcus aureus sortase A exists as a dimeric protein *in vitro*. *Biochemistry*. 2007;46(32):9346-9354. doi:10.1021/bi700519w.
- Zhu J, Xiang L, Jiang F, Zhang ZJ. Equilibrium of sortase A dimerization on Staphylococcus aureus cell surface mediates its cell wall sorting activity. *Exp Biol Med* (*Maywood*). June 2015. doi:10.1177/1535370215592122.
- Donahue EH, Dawson LF, Valiente E, et al. Clostridium difficile has a single sortase,
 SrtB, that can be inhibited by small-molecule inhibitors. *BMC Microbiol*. 2014;14(1):219.
 doi:10.1186/s12866-014-0219-1.
- Kruger RG, Otvos B, Frankel BA, Bentley M, Dostal P, McCafferty DG. Analysis of the Substrate Specificity of the Staphylococcus aureus Sortase Transpeptidase SrtA[†].
 Biochemistry. 2004;43(6):1541-1551. doi:10.1021/bi035920j.
- Fälker S, Nelson AL, Morfeldt E, et al. Sortase-mediated assembly and surface topology of adhesive pneumococcal pili. *Mol Microbiol*. 2008;70(3):595-607. doi:10.1111/j.1365-2958.2008.06396.x.

- El Mortaji L, Terrasse R, Dessen A, Vernet T, Di Guilmi AM. Stability and Assembly of Pilus Subunits of Streptococcus pneumoniae. *J Biol Chem.* 2010;285(16):12405-12415. doi:10.1074/jbc.M109.082776.
- El Mortaji L, Fenel D, Vernet T, Di Guilmi AM. Association of RrgA and RrgC into the Streptococcus pneumoniae pilus by sortases C-2 and C-3. *Biochemistry*. 2012;51(1):342-352. doi:10.1021/bi201591n.
- Jung ME, Clemens JJ, Suree N, et al. Synthesis of (2R,3S) 3-amino-4-mercapto-2-butanol, a threonine analogue for covalent inhibition of sortases. *Bioorg Med Chem Lett*. 2005;15(22):5076-5079. doi:10.1016/j.bmcl.2005.07.073.
- Liew CK, Smith BT, Pilpa R, et al. Localization and mutagenesis of the sorting signal binding site on sortase A from Staphylococcus aureus. *FEBS Lett.* 2004;571(1–3):221-226. doi:10.1016/j.febslet.2004.06.070.
- 79. Frankel BA, Kruger RG, Robinson DE, Kelleher NL, McCafferty DG. Staphylococcus aureus sortase transpeptidase SrtA: insight into the kinetic mechanism and evidence for a reverse protonation catalytic mechanism. *Biochemistry*. 2005;44(33):11188-11200. doi:10.1021/bi050141j.
- Bentley ML, Gaweska H, Kielec JM, McCafferty DG. Engineering the substrate specificity of Staphylococcus aureus Sortase A. The beta6/beta7 loop from SrtB confers NPQTN recognition to SrtA. *J Biol Chem.* 2007;282(9):6571-6581. doi:10.1074/jbc.M610519200.

- Kappel K, Wereszczynski J, Clubb RT, McCammon JA. The binding mechanism, multiple binding modes, and allosteric regulation of Staphylococcus aureus Sortase A probed by molecular dynamics simulations. *Protein Sci.* 2012;21(12):1858-1871. doi:10.1002/pro.2168.
- Moritsugu K, Terada T, Kidera A. Disorder-to-order transition of an intrinsically disordered region of sortase revealed by multiscale enhanced sampling. *J Am Chem Soc.* 2012;134(16):7094-7101. http://dx.doi.org/10.1021/ja3008402. Accessed July 21, 2015.
- 83. Biswas T, Pawale VS, Choudhury D, Roy RP. Sorting of LPXTG peptides by archetypal sortase A: role of invariant substrate residues in modulating the enzyme dynamics and conformational signature of a productive substrate. *Biochemistry*. 2014;53(15):2515-2524. http://dx.doi.org/10.1021/bi4016023. Accessed July 21, 2015.
- 84. Tian B-X, Eriksson LA. Catalytic Mechanism and Roles of Arg197 and Thr183 in the Staphylococcus aureus Sortase A Enzyme. *J Phys Chem B*. 2011;115(44):13003-13011. doi:10.1021/jp2058113.
- 85. Zong Y, Mazmanian SK, Schneewind O, Narayana SVL. The Structure of Sortase B, a Cysteine Transpeptidase that Tethers Surface Protein to the< i> Staphylococcus aureus</i> Cell Wall. *Structure*. 2004;12(1):105-112.
- Connolly KM, Smith BT, Pilpa R, Ilangovan U, Jung ME, Clubb RT. Sortase from Staphylococcus aureus Does Not Contain a Thiolate-Imidazolium Ion Pair in Its Active Site. *J Biol Chem.* 2003;278(36):34061-34065. doi:10.1074/jbc.M305245200.

- 87. Ton-That H, Liu G, Mazmanian SK, Faull KF, Schneewind O. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif. *Proc Natl Acad Sci U S A*. 1999;96(22):12424-12429. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=22937&tool=pmcentrez&rend ertype=abstract. Accessed June 10, 2015.
- Ton-That, H., Mazmanian, S. K., Alksne, L., and Schneewind O. Anchoring of Surface Proteins to the Cell Wall of Staphylococcus aureus. SORTASE CATALYZED *IN VITRO* TRANSPEPTIDATION REACTION USING LPXTG PEPTIDE AND NH2-GLY3 SUBSTRATES. *J Biol Chem.* 2000;275(13):9876-9881. doi:10.1074/jbc.275.13.9876.
- Huang X, Aulabaugh A, Ding W, et al. Kinetic mechanism of Staphylococcus aureus sortase SrtA. *Biochemistry*. 2003;42(38):11307-11315. doi:10.1021/bi034391g.
- 90. Kruger RG, Dostal P, McCafferty DG. Development of a high-performance liquid chromatography assay and revision of kinetic parameters for the Staphylococcus aureus sortase transpeptidase SrtA. *Anal Biochem*. 2004;326(1):42-48. doi:10.1016/j.ab.2003.10.023.
- 91. Cascioferro S, Totsika M, Schillaci D. Sortase A: an ideal target for anti-virulence drug development. *Microb Pathog*. 2014;77:105-112. doi:10.1016/j.micpath.2014.10.007.
- 92. Zhulenkovs D, Rudevica Z, Jaudzems K, Turks M, Leonchiks A. Discovery and structureactivity relationship studies of irreversible benzisothiazolinone-based inhibitors against

Staphylococcus aureus sortase A transpeptidase. *Bioorg Med Chem*. 2014;22(21):5988-6003. doi:10.1016/j.bmc.2014.09.011.

- 93. Oh K-B, Kim S-H, Lee J, Cho W-J, Lee T, Kim S. Discovery of diarylacrylonitriles as a novel series of small molecule sortase A inhibitors. *J Med Chem*. 2004;47(10):2418-2421. doi:10.1021/jm0498708.
- 94. Raj KK, Ganesh Kumar V, Leela Madhuri C, et al. Designing of potential inhibitors against Staphylococcus aureus sortase A: Combined analogue and structure based approach with *in vitro* validation. *J Mol Graph Model*. 2015;60:89-97. doi:10.1016/j.jmgm.2015.05.009.
- 95. Uddin R, Saeed K. An exhaustive yet simple virtual screening campaign against Sortase A from multiple drug resistant Staphylococcus aureus. *Mol Biol Rep.* 2014;41(8):5167-5175. doi:10.1007/s11033-014-3384-2.
- Bentley ML, Lamb EC, McCafferty DG. Mutagenesis Studies of Substrate Recognition and Catalysis in the Sortase A Transpeptidase from Staphylococcus aureus. *J Biol Chem*. 2008;283(21):14762-14771. doi:10.1074/jbc.M800974200.
- 97. Chen I, Dorr BM, Liu DR. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc Natl Acad Sci U S A*. 2011;108(28):11399-11404. doi:10.1073/pnas.1101046108.

- 98. Maresso AW, Chapa TJ, Schneewind O. Surface protein IsdC and sortase B are required for heme-iron scavenging of Bacillus anthracis. *J Bacteriol*. 2006;188(23):8145-8152. doi:10.1128/JB.01011-06.
- 99. Van Leeuwen HC, Klychnikov OI, Menks MAC, Kuijper EJ, Drijfhout JW, Hensbergen PJ. Clostridium difficile sortase recognizes a (S/P)PXTG sequence motif and can accommodate diaminopimelic acid as a substrate for transpeptidation. *FEBS Lett*. 2014;588(23):4325-4333. doi:10.1016/j.febslet.2014.09.041.
- 100. Marraffini L a, Schneewind O. Targeting proteins to the cell wall of sporulating Bacillus anthracis. *Mol Microbiol*. 2006;62(5):1402-1417. doi:10.1111/j.1365-2958.2006.05469.x.

Chapter 3

Solution Structure of the Sortase Required for Efficient Production

of Infectious Bacillus anthracis Spores

The work described in this chapter has been reproduced with permission from:

Robson, S. A., Jacobitz, A. W., Clubb, R. T., Phillips, M. L., Clubb, R. T. Solution Structure of the Sortase Required for Efficient Production of Infectious Bacillus anthracis Spores. *Biochemistry*. 2012 51(40):7953-7963.

doi:10.1021/bi300867t.http://pubs.acs.org/doi/abs/10.1021/bi300867t

Copyright 2011 American Chemical

Society

I contributed *in vitro* mass spectrometry experiments and analysis, as well as comparative structural and genetic analysis between the sortase structure presented here and existing sortase

enzymes.

3.1 Overview

Bacillus anthracis forms metabolically dormant endospores that upon germination can cause lethal anthrax disease in humans. Efficient sporulation requires the activity of the SrtC sortase (BaSrtC), a cysteine transpeptidase that covalently attaches the BasH and BasI proteins to the peptidoglycan of the forespore and predivisional cell, respectively. To gain insight into the molecular basis of protein display, we used nuclear magnetic resonance to determine the structure and backbone dynamics of the catalytic domain of BaSrtC (residues Ser56-Lys198). The backbone and heavy atom coordinates of structurally ordered amino acids have coordinate precision of 0.42 \pm 0.07 and 0.82 \pm 0.05 Å, respectively. BaSrtC_{A55} adopts an eight-stranded β barrel fold that contains two short helices positioned on opposite sides of the protein. Surprisingly, the protein dimerizes and contains an extensive, structurally disordered surface that is positioned adjacent to the active site. The surface is formed by two loops ($\beta 2-\beta 3$ and $\beta 4-H1$ loops) that surround the active site histidine, suggesting that they may play a key role in associating BaSrtC with its lipid II substrate. BaSrtC anchors proteins bearing a noncanonical LPNTA sorting signal. Modeling studies suggest that the enzyme recognizes this substrate using a rigid binding pocket and reveals the presence of a conserved subsite for the signal. This first structure of a class D member of the sortase superfamily unveils class-specific features that may facilitate ongoing efforts to discover sortase inhibitors for the treatment of bacterial infections.

3.2 Introduction

Bacterial surface proteins play key roles in microbial physiology and pathogenesis. Gram-positive bacteria display surface proteins using sortase enzymes, a large superfamily of cysteine transpeptidases, which covalently join proteins bearing an appropriate C-terminal cell wall sorting signal (CWSS) to strategically positioned amino groups located on the cell surface.(1-7) Typically, bacteria encode multiple sortases that either attach proteins to the crossbridge peptide of the cell wall or assemble pili, long proteinaceous structures that extend from the cell surface. Both processes occur through a related transpeptidation reaction that is best characterized for the sortase A enzyme from Staphylococcus aureus (SaSrtA). SaSrtA anchors proteins to the cell wall containing a CWSS that consists of an LPXTG motif (where X denotes any amino acid), followed by a segment of hydrophobic amino acids, and a tail that is comprised primarily of positively charged residues. The C-terminal charged tail presumably retards export, positioning the protein for processing by the extracellular membrane-associated SaSrtA enzyme. Catalysis occurs through a ping-pong mechanism that is initiated when the active site cysteine residue nucleophilically attacks the backbone carbonyl carbon of the threonine residue within the LPXTG motif, breaking the threonine–glycine peptide bond to create a sortase–protein complex in which the components are linked via a thioacyl bond.(8, 9) SaSrtA then transfers the protein to the cell wall precursor lipid II, when the amino group in this molecule nucleophilically attacks the thioacyl linkage to create a peptide-linked protein-lipid II product. The transglycosylation and transpeptidation reactions that synthesize the cell wall then incorporate this product into the peptidoglycan, resulting in a surface-displayed protein that is covalently attached to the crossbridge peptide. Instead of anchoring proteins to the cell wall, some sortases assemble pili.(2, 3) These sortases operate through a similar mechanism; however, they link pilin subunits together

by joining a lysine amino nucleophile located in one protein to the CWSS of another. A molecular-level understanding of sortase function could lead to new therapeutics for the treatment of bacterial infections, as many clinically significant pathogens are attenuated in their virulence when their sortase enzymes are genetically eliminated.(4, 10)

On the basis of primary sequence homology, most sortases in pathogenic Gram-positive bacteria can be grouped into four distinct families, called class A-D enzymes.(7, 11, 12) Class A enzymes are most closely related to SaSrtA. They appear to perform a housekeeping role in different species of bacteria as members of this group have been shown to anchor a large number of functionally distinct proteins to the cell wall that generally contain an LPXTG motif within their CWSS. In contrast, class B-D enzymes appear to have more specialized functions as representative members of these groups display a limited number of proteins that frequently contain noncanonical sorting signals. Class B enzymes are present in Firmicutes and can have distinct functions. Some members of this group attach heme receptors to the peptidoglycan,(13) while others function as polymerases that assemble pili. Class C enzymes are also broadly distributed in Gram-positive bacteria and function as pilin polymerases.(14, 15) Finally, class D enzymes are present in many bacilli (Bacillus cereus, Bacillus anthracis, and Bacillus thuringiensis). A single member of this family has been characterized, the class D enzyme from B. anthracis, which has been shown to attach proteins to the cell wall that facilitate sporulation.(16, 17) Atomic structures of representative class A-C enzymes have been determined, revealing a conserved eight-stranded β -barrel fold that houses three essential active site residues: His120, Cys184, and Arg197 (SaSrtA numbering).(1, 2, 7) The roles of these residues in catalysis have been defined for SaSrtA and are likely conserved in other sortases: Cys184 acts as a nucleophile that attacks the carbonyl atom in the scissile peptide bond of the

CWSS, Arg197 stabilizes the binding of the sorting signal substrate, and His120 may act as a general acid or base.

B. anthracis is a Gram-positive facultative anaerobe that causes lethal anthrax disease.(18) Like other bacterial species within the genus *Bacillus*, it forms dormant endospores (spores) that are capable of surviving for long periods of time under harsh conditions. In humans, anthrax caused by the inhalation of aerosolized B. anthracis spores has a high rate of mortality and has led to their use as a bioterrorism agent.(19) Efficient sporulation of B. anthracis is dependent upon the activity of SrtC (BaSrtC), a class D sortase that attaches the BasH and BasI proteins to distinct cellular structures. (16, 17) BasI is attached to the peptidoglycan of predivisional cells, while BasH is exclusively attached to the forespore presumably by BaSrtC inherited from the mother cell before the polar division takes place. B. anthracis also encodes a housekeeping class A enzyme (BaSrtA) that anchors a different set of proteins to the cell wall.(20, 21) Interestingly, BaSrtC and BaSrtA specifically recognize very closely related sorting signals.(16, 17) BaSrtA recognizes a canonical LPXTG-type sorting signal present in seven B. anthracis proteins (three of these proteins are involved in collagen adhesion, while the functions of the other proteins are not known). In contrast, BaSrtC anchors BasH and BasI that contain LPNTA sorting signals. B. anthracis also encodes a third sortase, BaSrtB.(22) This class B enzyme anchors the IsdC heme binding protein to the cell wall by recognizing its unique NPQTN sorting signal. While the sortase enzymes in *B. anthracis* anchor different proteins to the cell wall by recognizing distinct sorting signals, they are believed to attach these proteins to the same chemical group, the *m*-diaminopimelic acid (*m*-DAP) side chain of the cross-bridge peptide.(17, 23) In this paper, we describe the structure and dynamics of the *B. anthracis* BaSrtC enzyme, which is the first reported structure of a class D enzyme. The structure provides insight

into the mechanism of protein anchoring in pathogenic *B. anthracis*, as well as the evolutionary relationship between different types of enzymes within the sortase enzyme superfamily.

3.3 Results and Discussion

3.3.1 Design and Validation of Soluble BaSrtC

We studied the B. anthracis BaSrtC enzyme to gain insight into the molecular basis through which it selectively anchors proteins to the cell wall during bacterial sporulation. BaSrtC is 198 amino acids in length and contains a nonpolar N-terminal region (residues Met1–Tvr19) that likely functions to embed the protein in the membrane. This is followed by a C-terminal region (residues Lys20-Lys198) that has been shown to mediate the in vitro cleavage of an LPNTA peptide that contains the amino acid sequence of the CWSS found in its BasI protein substrate.(16) Inspection of the amino acid sequence of BaSrtC reveals that the C-terminal region contains residues that are homologous to the catalytic domain of SaSrtA as well as a less conserved polypeptide segment that connects the presumed catalytic domain to the N-terminal nonpolar region (Figure 1A). Structural studies have revealed that many sortases can contain ordered appendages that either precede or follow the amino acid sequence of the catalytic domain. We therefore used NMR to study two polypeptides, BaSrtC_{$\Delta 19$} (residues Lys20– Lys198), which contains the entire C-terminal region, and BaSrtC_{$\Delta 55$} (residues Ser56–Lys198), which contains only amino acids within the presumed catalytic domain. The ¹H-¹⁵N HSQC spectra of these proteins reveal a similar set of well-dispersed backbone amide chemical shifts indicating that these fragments adopt similar three-dimensional structures (Figure 1B). However, the spectrum of BaSrtC_{$\Delta 19$} also contains an additional ~30 cross-peaks that exhibit narrow line widths and random coil chemical shifts. Surprisingly, this indicates that unlike sortases from other classes, the linker preceding the catalytic domain in BaSrtC is unstructured in the isolated protein.

Previous studies have shown that the intact C-terminal region of BaSrtC selectively cleaves an LPNTA peptide derived from the CWSS of its BasI protein substrate.(16) To ascertain whether only the conserved catalytic domain is required for this enzymatic activity, the ability of BaSrtC $_{\Delta 55}$ to cleave a peptide containing the LPNTA sequence was determined. MALDI-TOF analysis of a reaction mixture containing BaSrtC $_{\Delta 55}$ and the peptide reveals that the isolated catalytic domain cleaves the peptide between the threonine and alanine residues (Figure 2). Importantly, BaSrtC $_{\Delta 55}$ exhibits specificity for the BasI sorting signal, as it is unable to cleave LPATG and LPETG peptides derived from proteins that are anchored to the cell wall by the BaSrtA sortase (data not shown and ref 16). Combined, these data indicate that *in vitro* the catalytic domain within BaSrtC is folded and sufficient for hydrolytic activity.

3.3.2 NMR Structure of BaSrtC_{A55}

The structure of BaSrtC_{Δ 55} was determined using multidimensional heteronuclear NMR and simulated annealing methods. A total of 2584 experimental restraints were employed, including 2330 interproton distance, 196 dihedral angle, and 58 hydrogen bond restraints. Figure 3A displays the ensemble of the 20 lowest-energy conformers of BaSrtC_{Δ 55}. Each structure exhibits good covalent geometry and has no NOE or dihedral angle violations greater than 0.5 Å or 5°, respectively (structural and restraint statistics are listed in Table 1). The structure of BaSrtC_{Δ 55} is generally well defined by the NMR data; the backbone and heavy atom coordinates of residues Lys62–Ala85, Val94–Gly115, Leu124–Thr172, and Arg185–Lys198 have rootmean-square deviations (rmsd's) from the average structure of 0.42 ± 0.07 and 0.82 ± 0.05 Å, respectively. The structure of $BaSrtC_{\Delta 55}$ also contains three disordered surface loops that surround the active site (colored red and orange in Figure 3A), two of which (red) likely mediate protein dimerization (*vide infra*).

BaSrtC_{$\Delta55$} adopts an eight-stranded β -barrel fold that contains two short helices positioned on opposite sides of the protein (Figure 3B). The structure is initiated by strand $\beta 1$ (Val67–Ile72), which after a short reverse turn interacts with strand β 2 (Lys77–Tyr82) in an antiparallel manner. A poorly ordered 11-amino acid loop then leads into strand β 3 (Val94– Met96), which is positioned parallel and antiparallel to strands β^2 and β^4 (His110–Gly115), respectively. The active site His116 residue is located immediately after strand β 4 and is followed by a large polypeptide segment containing helix H1 (Leu124–Gln126) that stretches to the opposite face of the protein. Strand β 5 (Thr132–Glu136) then joins with β 1 in an antiparallel manner before the peptide reverses direction to form strand $\beta 6$ (Thr141–Thr152). Strand $\beta 6$ wraps around the enzyme toward the active site and contains a sharp point of curvature created by a β -bulge at residues Gln147 and Lys148. A long loop containing helix H2 then leads into strand β 7 (Ile167–Thr172), which lies parallel with respect to strand β 4. A large loop then reverses the directionality of the chain before forming strand β 8 (Arg185–Tyr197) that lies antiparallel with respect to strands $\beta 6$ and $\beta 7$. Strand $\beta 8$ is extensive as it contains a β -bulge at residues Thr193 and Gly194 that allows it to form a continuous set of hydrogen bonding interactions with strand $\beta 6$ that together contribute residues to both faces of the protein structure.

3.3.3 Disordered Loops Positioned near the Active Site Might Mediate *in vitro* Dimerization

On the basis of primary sequence homology, the active site in $BaSrtC_{\Delta 55}$ is formed by three spatially adjacent amino acids: His116, Cys173, and Arg185. These residues are conserved in all sortases, and in BaSrtC_{$\Delta55$}, they are located at the end of a large cleft formed by strands β 4, β 7, and β 8. Cys173 is located at the end of strand β 7, and in the prototypical SaSrtA enzyme, the analogous residue mediates the nucleophilic attack on the threonine carbonyl carbon within the sorting signal. (38, 39) Surrounding the thiol are the side chains of Arg185 (strand β 8) and His116 (strand β 4), which in SaSrtA may stabilize binding of the sorting signal substrate and facilitate acid-base chemistry, respectively.(1) Surprisingly, unlike previously characterized sortases, BaSrtC $_{\Delta 55}$ contains three disordered surface loops that are positioned immediately adjacent to the active site (loops colored red and orange in Figure 3A). The loops are formed by residues that immediately follow the active site cysteine so as to connect strand β 7 to β 8 (the β 7- β 8 loop, residues Tyr173-Asp184), residues immediately following the active site histidine that connect strand β 4 to H1 (the β 4–H1 loop, residues His116–Asp123), and residues immediately adjacent to His116 that connect strands $\beta 2$ and $\beta 3$ (the $\beta 2-\beta 3$ loop, residues Trp83– Gly93). The loops are structurally disordered because an insufficient number of interproton distance restraints were identified in the NOESY spectra to define their position. For the $\beta 2 - \beta 3$ and β 4–H1 loops, few NOEs were observed because many of their residues exhibit resonance line broadening. The broadening is most severe for residues Leu90–Lys92 ($\beta 2-\beta 3$ loop) and His116–Phe121 (β 4–H1 loop), which are completely absent from the NMR spectra. This suggests that residues within the $\beta 2-\beta 3$ and $\beta 4-H1$ loops experience fluctuations in their

magnetic environments that occur on a micro- to millisecond time scale, presumably because they are flexible or because they reside within a surface that mediates protein aggregation.

To gain insight into the origin of structural disorder in the three active site loops, we quantitatively probed N-H bond motions in BaSrtC_{$\Delta55$} by measuring ¹⁵N spin-spin (T_2), ¹⁵N spin-lattice (T_1), and ${}^{1}H{}^{-15}N$ heteronuclear NOE relaxation parameters (Figure S1 of the Supporting Information). In general, the relaxation data are compatible with the structure of BaSrtC $_{\Delta 55}$ as residues within regular secondary structural elements whose coordinates are well defined in the ensemble of conformers have an average ${^{1}H}^{-15}N$ NOE value of 0.79, indicating that they undergo only small amplitude motions on the picosecond time scale. Interestingly, an analysis of the relaxation data suggests that the structurally disordered $\beta 2-\beta 3$, $\beta 4$ –H1, and $\beta 7-\beta 8$ loops may participate in transient protein oligomerization. As the chemical shifts of many residues within the disordered $\beta 2-\beta 3$ and $\beta 4-H1$ loops are unassigned, the relaxation data do not directly report on their dynamic status. However, several assigned surrounding residues within the $\beta 2-\beta 3$ and $\beta 4-H1$ loops, as well as in the adjacently positioned $\beta 7-\beta 8$ loop, exhibit substantially shortened T_2 values compared to those of other residues in the protein (Figure S1A of the Supporting Information). Intriguingly, this indicates that their backbone amide atoms experience fluctuating magnetic environments on the micro- to millisecond time scale that could result from motions within the loops themselves and/or because the loops reside within or near a molecular surface involved in transient oligomerization.

A more detailed analysis of the NMR relaxation data indicates that active site loops likely form a surface that mediates protein oligomerization. Initially, we attempted to model the dynamics of BaSrtC using the Lipari–Szabo model-free formalism, which yields for each backbone N–H bond vector a generalized order parameter (S^2).(35) S^2 ranges from 0 to 1 and

describes the rigidity of the amide bond on the picosecond time scale, with a value of 1 indicating that it is completely immobilized. Surprisingly, fitting of our data consistently yielded many S^2 values of precisely 1.0, which is suggestive of protein aggregation. This is illustrated in Figure 4A, which shows a scatter plot of experimentally derived T_1 and T_2 values for each residue overlaid with model-free predicted values of T_1 and T_2 for a range of S^2 values [colored lines indicate values of T_1 and T_2 expected for backbone amide nitrogens containing S^2 values ranging from 0.5 to 1 (see the key)]. The plot reveals that data for most residues are inconsistent with all reasonable values of S^2 because their T_2 and T_1 times are conjointly too short; this results in the software attempting a "best fit" by assigning values of 1 to S^2 . We hypothesized that these anomalous results are caused by protein aggregation. As protein aggregation is concentrationdependent, we remeasured the T_1 and T_2 times using a sample of BaSrtC_{$\Delta55$} that was diluted to 0.125 mM. As shown in Figure 4B, the data for the dilute sample are in better agreement with values predicted by the model-free formalism primarily because the residues now have longer T_2 values. This is most likely because in the dilute sample a smaller fraction of the protein forms higher-molecular weight oligomers that influence the measured relaxation parameters.

To measure the stoichiometry and affinity of protein aggregation, we performed sedimentation equilibrium ultracentrifugation experiments (Figure 4C). The resultant data are best fit by a monomer–dimer equilibrium model that yields a dimer dissociation constant (KD) of 89 μ M. Thus, at the protein concentrations used in our NMR studies, ~90% of BaSrtC_{Δ 55} likely exists in its dimeric form. This explains our difficulty in fitting the relaxation data with the model-free formalism and why analysis of the relaxation data yielded an overall molecular correlation time of ~16 ns, a value that is significantly longer than expected for a protein with BaSrtC_{Δ 55}'s molecular mass. Taken together, the NMR and centrifugation data indicate that

residues within the $\beta 2-\beta 3$ and $\beta 4$ –H1 loops likely form a dimerization interface and that these residues are broadened beyond detection in the NMR data because of monomer–dimer exchange that occurs at a rate that is intermediate on the chemical shift time scale. This notion is compatible with the positioning of many residues that exhibit anomalously short T_2 times at high protein concentrations as they are positioned proximal to the $\beta 2-\beta 3$ and $\beta 4$ –H1 loops (orange points in Figure 4B and orange spheres in Figure 4D). Apparently, these residues also experience monomer–dimer exchange broadening, however to a lesser extent, allowing us to still detect them. In addition to forming a dimer interface, residues within the three disordered loops may also exhibit elevated mobility as several residues within the $\beta 2-\beta 3$ and $\beta 7-\beta 8$ loops exhibit depressed {¹H}–¹⁵N heteronuclear NOEs indicating that they undergo faster picosecond motions.

The biological significance of the weak *in vitro* dimerization surface detected by NMR and centrifugation is unclear. As BaSrtC is embedded in the membrane, it may indeed dimerize on the cell surface because it undergoes limited two-dimensional diffusion. If this is the case, then class D enzymes might be unique, as previously reported structures of class A–C sortases have shown that they are monomeric(40-55) (it should be noted that biochemical studies have shown that the *S. aureus* SaSrtA enzyme dimerizes with weak affinity, but this finding is controversial(56)). Alternatively, it is also possible that the *in vitro* dimerization we have observed occurs only because BaSrtC is being studied in isolation. In this scenario, dimerization via the large structurally disordered surface occurs because the appropriate binding partner that normally interacts with this surface is missing. On the cell surface, the missing binding partner could be, among others, protein factors involved in secretion (e.g., the SEC translocon), a component of the cell wall, or BaSrtC's intact lipid II and protein substrates. This is supported by microscopy studies that have shown that the SrtA and SecA proteins in Streptococcus pyogenes

colocalize at the cross wall compartment where the cell wall is synthesized, and at polar sites where surface protein anchoring also occurs.(57) It would also explain the in vitro catalytic properties of BaSrtC as we and others have been able to demonstrate only that it can hydrolyze its sorting signal (Figure 2 and ref 16). The full transpeptidation reaction using *m*-DAP as a nucleophile has never been demonstrated biochemically. This is presumably because to be fully active, BaSrtC needs to either associate with other factors on the cell surface or bind to its intact lipid II and protein substrates. The notion that the disordered surface mediates substrate recognition on the cell surface is supported by NMR and X-ray studies of SaSrtA and SaSrtB that have implicated the $\beta7-\beta8$ loop in binding the cross-bridge peptide. (49, 53) Furthermore, it is interesting to note that of all the sortase enzymes whose structures have been determined only the sortases from *B. anthracis* (BaSrtA, BaSrtB, and BaSrtC) contain a disordered $\beta7-\beta8$ loop and only these enzymes attach proteins to lipid II molecules with a *m*-DAP group.(50, 51) Thus, on the cell surface, the large disordered loop regions identified by NMR may play a key role in associating BaSrtC with protein factors and substrates that facilitate cell surface protein anchoring.

3.3.4 Class A and D Sortases Are Structurally Related

Structures of representative class A–C sortases have been determined. Despite a limited degree of sequence homology, BaSrtC is most structurally similar to class A enzymes based on a DALI analysis. Three structures of class A enzymes have been reported: SaSrtA, BaSrtA, and the SrtA enzyme from *S. pyogenes* (SpSrtA).(41, 48-50, 52) The sequence of BaSrtC is only 23–24% identical with the sequences of these enzymes; however, the coordinates of its carbon atoms that participate in regular secondary structures can be superimposed with rmsd's of 2.4, 1.9, and
2.0 Å, respectively. Notably, both class A and D sortases contain a short 310-helix located within the $\beta 6-\beta 7$ loop (Figure 5, yellow). In SaSrtA, NMR studies have shown that this loop interacts with the leucine-proline portion of the LPXTG sorting signal, suggesting that BaSrtC will also recognize its distinct LPNTA sorting signal through a generally similar mechanism (vide infra).(49) Interestingly, class A enzymes also exhibit structural heterogeneity in their N-termini, because BaSrtA possesses a long N-terminal appendage that contacts the active site histidine. This N-terminal appendage is disordered in the structures of BaSrtC, SaSrtA, and SpSrtA (Figure 5, green). Although the BaSrtC enzyme adopts a canonical sortase fold, its structure differs markedly from class B and C enzymes. For example, in contrast to class B enzymes that possess a large structured $\beta 6-\beta 7$ loop that contains an α -helix, the analogous loop in BaSrtC is substantially shorter and lacks a similarly positioned helix (Figure 5, yellow).(42, 51, 53) Unlike class B enzymes, BaSrtC_{$\Delta55$} is also missing the N-terminal α -helices that precede the catalytic domain (Figure 5, green). The structure of BaSrtC $_{\Delta 55}$ also differs substantially from that of class C enzymes as it is missing an N-terminal "lid" appendage that has been proposed to regulate access to the active site in several members of this group. (44, 46, 54, 58)

3.3.5 Model of the BaSrtC₄₅₅–Sorting Signal Complex

To gain insight into how $BaSrtC_{\Delta 55}$ recognizes its sorting signal, we modeled the structure of its substrate complex by superimposing the backbone coordinates of $BaSrtC_{\Delta 55}$ onto the coordinates of our previously determined solution structure of the SaSrtA enzyme covalently bound to an LPAT peptide analogue.(49, 59) Inspection of the model reveals that BaSrtC and SaSrtA can bind to their respective sorting signals in a similar manner as there is minimal atomic overlap between the peptide and BaSrtC atoms in the model of the BaSrtC–peptide complex

(Figure 6A). In the model, the sorting signal peptide adopts a kinked structure and rests in a groove whose base is formed by residues in strands $\beta 4$ and $\beta 7$ and whose walls are formed by the $\beta 6-\beta 7$, $\beta 7-\beta 8$, $\beta 3-\beta 4$, and $\beta 2$ -H2 surface loops. Because the experimentally determined structure of the apo form of BaSrtC can readily accommodate the peptide, this suggests that the enzyme may recognize its sorting signal substrate through a lock-and-key mechanism. This is in marked contrast to SaSrtA, which undergoes major changes in the structure and dynamics of the $\beta 6-\beta 7$ loop upon signal binding.(49, 60) This can be seen in Figure 6B. The left panel shows apo-SaSrtA (red) aligned with our BaSrtC structure (green). While the active site residues of apo-SaSrtA (pink) align well with the active site residues of BaSrtC (gray), the agreement is significantly better than that of the holo-SaSrtA (SaSrtA–LPAT*) structure in Figure 6B [middle panel (blue)]. Furthermore, the BaSrtC active site residues more closely resemble the position of holo-SaSrtA (Figure 6B, right panel, orange). This result further suggests that BaSrtC exists in an "active-ready" form prior to binding its cognate sorting signal.

To facilitate further discussion of the recognition process, hereafter we refer to amino acids within the signal as P and P' if they precede and follow the scissile peptide bond, respectively [e.g., the LPNTA signal-recognized BaSrtC is Leu (P4)-Pro (P3)-Asn (P2)-Thr (P1)-Ala (P'1)]. Corresponding binding sites on the enzyme for these residues are termed S and S', respectively. The model clearly defines enzyme subsites that can accommodate the leucine (P4) and proline (P3) residues in the LPNTA signal. The leucine residue is recognized by a small pocket on the enzyme that is formed by the $\beta 6-\beta 7$ loop and residues in the underlying β -sheet. Within the S4 subsite, the leucine is contacted by the side chains of Val166, Pro168, Val173, and Val174 on the $\beta 6-\beta 7$ loop and Val198 on strand $\beta 8$. The S3 subsite, which recognizes the LPNTA signal's proline residue, is formed by residues in the underlying β -sheet [Ala124 (β 4)] and Ile185 (β 7)] and Val110 within the adjacent β 3- β 4 loop. The structures of the class A BaSrtA and SpSrtA enzymes also contain similarly shaped nonpolar subsites (48, 50) which is compatible with bioinformatics studies that have predicted that the majority of sorting signals processed by class A and D enzymes contain proline and leucine amino acids at positions P3 and P4, respectively.(11) In B. anthracis, the class A BaSrtA and class D BaSrtC enzymes anchor proteins to the cell wall that contain closely related LPXTG and LPNTA sorting signals, respectively. Our modeling studies suggest both enzymes recognize the conserved leucineproline element of these signals through structurally conserved subsites that do not require substrate binding to form. How each distinguishes differences that occur at the P1' position cannot be predicted as the coordinate precision of amino acids that contact this residue is poorly defined in the NMR ensembles of both enzymes. However, BaSrtC may in part distinguish the unique signal present in its BasH and BasI substrates by interacting with the P2 amino acid that is invariably an asparagine, as the bottom of BaSrtC's S2 subsite contains Ser114, which can presumably favorably interact with the asparagine amide group through hydrogen bonding (Figure 6A, gray), whereas BaSrtA contains a hydrophobic residue (Ala124).

At present, more than 800 sortase genes have been identified in nearly 300 species of bacteria. The functions of the vast majority of these enzymes are not known because most bacteria contain multiple sortases with unknown substrate specificities. Work described here lays the foundation for more detailed structure–function studies that will elucidate the molecular basis of substrate specificity and could facilitate future efforts to predict sortase function and to discover therapeutically useful sortase inhibitors. Structural data could aid inhibitor discovery efforts in several ways.(61) First, atomic structures can be used to discover new chemical

scaffolds that are likely to bind and inhibit sortases. In this approach, an *in silico* screen of compound libraries could be performed to identify small molecules that have the appropriate physicochemical properties to efficiently interact with the enzyme active site. These molecules serve as new leads for further development after their inhibitory properties have been confirmed experimentally. Second, the structures can be used to optimize the binding affinity and selectivity of established sortase inhibitors. In this approach, computational methods are used to dock the lead inhibitor molecule to the structure of the enzyme. Analogues that are likely to have increased potency are then identified in silico by modifying the bound inhibitor and calculating an algorithm-dependent binding energy. The most promising of these analogues are then tested biochemically to ascertain whether they have improved activity. Finally, if the structure of the enzyme-substrate complex is known, computational methods can be used to identify small molecules with physicochemical features similar to those of the substrate. These molecules presumably bind and inhibit the enzyme that can be tested experimentally. As several sortases with known structures recognize sorting signals with distinct amino acid sequences, their unique active site features in combination with the aforementioned computational methods may allow class- and/or bacterial species-specific inhibitors to be developed. Such molecules are urgently needed as infections caused by MRSA and other multidrug-resistant bacteria are a major health concern.

3.4 Experimental Procedures

3.4.1 Cloning, Protein Expression, and Purification

BaSrtC_{Δ 19} and BaSrtC_{Δ 55} were amplified via polymerase chain reaction (PCR) from genomic *B. anthracis* DNA (*Sterne* strain) with primers that placed an NdeI restriction site and a

BamHI restriction site on the 5' and 3' ends of the PCR product, respectively. Each PCR product was digested with NdeI and BamHI restriction enzymes, as was empty vector pET15B (Qiagen). Digested PCR products and pET15B plasmid were ligated together and transformed into *Escherichia coli* XL-1 cells (Stratagene). Successful transformants were confirmed by DNA sequencing. Plasmids where then transformed into *E. coli* BL21(DE3) (Stratagene) for expression.

Protein for enzymatic assays was expressed in BL21(DE3) cells in standard Luria-Bertani broth (LB), at 37 °C. Isotopically labeled protein for nuclear magnetic resonance (NMR) studies was expressed in BL21(DE3) cells grown in M9 medium supplemented with ¹⁵NH₄Cl and/or $[^{13}C_6]$ glucose. At an A_{600} of ~0.7, expression was induced by the addition of 1 mM isopropyl β d-thiogalactopyranoside (IPTG). In the case of LB cultures, cells were harvested after 4 h by centrifugation at 6000g and stored at -80 °C. Cells grown in M9 media where shifted to 22 °C after induction, with induction allowed to progress overnight. Pellets were resuspended in BugBuster (Novagen) with the addition of phenylmethanesulfonyl fluoride (PMSF) and benzamidine (the final concentration of each was 2 mM) and allowed to lyse at room temperature for 30 min. The lysate was cleared by centrifugation at 13000g for 30 min at 4 °C. The soluble fraction was incubated with TALON His-affinity resin (Clontech). The resin was washed with 20 mL of lysis buffer followed by 10 mL of 10 mM imidazole followed by a final washing with thrombin cleavage buffer [20 mM Tris (pH 8.0), 150 mM NaCl, and 2.5 mM CaCl₂]. Resin was then treated with thrombin at 37 °C for 2 h, releasing untagged BaSrtC proteins. Protein was further purified by Sephacryl-100 gel filtration in either NMR buffer [20 mM HEPES (pH 6.0)] or peptide cleavage assay buffer (see below).

3.4.2 NMR Spectroscopy and Structure Calculations

BaSrtC_{A55} was concentrated to 1 mM in NMR buffer [20 mM HEPES (pH 6.0) in 7% D2O]. In addition, PMSF and ethylenediaminetetraacetic acid (EDTA) were added to final concentrations of 1 mM each. A second sample was produced by lyophilizing BaSrtC_{$\Delta55$} in NMR buffer and resolubilizing it with an equal volume of 99.999% D₂O. NMR spectra were recorded at 298 K on Bruker Avance 500, 600, and 800 MHz spectrometers equipped with tripleresonance cryogenic probes. All NMR spectra were processed using NMRPipe(24) and analyzed using CARA (version 1.8.4).(25) Chemical shift assignments (¹H, ¹³C, and ¹⁵N) were obtained by analyzing the following experiments: HNCA, HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO, HNHB. HNHA, HBHA(CO)NH, CC(CO)NH, HCCH-TOCSY, HCCH-COSY, (HB)CB(CGCDCE)HE, and (HB)CB(CGCD)HD.(26, 27) TALOS+(28) was used to obtain φ and ψ dihedral angle restraints. Additional φ and ψ dihedral angle restraints were determined by measuring relative HNHA_i and HNHA_{i-1} NOE intensities. Distance restraints were obtained from three-dimensional ¹⁵N- and ¹³C-edited NOESY spectra. The collection of residual dipolar couplings (RDCs) was attempted in a PEG/hexanol mixture; however, protein alignment was not achieved. Initial NOE assignments were determined using ATNOS/CANDID(29, 30) and simulated annealing in NIH-XPLOR.(31) These NOE assignments were subsequently transferred back to CARA for validation and additional manual picking of NOE assignments. Hydrogen bonds were identified initially on the basis of calculated structures and NOE patterns, verified using deuterium exchange of backbone amides, and included in NIH-XPLOR calculations using the HBDB algorithm.(32) A total of 200 final structures were calculated. Structures with no NOE or dihedral angle violations greater than 0.5 Å or 5°, respectively, were selected with the 20 lowest-energy conformers presented. The BaSrtC-LPAT* complex was modeled by aligning the

SaSrtA–LPAT* structure [Protein Data Bank (PDB) entry 2KID] with the BaSrtC structure in PyMOL.(33) The enzyme portion of the SaSrtA–LPAT* complex was removed, leaving behind the substrate only. Manual adjustment of the remaining LPAT* substrate was done to remove atom clashes. The NMR structure of the apoenzyme has been deposited as PDB entry 2LN7.

3.4.3 ¹⁵N Relaxation Measurements

¹⁵N relaxation data (T_1 and T_2) and heteronuclear {¹H}-¹⁵N NOE data were collected on a Bruker Avance 600 MHz spectrometer equipped with a triple-resonance cryogenic probe at an initial protein concentration of 1 mM. Relaxation data and parameters were analyzed in SPARKY,(34) which generated raw relaxation and NOE parameters. Relaxation parameters were analyzed using programs kindly provided by Arthur G. Palmer III (Columbia University, New York, NY) in an attempt to perform "ModelFree" analysis.(35) Using the data from a protein concentration of 1 mM, the ModelFree formalism failed to generate meaningful results. However, we noticed that the R2R1_tm software that utilizes the R_2/R_1 ratio to calculate an approximate correlation time (τ_c) resulted in an unusually high number for a protein of ~16.5 kDa (τ_c at 1 mM ~ 16 ns). We conjectured that monomer–dimer exchange phenomena might result in aberrant tumbling and relaxation behavior, so we measured T_2 and T_1 values at an additional protein concentration 0.125 mM. To compare our relaxation data to the ModelFree formalism of Lipari and Szabo,(36, 37) we used eqs 7a and 7b from ref 36 to plot T_1 versus T_2 for order parameters from $S^2 = 1$ to $S^2 = 0.6$.

3.4.4 Analytical Ultracentrifugation and Enzyme Assays

Runs of sedimentation equilibrium were performed at 25 °C on a Beckman Optima XL-A analytical ultracentrifuge in 12 mm path length double-sector cells. All samples were in 20 mM HEPES (pH 6.0) and 100 mM NaCl with the addition of 5 mM TCEP. Absorption was monitored at 280 nm at sample concentrations of 0.03, 0.015, and 0.0075 mM. Sedimentation equilibrium profiles were measured at 40000 and 50000 rpm. The data were initially fit with a nonlinear least-squares exponential fit for a single ideal species using the Beckman Origin-based software (version 3.01). Analysis of the association behavior used the global analysis software (the multifit option of the software mentioned above) to analyze four scans simultaneously, corresponding to protein at a concentration of 0.03 mM at 40000 and 50000 rpm. Partial specific volumes were calculated from the amino acid composition.

Peptides were synthesized by NEO Biosciences and used without further purification. BaSrtC (10 μ M) was incubated with 200 μ M peptide substrate in 5 mM CaCl₂ and 20 mM HEPES (pH 7.5) at room temperature. Samples (1 μ L) were removed and plated with 0.5 μ L of 2,5-dihydroxybenzoic acid (DHB) matrix (Acros Organics) dissolved in 50% ethanol and 0.1% trifluoroacetic acid (TFA) and analyzed by MALDI-TOF immediately upon addition of enzyme, and again after 24 h. Each reaction was performed in triplicate, and all samples were spotted and analyzed in triplicate.

3.5 Figures



Figure 3.1 NMR spectra and amino acid sequence of BaSrtC.

(A) Sequence alignment of *B. anthracis* (Ames strain) class A sortase (BaSrtA), class B sortase (BaSrtB), and class D sortase (BaSrtC). The sequence alignment was performed by ClustalW.63 Conserved residues are colored *orange*, while identical residues are colored *red*. The predicted

transmembrane region (TM) is indicated by a cylinder labeled TM. Secondary structure features of BaSrtC are indicated by cylinders or arrows above the sequence. **(B)** $^{1}H^{-15}N$ HSQC spectra of BaSrtC_{$\Delta 19$} (left) and BaSrtC_{$\Delta 55$} (right). BaSrtC_{$\Delta 19$} contains approximately 30 more peaks, the majority of which reside toward the center of the spectrum. The BaSrtC_{$\Delta 55$} HSQC spectrum has its residue assignments indicated.



Figure 3.2 MALDI-TOF demonstrating BaSrtCA55 can cleave the LPNTA-containing peptide.

A peptide substrate, VQGEKLPNTASNN, was incubated with BaSrtC_{$\Delta 55$} for 24 h at room temperature. MALDI spectra of the peptide were taken immediately upon addition of the enzyme **(A)** and after incubation with BaSrtC_{$\Delta 55$} for 24 h **(B)**.



Figure 3.3 Structures of *B. anthracis* class D sortase.

(A) Stereoview of the overlay of the ensemble of NMR structures of BaSrtC_{$\Delta 55$} (PDB entry 2LN7). The unassigned $\beta 2$ – $\beta 3$ and $\beta 4$ –H1 loops are colored *red*. The assigned but poorly ordered $\beta 7$ – $\beta 8$ loop is colored *orange*. (B) Ribbon diagram with secondary structure elements labeled. Loop structures are labeled with text. Active site residue side chains for Arg185, Cys173, and His86 are colored *green*.

	(SA)	$(\overline{SA})_r$
rmsd from NOE restraints (Å) ^{b}	0.029 ± 0.001	0.046
all (2330)	0.032 ± 0.001	0.058
sequential $(i - j = 1)$ (617)	0.039 ± 0.002	0.054
medium-range ($ i - j \le 4$) (292)	0.027 ± 0.001	0.040
long-range $(i - j \ge 5)$ (977)	0.021 ± 0.001	0.034
intraresidue (444)	0.021 ± 0.001	0.034
rmsd from dihedral angles restraints (deg) ^c (196)	0.182 ± 0.087	0.095
deviation from idealized covalent geometry		
bonds (Å)	0.007 ± 0.003	0.004
angles (deg)	0.414 ± 0.172	0.523
impropers (deg)	0.191 ± 0.081	0.493
PROCHECK results $(\%)^d$		
most favorable region	85.6 ± 2.6	89.0
additionally allowed region	14.4 ± 2.6	11.0
generously allowed region	0.0 ± 0.0	0.0
disallowed region	0.0 ± 0.0	0.0
coordinate precision $(Å)^{e,f}$		
protein backbone	0.42 ± 0.07	
protein heavy atoms	0.82 ± 0.05	

Table 3.1Structural Statistics for the Solution Structure of BaSrtC $_{\Delta 55}^{a}$

^{*a*}The notation of the NMR structures is as follows. (SA) represents the final 20 simulated annealing structures. $(\overline{SA})_r$ represents the average energy-minimized structure. The number of terms for each restraint is given in parentheses.

^{*b*}None of the structures exhibited distance violations of >0.5 Å or dihedral angle violations of $>5^{\circ}$.

^{*c*}The experimental dihedral angle restraints were as follows: 95 ϕ , 94 ψ , and 39 χ_1 angular restraints.

^{*d*}Determined using PROCHECK. (62)

^{*e*}The coordinate precision is defined as the average atomic rmsd of the 20 individual SA structures and their mean coordinates. These values are for residues 63–85, 94–113, 124–172, and 182–198 of BaSrtC_{$\Delta 55$}. Backbone atoms are N, C α , and C'. Assignments were made for residues 58–89, 93–115, 122–183, and 184–198 of BaSrtC_{$\Delta 55$}.

^{*f*}Structure calculations also included 58 hydrogen bonds. These bonds were included in xplor structure calculations as HBDB terms as described in ref 32.



Figure 3.4 Dynamics and ultracentrifugation analysis.

(A) Plot of experimentally determined T_2 vs T_1 data (black dots) for 1 mM BaSrtC_{$\Delta55$} (assignable backbone ¹⁵N resonances). Also plotted are lines that correspond to calculated T_2 and T_1 values for various S^2 order parameter values and correlation times (correlation times increase along each line in a clockwise fashion). (B) Same as panel A, but T_2 and T_1 data were collected at a protein concentration of 0.125 mM. Residues with T_2 times of <0.07 s are colored *orange*. (C) Analytical ultracentrifugation equilibrium data demonstrating that BaSrtC is a dimer with a KD

of 89 μ M. (**D**) Structure of BaSrtC with unassigned residue backbone nitrogens colored *red*. Residue backbone nitrogens with short T_2 times identified in panel B are colored *orange*.



Figure 3.5 Comparison of representative structures from class A–D sortases.

B. anthracis class D enzymes are structurally similar to class A enzymes. BaSrtC lacks the long, structured, $\beta 6-\beta 7$ loop (*yellow*) present in class B enzymes, but the appearance of a short 3₁₀-helix in this region likely indicates the existence of a rigid substrate-binding pocket, similar to BaSrtA. BaSrtC also lacks the structured N-terminal lid present in class C enzymes (*green*).



Figure 3.6 Substrate binding and active site architecture.

(A) Model of the CWSS peptide structure from the SaSrtA–LPAT* complex on the substrate binding surface of BaSrtC. The peptide (green) is taken from the SaSrtA–LPAT* structure without modification. Active site residues are indicated along with the potential Ser114 binding site. The hydroxyl oxygen of Ser114 (Ser114 is colored gray, while its OH group is colored white) comes into close contact with the side chain of position *X*. (B) Close-up overlay of the BaSrtC structure (green) with apo-SaSrtA (red, left), holo-SaSrtA (blue, center), and apo-BaSrtA (orange, right). Active site arginine, cysteine, and histidine residues are denoted. The active site residues of apo-BaSrtC overlay best with those of holo-SaSrtA (SaSrtA–LPAT*).

3.6 References

- Clancy, K. W., Melvin, J. A., and McCafferty, D. G. (2010) Sortase transpeptidases: Insights into mechanism, substrate specificity, and inhibition Biopolymers 94, 385–396.
- Hendrickx, A. P., Budzik, J. M., Oh, S. Y., and Schneewind, O. (2011) Architects at the bacterial surface: Sortases and the assembly of pili with isopeptide bonds Nat. Rev. Microbiol. 9, 166–176.
- Mandlik, A., Swierczynski, A., Das, A., and Ton-That, H. (2008) Pili in Gram-positive bacteria: Assembly, involvement in colonization and biofilm development Trends Microbiol. 16, 33–40.
- Maresso, A. W. and Schneewind, O. (2008) Sortase as a target of anti-infective therapy Pharmacol. Rev. 60, 128–141.
- Marraffini, L. A., Dedent, A. C., and Schneewind, O. (2006) Sortases and the art of anchoring proteins to the envelopes of Gram-positive bacteria Microbiol. Mol. Biol. Rev. 70, 192–221.
- Scott, J. R. and Barnett, T. C. (2006) Surface proteins of Gram-positive bacteria and how they get there Annu. Rev. Microbiol. 60, 397–423.
- Spirig, T., Weiner, E. M., and Clubb, R. T. (2011) Sortase enzymes in Gram-positive bacteria Mol. Microbiol. 82, 1044–1059.

- Frankel, B. A., Kruger, R. G., Robinson, D. E., Kelleher, N. L., and McCafferty, D. G. (2005) Staphylococcus aureus sortase transpeptidase SrtA: Insight into the kinetic mechanism and evidence for a reverse protonation catalytic mechanism Biochemistry 44, 11188–11200.
- Huang, X., Aulabaugh, A., Ding, W., Kapoor, B., Alksne, L., Tabei, K., and Ellestad, G.
 (2003) Kinetic mechanism of Staphylococcus aureus sortase SrtA Biochemistry 42, 11307–11315.
- Suree, N., Jung, M. E., and Clubb, R. T. (2007) Recent advances towards new antiinfective agents that inhibit cell surface protein anchoring in Staphylococcus aureus and other Gram-positive pathogens Mini-Rev. Med. Chem. 7, 991–1000.
- Comfort, D. and Clubb, R. T. (2004) A comparative genome analysis identifies distinct sorting pathways in Gram-positive bacteria Infect. Immun. 72, 2710–2722.
- Dramsi, S., Trieu-Cuot, P., and Bierne, H. (2005) Sorting sortases: A nomenclature proposal for the various sortases of Gram-positive bacteria Res. Microbiol. 156, 289–297.
- Mazmanian, S. K., Ton-That, H., Su, K., and Schneewind, O. (2002) An iron-regulated sortase anchors a class of surface protein during Staphylococcus aureus pathogenesis Proc. Natl. Acad. Sci. U.S.A. 99, 2293–2298.
- Ton-That, H., Marraffini, L. A., and Schneewind, O. (2004) Protein sorting to the cell wall envelope of Gram-positive bacteria Biochim. Biophys. Acta 1694, 269–278.

- Ton-That, H. and Schneewind, O. (2003) Assembly of pili on the surface of Corynebacterium diphtheriae Mol. Microbiol. 50, 1429–1438.
- Marraffini, L. A. and Schneewind, O. (2006) Targeting proteins to the cell wall of sporulating Bacillus anthracis Mol. Microbiol. 62, 1402–1417.
- Marraffini, L. A. and Schneewind, O. (2007) Sortase C-mediated anchoring of BasI to the cell wall envelope of Bacillus anthracis J. Bacteriol. 189, 6425–6436.
- Koch, R. (1876) Die Ätiologie der Milzbrand-Krankheit, begründet auf die Entwicklungsgeschichte des Bacillus anthracis Beitr. Biol. Pflanz. 2, 277–310.
- Brookmeyer, R. and Blades, N. (2002) Prevention of inhalational anthrax in the U.S. outbreak Science 295, 1861.
- Gaspar, A. H., Marraffini, L. A., Glass, E. M., Debord, K. L., Ton-That, H., and Schneewind, O. (2005) Bacillus anthracis sortase A (SrtA) anchors LPXTG motifcontaining surface proteins to the cell wall envelope J. Bacteriol. 187, 4646–4655.
- 21. Aucher, W., Davison, S., and Fouet, A. (2011) Characterization of the sortase repertoire in Bacillus anthracis PLoS One 6, e27411.
- Maresso, A. W., Chapa, T. J., and Schneewind, O. (2006) Surface protein IsdC and Sortase B are required for heme-iron scavenging of Bacillus anthracis J. Bacteriol. 188, 8145–8152.
- Budzik, J. M., Oh, S. Y., and Schneewind, O. (2008) Cell wall anchor structure of BcpA pili in Bacillus anthracis J. Biol. Chem. 283, 36676–36686.

- Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995)
 NMRPipe: A multidimensional spectral processing system based on UNIX pipes J.
 Biomol. NMR 6, 277–293.
- Keller, R. (2004) The Computer Aided Resonance Assignment Tutorial, 1st ed., CANTINA Verlag, Goldav, Switzerland.
- 26. Palmer, A. G., III (1995) Protein NMR Spectroscopy, Academic Press, San Diego.
- 27. Teng, Q. (2005) Structural Biology: Practical NMR Applications, Springer, Berlin.
- Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts J. Biomol. NMR 44, 213–223.
- Herrmann, T., Guntert, P., and Wuthrich, K. (2002) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS J. Biomol. NMR 24, 171–189.
- 30. Herrmann, T., Guntert, P., and Wuthrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA J. Mol. Biol. 319, 209–227.
- Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) The Xplor-NIH NMR molecular structure determination package J. Magn. Reson. 160, 65–73.

- Grishaev, A. and Bax, A. (2004) An empirical backbone-backbone hydrogen-bonding potential in proteins and its applications to NMR structure refinement and validation J. Am. Chem. Soc. 126, 7281–7292.
- Delano, W. L. (2006) The PyMOL Molecular Graphics System, version 0.99, DeLano Scientific, South San Francisco, CA.
- Goddard, T. D. and Kneller, D. G. (2001) Sparky NMR Analysis Software, University of California, San Francisco.
- Palmer, A. G., III (2001) NMR probes of molecular dynamics: Overview and comparison with other techniques Annu. Rev. Biophys. Biomol. Struct. 30, 129–155.
- Lipari, G. and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results J. Am. Chem. Soc. 104, 4559–4570.
- Lipari, G. and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity J. Am. Chem. Soc. 104, 4546–4559.
- Liew, C. K., Smith, B. T., Pilpa, R., Suree, N., Ilangovan, U., Connolly, K. M., Jung, M. E., and Clubb, R. T. (2004) Localization and mutagenesis of the sorting signal binding site on sortase A from Staphylococcus aureus FEBS Lett. 571, 221–226.
- Ton-That, H., Liu, G., Mazmanian, S. K., Faull, K. F., and Schneewind, O. (1999)
 Purification and characterization of sortase, the transpeptidase that cleaves surface

proteins of Staphylococcus aureus at the LPXTG motif Proc. Natl. Acad. Sci. U.S.A. 96, 12424–12429.

- 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) Structure analysis and site-directed mutagenesis of defined key residues and motives for pilus-related sortase C1 in group B Streptococcus FASEB J. 25, 1874–1886.
- Ilangovan, U., Ton-That, H., Iwahara, J., Schneewind, O., and Clubb, R. T. (2001) Structure of sortase, the transpeptidase that anchors proteins to the cell wall of Staphylococcus aureus Proc. Natl. Acad. Sci. U.S.A. 98, 6056–6061.
- Kang, H. J., Coulibaly, F., Proft, T., and Baker, E. N. (2011) Crystal structure of Spy0129, a Streptococcus pyogenes class B sortase involved in pilus assembly PLoS One 6, e15969.
- Lu, G., Qi, J., Gao, F., Yan, J., Tang, J., and Gao, G. F. (2011) A novel 'open-form' structure of sortaseC from Streptococcus suis Proteins: Struct., Funct., Bioinf. 79, 2764– 2769.
- Manzano, C., Contreras-Martel, C., El Mortaji, L., Izore, T., Fenel, D., Vernet, T., Schoehn, G., Di Guilmi, A. M., and Dessen, A. (2008) Sortase-mediated pilus fiber biogenesis in Streptococcus pneumoniae Structure 16, 1838–1848.
- 45. Manzano, C., Izore, T., Job, V., Di Guilmi, A. M., and Dessen, A. (2009) Sortase activity is controlled by a flexible lid in the pilus biogenesis mechanism of Gram-positive pathogens Biochemistry 48, 10549–10557.

- 46. Neiers, F., Madhurantakam, C., Falker, S., Manzano, C., Dessen, A., Normark, S., Henriques-Normark, B., and Achour, A. (2009) Two crystal structures of pneumococcal pilus sortase C provide novel insights into catalysis and substrate specificity J. Mol. Biol. 393, 704–716.
- Persson, K. (2011) Structure of the sortase AcSrtC-1 from Actinomyces oris Acta Crystallogr. D67, 212–217.
- Race, P. R., Bentley, M. L., Melvin, J. A., Crow, A., Hughes, R. K., Smith, W. D., Sessions, R. B., Kehoe, M. A., McCafferty, D. G., and Banfield, M. J. (2009) Crystal structure of Streptococcus pyogenes sortase A: Implications for sortase mechanism J. Biol. Chem. 284, 6924–6933.
- 49. Suree, N., Liew, C. K., Villareal, V. A., Thieu, W., Fadeev, E. A., Clemens, J. J., Jung, M. E., and Clubb, R. T. (2009) The structure of the Staphylococcus aureus sortase-substrate complex reveals how the universally conserved LPXTG sorting signal is recognized J. Biol. Chem. 284, 24465–24477.
- 50. Weiner, E. M., Robson, S., Marohn, M., and Clubb, R. T. (2010) The Sortase A enzyme that attaches proteins to the cell wall of Bacillus anthracis contains an unusual active site architecture J. Biol. Chem. 285, 23433–23443.
- 51. Zhang, R., Wu, R., Joachimiak, G., Mazmanian, S. K., Missiakas, D. M., Gornicki, P., Schneewind, O., and Joachimiak, A. (2004) Structures of sortase B from Staphylococcus aureus and Bacillus anthracis reveal catalytic amino acid triad in the active site Structure 12, 1147–1156.

- Zong, Y., Bice, T. W., Ton-That, H., Schneewind, O., and Narayana, S. V. (2004) Crystal structures of Staphylococcus aureus sortase A and its substrate complex J. Biol. Chem. 279, 31383–31389.
- 53. Zong, Y., Mazmanian, S. K., Schneewind, O., and Narayana, S. V. (2004) The structure of sortase B, a cysteine transpeptidase that tethers surface protein to the Staphylococcus aureus cell wall Structure 12, 105–112.
- 54. Khare, B., Fu, Z. Q., Huang, I. H., Ton-That, H., and Narayana, S. V. (2011) The crystal structure analysis of group B Streptococcus sortase C1: A model for the "lid" movement upon substrate binding J. Mol. Biol. 414, 563–577.
- 55. Khare, B., Krishnan, V., Rajashankar, K. R., I-Hsiu, H., Xin, M., Ton-That, H., and Narayana, S. V. (2011) Structural differences between the Streptococcus agalactiae housekeeping and pilus-specific sortases: SrtA and SrtC1 PLoS One 6, e22995.
- Lu, C., Zhu, J., Wang, Y., Umeda, A., Cowmeadow, R. B., Lai, E., Moreno, G. N., Person, M. D., and Zhang, Z. (2007) Staphylococcus aureus sortase A exists as a dimeric protein *in vitro* Biochemistry 46, 9346–9354.
- Raz, A. and Fischetti, V. A. (2008) Sortase A localizes to distinct foci on the Streptococcus pyogenes membrane Proc. Natl. Acad. Sci. U.S.A. 105, 18549–18554.
- 58. Lu, G., Qi, J., Gao, F., Yan, J., Tang, J., and Gao, G. F. (2011) A novel "open-form" structure of sortaseC from Streptococcus suis Proteins 79, 2764–2769.

- Jung, M. E., Clemens, J. J., Suree, N., Liew, C. K., Pilpa, R., Campbell, D. O., and Clubb, R. T. (2005) Synthesis of (2R,3S)-3-amino-4-mercapto-2-butanol, a threonine analogue for covalent inhibition of sortases Bioorg. Med. Chem. Lett. 15, 5076–5079.
- 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) Staphylococcus aureus Sortase A transpeptidase. Calcium promotes sorting signal binding by altering the mobility and structure of an active site loop J. Biol. Chem. 281, 1817–1826.
- Taboureau, O., Baell, J. B., Fernández-Recio, J., and Villoutreix, B. O. (2012)
 Established and Emerging Trends in Computational Drug Discovery in the Structural
 Genomics Era Chem. Biol. 19, 29–41.
- Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR J. Biomol. NMR 8, 477–486.
- Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2002) Multiple sequence alignment using ClustalW and ClustalX. Current Protocols in Bioinformatics, Chapter 2, Unit 2, 3, Wiley, New York.

3.7 Supplementary Materials



Figure S3.1 Residue by residue T_2 , T_1 and ${}^{15}N{}^{1}H$ Heteronuclear NOE data for BaSrtC_{A55}.

Panels A and B are T_2 relaxation data at 1 mM and 0.125 mM respectively.

Panels C and D are T₁ relaxation at 1 mM and 0.125 mM respectively. Panel contains the

 $^{15}\mathrm{N}\{^{1}\mathrm{H}\}$ heteronuclear NOE data. Labeled bars indicate flexible loops. Absence of data

for a particular residue indicates either a proline residue or unassigned residues.

Chapter 4

Structural and Computational Studies of the Staphylococcus aureus

Sortase B-Substrate Complex Reveal a Substrate-Stabilized

Oxyanion Hole

The work described in this chapter has been reproduced with permission from:

Jacobitz, A. W., Wereszczynski, J., Yi, S. W., Amer, B. R., Huang, G. L., Nguyen, A. V., Sawaya, M. R., Jung, M. E., McCammon, J. A., Clubb, R. T. (2014). Structural and computational studies of the Staphylococcus aureus Sortase B-substrate complex reveal a substrate-stabilized oxyanion hole. *Journal of Biological Chemistry*, jbc.M113.509273. <u>http://doi.org/10.1074/jbc.M113.509273</u>

Copyright 2014

Journal of Biological Chemistry

I was the primary author of this work. I designed experiments, performed biochemical assays, crystallography, computational model building, and analyzed data.

4.1 Overview

Sortase cysteine transpeptidases covalently attach proteins to the bacterial cell wall or assemble fiber-like pili that promote bacterial adhesion. Members of this enzyme superfamily are widely distributed in Gram-positive bacteria that frequently utilize multiple sortases to elaborate their peptidoglycan. Sortases catalyze transpeptidation using a conserved active site His-Cys-Arg triad that joins a sorting signal located at the C terminus of their protein substrate to an amino nucleophile located on the cell surface. However, despite extensive study, the catalytic mechanism and molecular basis of substrate recognition remains poorly understood. Here we report the crystal structure of the *Staphylococcus aureus* sortase B enzyme in a covalent complex with an analog of its NPQTN sorting signal substrate, revealing the structural basis through which it displays the IsdC protein involved in heme-iron scavenging from human hemoglobin. The results of computational modeling, molecular dynamics simulations, and targeted amino acid mutagenesis indicate that the backbone amide of Glu224 and the side chain of Arg233 form an oxyanion hole in sortase B that stabilizes high energy tetrahedral catalytic intermediates. Surprisingly, a highly conserved threenine residue within the bound sorting signal substrate facilitates construction of the oxyanion hole by stabilizing the position of the active site arginine residue via hydrogen bonding. Molecular dynamics simulations and primary sequence conservation suggest that the sorting signal-stabilized oxyanion hole is a universal feature of enzymes within the sortase superfamily.

4.2 Introduction

Surface proteins in bacteria play key roles in the infection process by promoting microbial adhesion to host tissues, nutrient acquisition, host cell entry, and the suppression of the immune response. In Gram-positive bacteria, virulence factors are displayed by sortase enzymes, a superfamily of cysteine transpeptidases that join proteins bearing a cell wall sorting signal to the cell wall or to other proteins to construct pili (1-6). Sortases have proven to be useful molecular biology tools to site-specifically attach proteins to a variety of biomolecules and are considered a potential drug target because they display virulence factors. The clinically important pathogen, Staphylococcus aureus, displays surface proteins using two sortase enzymes, sortase A (SrtA)(2) and sortase B (SrtB). srtA- strains of S. aureus are significantly attenuated in virulence, whereas srtB- strains establish less persistent infections (7–9). SrtA plays a "housekeeping" role in the cell, covalently mounting a variety of proteins to the cell wall, whereas SrtB anchors the heme transporter IsdC, a key component of the iron-regulated surface determinant system that captures heme-iron from hemoglobin (10-12). The mechanism of catalysis is best understood for SrtA, the archetypal member of the sortase superfamily. Proteins anchored by SrtA possess a C-terminal cell wall sorting signal that consists of an LPXTG motif that is processed by the enzyme, followed by a hydrophobic transmembrane segment and a positively charged C-terminal tail (13). SrtA operates through a ping-pong mechanism that begins when its active site cysteine residue nucleophilically attacks the carbonyl carbon of the threonine in the LPXTG motif. This results in a tetrahedral intermediate that, after cleavage of the threonine-glycine peptide bond, generates a semistable enzyme-substrate thioacyl intermediate (14, 15). The protein is then transferred by SrtA to the cell wall precursor, lipid II, when the amino group in this molecule nucleophilically attacks the thioacyl linkage creating a

second tetrahedral intermediate that collapses to form the covalently linked protein-lipid II product (16–18). Transglycosylation and transpeptidation reactions that synthesize the cell wall then incorporate this product into the peptidoglycan. SrtB anchors the IsdC protein to the cell wall through a similar mechanism. However, unlike SrtA, SrtB recognizes a unique NPQTN sorting signal, and it attaches IsdC to un-cross-linked peptidoglycan instead of heavily cross-linked peptidoglycan (19, 20).

Sortase enzymes adopt an α/β sortase fold that contains three proximally positioned active site residues: His130, Cys223, and Arg233 (SrtB numbering). Although the nucleophilic role of the cysteine residue is well established, various catalytic functions have been proposed for the histidine and arginine residues. His130 was originally thought to activate the cysteine by forming a histidine-cysteine ion pair (21), but more recent data suggest that it instead functions as a general acid/base (14, 22). Arg233 has been proposed to either stabilize substrate binding (23-26), function as a general base (27), or directly stabilize the tetrahedral catalytic intermediates (14, 22, 23). Other residues within the active sites of sortases have also been proposed to participate in catalysis, including Asp225 in SrtB, which was postulated to participate in a His130-Cys223-Asp225 catalytic triad (28). The catalytic mechanism has remained poorly understood because the tetrahedral and acyl intermediates of catalysis are too short lived to be characterized by either NMR or x-ray crystallography. Several sortase structures have been determined in the absence of their substrates (28–32) or covalently bound to generic sulfhydryl modifiers (33). However, only a single structure of a sortase enzyme covalently bound to its sorting signal substrate has been reported (the NMR structure of SrtA bound to an LPAT substrate analog) (1, 24). This structure revealed that the active site in this enzyme undergoes substantial changes in its structure and dynamics that facilitate specific recognition of the sorting

signal, but it did not provide an atomic level view of the positioning of atoms within the active site because their coordinates were not well defined in the NMR structure because of resonance line broadening. Thus, the structural features utilized by sortases to stabilize key tetrahedral and thioacyl reaction intermediates remain poorly understood.

Here we report the 2.5 Å crystal structure of SrtB covalently bound to an analog of its NPQTN sorting signal. The structure of the complex closely resembles the thioacyl intermediate formed during catalysis, laying the groundwork for MD simulations to investigate the catalytic mechanism. The results of these simulations and *in vitro* transpeptidation measurements suggest that Arg233 and the backbone amide of Glu224 form an oxyanion hole that stabilizes high energy tetrahedral catalytic intermediates. Interestingly, a highly conserved threonine residue within the sorting signal actively participates in constructing the oxyanion hole by hydrogen bonding to the active site arginine residue. MD simulations of SrtA, as well as primary sequence conservation, suggest that all sortases will use a similar substrate-stabilized mechanism to anchor proteins to the cell wall or to assemble pili.

4.3 **Results**

4.3.1 Crystal Structure of the SrtB-NPQT* Complex and Computational Modeling of the Thioacyl Intermediate

Sortase catalyzed transpeptidation reactions occur via covalent enzyme-substrate acyl and tetrahedral intermediates that are too short-lived to be resolved by NMR or x-ray crystallography (2, 14, 43). To overcome this problem, we synthesized a Cbz-NPQT* sorting signal analog, where Cbz is a carbobenzyloxy protecting group, and T* is a threonine derivative that replaces the carboxyl group with $-CH_2$ -SH ((2*R*,3*S*)-3-amino-4-mercapto-2-butanol). The peptide contains

the sorting signal sequence recognized by the S. aureus SrtB sortase and forms a disulfide bond via its T* moiety to the thiol of Cys223, generating a SrtB-NPQT* complex that structurally mimics the thioacyl catalytic intermediate (Fig. 1a). A soluble version of SrtB lacking its 30amino acid N-terminal membrane anchor (SrtB, residues 31-244) was disulfide-bonded to the NPQT* substrate analog, and a 2.5 Å crystal structure of the SrtB-NPQT* complex was determined (Table 1). The data were refined to 2.5 Å resolution in accordance compatible with $CC_{1/2}$ statistics (36, 37), although the data extended only to 2.9 Å using more conventional statistics such as R_{merge} and I/σ . In the complex, SrtB adopts an α/β sortase fold containing eight β -strands that are flanked by five α -helices (Fig. 1b). The positioning of the sorting signal substrate is well defined, as evidenced by an $F_{o} - F_{c}$ omit map of the complex (Fig. 2a). The solvent-exposed Cbz group at the N terminus of the peptide is only partially modeled in the structure of the complex because its electron density was not strong enough to fully define its position. The signal adopts an L-shaped structure and is connected via the sulfhydryl of the T* moiety to the active site cysteine, Cys223. It is nestled within a narrow groove whose base is formed by residues within strands β 4 and β 5 and whose walls are formed by residues projecting from loops connecting strands $\beta 6$ to $\beta 7$ (the $\beta 6/\beta 7$ loop), $\beta 7$ to $\beta 8$ (the $\beta 7/\beta 8$ loop), and $\beta 2$ to $\beta 3$ (the $\beta 2/\beta 3$ loop) (Fig. 1, b and c). To facilitate a discussion of the molecular basis of substrate recognition, we henceforth utilize the nomenclature developed by Schechter and Berger (56), where P and P' refer to amino acids on the N-terminal and C-terminal sides of the scissile peptide bond of the sorting signal, respectively. For the NPQTN sorting signal, the N-terminal N is P4, P is P3, Q is P2, T is P1, and the C-terminal N, occurring after the scissile bond, is P1'. The base of the NPQT binding pocket is defined by residues Asn92, Tyr128, Tyr181, Ile182, and Ser221, with the walls defined by residues Leu96, Thr177, Cys223, Glu224, and Arg233 (Fig. 2, b and
c). The side chain of the disulfide linked T* residue is buried inside of a deep groove where its methyl group contacts the side chains of Tyr128 and Ile182. This positions the hydroxyl oxygen on the threonine residue to accept two hydrogen bonds from the side chain of Arg233 in the active site: a 3.0 Å hydrogen bond to the ϵ -nitrogen atom and a 3.3 Å hydrogen bond to the η nitrogen atom in the guanidino group (Fig. 2, b and c). The side chain of the glutamine residue at position P2 points out of the binding pocket where it is packed against the side chain of Leu96 and donates a hydrogen bond to the backbone carbonyl group of Glu224. Additional enzyme interactions to the backbone of P2 Gln residue stabilize the positioning of the sorting signal. The backbone carbonyl group is held in place by a hydrogen bond from the side chain amide nitrogen of Asn92, and its backbone nitrogen donates a hydrogen bond to a sulfate ion, which in turn is coordinated by the side chain of His93 and the backbone amide of Asn92. The P3 Pro residue rests on top of Ile182 and forms a kink that causes the bound peptide to adopt an L-shaped structure that positions the side chain of the P4 Asn to donate a hydrogen bond to the backbone carbonyl of Thr177 within the $\beta 6/\beta 7$ loop. Substrate binding induces only small changes in the structure of SrtB as the C α coordinates of the SrtB-NPQT* complex can be superimposed with the previously determined structure of the unmodified enzyme (28) with a root mean square deviation of 0.44 Å.

To gain insight into how sortase stabilizes reaction intermediates, an energy-minimized model of the thioacyl complex was generated by replacing the disulfide link in the structure of the SrtB-NPQT* complex with a thioacyl bond (Fig. 3). This required only small changes including the removal of the methylene group in between the Cys223 thiol and the P1 Thr residue of the sorting signal that decreased their separation by ~1 Å. The coordinates of the thioacyl intermediate were then energy-minimized while the restraints on the initial atom

positions were gradually removed. To verify that the final model contained the proper orientation of the thioacyl bond, two initial models of the intermediate were energy-minimized, one in which the carbonyl oxygen pointed toward Arg233 and a second model in which the thioacyl bond was rotated by 180° (carbonyl oxygen pointing away from Arg233). After minimization, both starting models converged to nearly identical structures (root mean square deviation = 0.71 Å for all $C\alpha$) that closely resembled the SrtB-NPQT* complex. Importantly, in both refined models, the thioacyl linkage adopts a similar conformation in which the P1 threonine residue in the signal is poised to accept hydrogen bonds from the active site Arg233 residue and the backbone amide from Glu224 (Fig. 3). This key sorting signal-enzyme interaction may hold the active site Arg233 side chain in a catalytically competent conformation that also stabilizes higher energy tetrahedral reaction intermediates that form during catalysis (see "Discussion").

4.3.2 Identification of Enzyme-Substrate Interactions Required for Catalysis

The catalytic importance of active site and sorting signal amino acids was investigated *in vitro* using SNKDKVENPQTNAGT (sorting signal in bold) and GGGGG peptides that mimic the sorting signal and secondary lipid II substrates, respectively. HPLC separation of the reaction products indicates that SrtB used for crystallographic studies is fully functional *in vitro* (k_{cat} and K_m values for SrtB of 1.0×10^{-4} s⁻¹ and 1.8 mM, respectively) (Fig. 4a) (57). Based on sequence conservation, residues His130, Cys223, and Arg233 in SrtB have been postulated to form a triad that mediates catalysis. In addition, it has been proposed that the side chain of Asp225, positioned near the active site, may also play a critical role in catalysis by stabilizing and activating His130 (28). To investigate the relative importance of these residues, we purified four single amino acid mutants of SrtB and assayed them for their ability to catalyze transpeptidation.

H130A, C223A, and R233A mutant enzymes had no detectable activity after 24 h, whereas D225A exhibited nonspecific proteolytic activity (Fig. 4b). This suggests that unlike the conserved active site residues (His130, Cys223, and Arg233), Asp225 is not required for early steps in transpeptidation that involve the formation of the first thioacyl intermediate. The D225A mutation may disrupt the active site architecture of SrtB, allowing recognition of various sequences as primary substrates, but it does not appear to play a direct role in the catalytic mechanism. Interestingly, similar promiscuous activity has been observed in several SrtA mutants (57, 58), as well as the wild-type SrtB enzyme from Listeria monocytogenes (59).

The importance of enzyme-substrate interactions visualized in the structure of the complex was also tested using the HPLC assay. Sorting signal peptides containing alanine substitutions at sites P1 (NPQAN), P3 (NAQTN), and P4 (APQTN) were unreactive (Fig. 4c), compatible with extensive enzyme contacts to the side chains of these residues in the structure of the complex. In contrast, peptides containing an alanine substitution at either site P2 (NPATN) or P1' (NPQTA) could be processed by the enzyme to yield transpeptidation products, although they were less reactive than the native signal. The ability of SrtB to process the P2 mutant peptide (NPATN) is compatible with the structure of the complex as the side chain of residue P2 projects out of the binding pocket toward the $\beta 7/\beta 8$ loop. The SrtA enzyme also processes signals containing a range of amino acids at site P2 in its LPXTG sorting signal (60), suggesting that this promiscuity is evolutionarily conserved.

The structure of the SrtB-NPQT* complex and computational model of the thioacyl intermediate reveal that the threonine side chain within the sorting signal likely plays a key role in catalysis by stabilizing the positioning of Arg233 through hydrogen bonding (Figs. 2b and 3). To determine the importance of this interaction in catalysis, we tested how efficiently NPQSN

and NPQVN peptides were used as transpeptidation substrates. These peptides are identical to the native sorting signal peptide but contain threonine to serine and threonine to valine mutations at the P1 position of the signal, respectively. The threonine to serine substitution preserves the hydroxyl group that hydrogen bonds to Arg233 but removes the methyl group of threonine that interacts with the side chains of Tyr128 and Ile182. In contrast, introduction of a valine substitution eliminates the hydroxyl group but does not significantly alter the size or shape of the P1 residue. Both peptides were unreactive, suggesting that each type of substrate-enzyme interaction is important for catalysis (Fig. 4c).

4.3.3 Molecular Dynamics Simulations of SrtA and SrtB Reveal a Conserved Mechanism through which the Substrate Stabilizes the Positioning of the Active Site Arginine Residue

The atomic structures of only two sortases covalently bound to their substrates have been determined: the NMR structure of SrtA bound to an LPAT* peptide (24) and the structure of the SrtB complex reported here (Fig. 5). A comparison reveals a generally similar mode of binding in which the signals adopt an L-shaped structure enabling extensive enzyme contacts to the side chains of residues located at positions P4 and P3 (Leu-Pro and Asn-Pro in the SrtA and SrtB substrates, respectively). Interestingly, although both bound sorting signals contain a conserved threonine residue at the P1 position, the side chain of this amino acid is oriented differently in the structures of the SrtA-LPAT* and SrtB-NPQT* complexes. As described above, in SrtB the side chain of the P1 Thr residue faces "in" and interacts with the active site arginine, and the side chain of the P2 Gln residue projects "out" toward the $\beta 7/\beta 8$ loop. In contrast, in the SrtA-LPAT* structure the P1 Thr points out toward the solvent and the P2 Ala residue points in toward the

bottom of the binding pocket. This conformational difference occurs because the P1 and P2 residues in each signal have distinct backbone torsional angles; the P2 ϕ and ψ angles are 51 and 98° for SrtA-LPAT*, and -75 and 127° for SrtB-NPQT*, respectively, and the P1 T* pseudo ϕ and ψ angles are -71 and -17° for SrtA-LPAT* and -119 and 25° for SrtB-NPQT*, respectively. This key structural difference is not caused by a lack of coordinate precision in the NMR structure of the SrtA-LPAT* complex because several NOEs define the positioning of the side chains of the P1 and P2 residues (see Fig. 1c in Ref. 24).

It is possible that the sorting signals bound to SrtA and SrtB are flexible and thus capable of undergoing motions in which the P1 Thr side chain moves into, and out of, the active site. To investigate this issue, we performed MD simulations of both complexes using the method of umbrella sampling with Hamiltonian replica exchange (54). The free energy profile for transitions between the Thr-in (SrtB-like) and Thr-out (SrtA-like) states for the bound signals in each complex was then calculated. MD calculations using the coordinates of the SrtA-peptide complex indicate that the peptide can transition from the Thr-out conformation observed in the NMR structure of the SrtA-LPAT* complex (Fig. 6a, middle panel) to a Thr-in orientation observed in the structure of the SrtB-NPQT complex (Fig. 6a, left panel). To investigate the energetics associated with this transition, we calculated the free energies of intermediate structures on this pathway, which are represented by a two-dimensional coordinate system (Fig. 6b). The first coordinate (x axis) reports on the positioning of the P1 and P2 residues, and the second reaction coordinate (y axis) reports on the positioning of the remainder of the sorting signal relative to the enzyme. The former is defined as a collective coordinate that describes the structure of residues P1 and P2 relative to the catalytic cysteine, and the latter is defined as the radius of gyration of residues P3 and P4 with select atoms in the β sheet of the sortase molecule

(described further under "Experimental Procedures"). In the free energy profile for the covalent SrtA-LPAT complex, there are three dominant energy wells all with minima within 0.4 kcal/mol of one another. The region we refer to as well 1 corresponds to a sorting signal structure similar to that observed in the SrtB-NPQT* complex in which the sorting signal contacts the arginine (Thr-in) (Fig. 6a, left panel), whereas well 2 conformers resemble the NMR structure of the SrtA-LPAT* complex in which the P1 threonine side chain points away from the active site (Thr-out) (Fig. 6a, middle panel). Interestingly, this analysis reveals that a low energy pathway exists between these two states (Fig. 6b), suggesting that the P1 Thr and P2 Ala residues can alter their conformation within the active site of SrtA. This structural transition is documented in supplemental Video S1, which shows select snapshots from the MD trajectory in which the P2 Thr transitions from the Thr-out to Thr-in state where it engages the active site arginine residue. Interestingly, the Thr-out to Thr-in transition can also occur through a third low energy intermediate (well 3) in which the hydrophobic residues P3 and P4 do not contact SrtA but are instead exposed to solution. This entropically stabilized state is presumably not significantly populated *in vivo* when the enzyme contacts larger protein substrates that contain a full cell wall sorting signal.

In contrast to SrtA, MD simulations of the SrtB-NPQT complex reveal only a single, narrow free energy minimum in which the threonine remains projected into the active site (Thrin) where it contacts Arg233 (Fig. 6, a, right panel, and c). This indicates that the Thr-out conformer observed in SrtA is disfavored in SrtB. The larger size of the P2 residue in the SrtB sorting signal could, in principle, cause this difference. In the SrtA-LPAT complex, the P2 Ala residue adopts a positive ϕ angle, and the side chain projects toward the base of the binding pocket, whereas the larger Gln P2 residue in the SrtB bound peptide adopts a negative ϕ angle and rests on the surface of the enzyme. Because reduced steric stress enables amino acids with smaller side chains to more readily adopt positive ϕ angles, it is possible that the smaller size of the alanine side chain in the LPAT signal facilitates formation of the Thr-out state. To test this hypothesis, a third free energy profile was computed for SrtB bound to NPAT, which changes its P2 residue to alanine (Fig. 6d). This change expanded the range of conformers accessible to the peptide bound to SrtB (Fig. 6, compare c and d), but it did not encourage sampling of the Thr-out state observed in SrtA. This suggests that features of the SrtB-peptide complex other than the identity of its P2 residue are important for dictating how the P1 Thr residue is positioned (described below). In sum, our MD simulations indicate that sorting signals bound to both SrtA and SrtB can adopt conformations in which the P1 Thr side chain is located within the active site for stabilizing interactions with the active site arginine. In SrtB this is the predominant state of the signal, whereas in SrtA, it is one of two possible low energy binding conformations.

4.4 Discussion

Members of the sortase superfamily catalyze transpeptidation reactions that covalently attach proteins to the bacterial cell wall or assemble pili (1–3). At present, over a thousand sortase enzymes have been identified that, based on their primary sequences and functions, can be partitioned into at least six distinct families (called class A to F enzymes) (1). The structure of the SrtB-NPQT* complex reveals how class B sortases recognize their substrates. In bacterial pathogens, class B enzymes typically anchor heme-binding proteins to the cell wall that enable the bacterium to utilize host derived heme-iron as a nutrient (1, 10–12). Our results indicate that SrtB recognizes its NPQTN sorting signal via a large groove adjacent to the active site formed by residues in the $\beta 6/\beta 7$ loop and residues within strands $\beta 4$ and $\beta 7$. Structurally, class B enzymes

are distinguished from other members of the sortase superfamily by the presence of a large $\beta 6/\beta 7$ loop that contains a long α -helix. In the complex, this loop plays a major role in signal recognition because it contacts the Asn (P4), Pro (P3), and Thr (P1) residues of the peptide, which are highly conserved in sorting signals recognized by class B enzymes. These substrateenzyme contacts are important for catalysis because alanine substitutions at these sites in the sorting signal disrupt transpeptidation (Fig. 4c). As in the prototypical SrtA enzyme, site P2 in the SrtB sorting signal is tolerant to alanine substitution, which is consistent with the positioning of the glutamine residue at this site, which rests on the surface of the enzyme. Although not visualized in our structure, the P1' residue following the scissile peptide bond is also tolerant to alanine mutation (Fig. 4c), distinguishing SrtB from the prototypical class A SrtA enzyme, which only recognizes glycine at the P1' position (61). Given that the signal sequences and known structures of class B enzymes are highly similar, it is likely that they all recognize their substrates in a fashion similar to what is seen in the SrtB-NPQT* complex. Interestingly, our results indicate that SrtA and SrtB enzymes recognize their cognate sorting signals in a generally similar manner, despite the fact that the enzymes share only 26% sequence identity and that they recognize distinct LPXTG and NPQTN sorting signals, respectively. A comparison of the SrtB-NPQT* and previously reported NMR structure of the SrtA-LPAT* complex reveals that the bound sorting signals both adopt an L-shaped conformation in which the proline residue at position P3 redirects the polypeptide to position the side chain of the P4 residue so that it contacts the $\beta 6/\beta 7$ loop. This general mode of binding is likely a conserved feature of substrate recognition by sortase enzymes as an inspection of sorting signals predicted to be processed by these enzymes shows that over 90% of them contain a proline at P3 (60).

Computational modeling of the thioacyl enzyme-substrate reaction intermediate suggests that SrtB facilitates catalysis by forming a substrate-stabilized oxyanion hole. In both the experimental and energy-minimized model of this reaction intermediate, the conserved active site arginine (Arg233) is held in place near the scissile bond by donating a hydrogen bond from its ϵ -nitrogen to the hydroxyl group of the threenine located at the P1 position in the sorting signal (Figs. 2b and 3). In the model of the thioacyl complex, this interaction positions the arginine so that its guanidino group and the backbone amide of Glu224 donate hydrogen bonds to the oxygen atom in the thioacyl bond. During catalysis, two oxyanionic transition states form; the first precedes the generation of the thioacyl intermediate emulated by the SrtB-NPQT* structure, and the second occurs after nucleophilic attack of the thioacyl bond by the amino group present in lipid II. It seems likely that Arg233 and Glu224 form an oxyanion hole that stabilizes these high energy reaction intermediates because only small changes in the positioning of the oxygen atom within the thioacyl bond are expected to occur when the carbonyl carbon transitions from its planar sp2 configuration to its tetrahedral sp3 state. This is distinct from the oxyanion hole discovered in penicillin binding proteins that perform a similar transpeptidation reaction but use two backbone amides to stabilize the tetrahedral intermediate (62). The threonine residue in the sorting signal appears to play a significant role in stabilizing this oxyanion hole because even conservative mutations to either serine or valine disrupt transpeptidation (Fig. 4c). Its catalytic role is also substantiated by the high degree of sequence conservation at site P1 in all known SrtB substrates. The putative oxyanion hole reported here is compatible with results obtained by McCafferty and co-workers (22, 23), who demonstrated that the analogous arginine in the SrtA enzyme is intolerant to mutation, except when substituted with citrulline, an arginine isostere that lacks a formal positive charge.

MD simulations indicate that the SrtA enzyme can also form a substrate stabilized oxyanion hole that could facilitate catalysis. In the structures of the SrtB-NPQT* and SrtA-LPAT* complexes, the bound sorting signals adopt similar L-shaped conformations. However, the positioning of the P1 Thr residue in each sorting signal differs substantially; in SrtB-NPQT*, the threonine side chain interacts with the active site arginine residue (Thr-in conformation) (described above), whereas in the SrtA complex, the analogous threonine residue projects away from the active site (Thr-out conformation). Because the threonine in the SrtA complex is not positioned to form stabilizing interactions with the active site arginine residue, we wondered whether the Thr-in conformer observed in the SrtB complex could also be sampled by the bound SrtA sorting signal to stabilize its tetrahedral reaction intermediates. To investigate this issue, MD simulations of the SrtA thioacyl complex were performed, revealing that the P1 threonine residue can transition from the Thr-out to the Thr-in state presumably needed to construct the oxyanion hole (supplemental Video S1 and Fig. 6). Conformations of the SrtA-LPAT complex in which the threenine side chain of the sorting signal forms interactions with the active site arginine were obtained without major rearrangement of the structure of the enzyme. Interestingly, a previous MD study based on the SrtA-LPAT* NMR structure reports that the active site arginine residue functions only to position the sorting signal substrate by hydrogen bonding to its backbone carbonyl atoms at positions P2 and P4 (25). Our work is compatible with this conclusion but suggests that this stabilizing interaction will only occur when the P1 Thr samples the out position that is presumably not catalytically active. Thus, based on primary sequence conservation and the demonstrated importance of the P1 Thr in catalysis in both SrtA and SrtB, we conclude that the Thr-in conformation observed in the SrtB structure, and accessible to SrtA, represents a catalytically competent form of the peptide that is essential for stabilizing the tetrahedral transition state.

The active site of SrtB appears to be more conformationally restrictive than SrtA because MD simulations of the SrtB-NPQT thioacyl complex reveal that only the Thr-in conformer is accessible to the bound peptide. This is compatible with structural and NMR relaxation data, which have shown that the active site of SrtA contains a flexible $\beta 6/\beta 7$ loop that becomes ordered upon substrate binding (24, 63), whereas SrtB contains a preformed, rigid binding pocket for its sorting signal substrate. Interestingly, inspection of the MD data suggests that SrtA can form unique contacts to the sorting signal that may stabilize the Thr-out conformer. In the NMR structure of the SrtA-LPAT* structure, the conserved active site His101 residue is positioned to form a 3.1 Å hydrogen bond from its ϵ -nitrogen to the P1 Thr hydroxyl. In contrast, although the SrtB active site histidine is about the same distance from the active site cysteine as the analogous residues in SrtA (~5.2 Å from His δ -N to Cys S), this potential stabilizing interaction for the Thr-out conformer is obstructed by the side chain of Leu96, which is inserted between His130 and Cys223 in SrtB. Thus, the more restrictive active site of SrtB and the lack of stabilizing interactions may prevent the signal bound to SrtB from adopting the catalytically nonproductive Thr-out conformer. It is unlikely that the ability of the SrtA peptide to sample the less reactive Thr-out state impacts the kinetics of transpeptidation because the half-life of the long-lived thioacyl intermediate presumably far exceeds the time needed for the threonine residue to transition between its Thr-in and Thr-out conformers.

It seems likely that nearly all members of the sortase superfamily will employ a substrate stabilized oxyanion hole to anchor proteins to the cell wall or to assemble pili. Based on structural and mutagenesis data, the transpeptidation reaction will be initiated when the sorting signal of the partially secreted protein substrate binds to the groove on sortase formed by residues in the $\beta 6/\beta 7$ loop and residues within strands $\beta 4$ and $\beta 7$. Most sorting signals can be expected to adopt an L-shaped structure when bound to the enzyme because they contain a proline residue at position P3 (~90% of predicted sorting signals) (60). For catalysis to proceed, the sortase must contain a properly charged active site in which the cysteine and histidine residues are in their thiolate and imidazolium ionization states, respectively. In isolation, only a small fraction of the enzyme may be properly ionized based on pKa measurements of the SrtA enzymes from S. aureus and Bacillus anthracis (14, 29, 64). If properly ionized, the cysteine thiolate attacks the carbonyl carbon of the P1 residue, forming the first tetrahedral intermediate. Our results suggest that the oxyanion in this intermediate is stabilized by hydrogen bonding from the active site arginine residue and a backbone amide group located in the $\beta 7/\beta 8$ loop (in SrtB Arg233 and the backbone amide of Glu224) (Fig. 7). The threonine residue within the sorting signal, at position P1, plays a key role in constructing the oxyanion hole by stabilizing the positioning of the arginine side chain via hydrogen bonding. This oxyanion hole is presumably used by most sortase enzymes to facilitate catalysis because they all contain conserved active site arginine residues, and $\sim 95\%$ of their predicted sorting signal substrates contain a threonine at the P1 position (60). Breakage of the scissile bond is then facilitated by protonation of the amide group by His130, resulting in a semistable thioacyl intermediate. It remains unclear where the amino nucleophile on lipid II enters the active site in SrtB. A crystal structure of SrtB noncovalently bound to a triglycine peptide, meant to mimic the lipid II substrate, localized the binding site for the peptide to the $\beta 7/\beta 8$ loop (33). However, the specificity of this interaction is suspect because the peptide is expected to bind weakly (K_m for GGGGG binding is 140 μ M (43)) and because the complex was not co-crystallized (the peptide was soaked into the crystal). It is

also important to note that in this and all crystal structures of SrtB solved to date, side chains from residues Asp225–Tyr227 in the $\beta 7/\beta 8$ loop are involved in crystal lattice contacts to symmetry related molecules in the crystal. Although it is possible that these contacts simply reinforce the existing, predominant conformation of this loop, it is also possible that they have captured one of many possible conformations that could be involved in mediating substrate access to the active site. Alternatively, as originally proposed by Joachimiak and co-workers (28), the pentaglycine peptide in lipid II may enter the active site via a groove located between the $\beta 7/\beta 8$ and $\beta 2/\beta 3$ loops. This is compatible with NMR chemical shift perturbation studies of SrtA (24), high resolution crystal structures of other sortase enzymes, which also contain a similarly positioned groove (30, 32), and the presence of the highly conserved histidine residue in this groove that has been proposed to function as a general base that deprotonates lipid II. Our model of the thioacyl intermediate does not rule out either of these entry points. However, it is most compatible with lipid II entering via the groove between the $\beta 7/\beta 8$ and $\beta 2/\beta 3$ loops because from this direction, attack of the carbonyl carbon by the amino nucleophile will generate a tetrahedral intermediate whose negative charge is positioned to be stabilized by the oxyanion hole formed by Arg233. The transpeptidation reaction would then be completed by the collapse of the second tetrahedral intermediate into the final, covalently linked, protein-lipid II product. Additional hybrid quantum mechanics/molecular mechanics simulations are currently underway to quantitatively investigate the role of the oxyanion hole in catalysis. Beyond providing fundamental insight into the process of protein display and pilin assembly in bacteria, the new mechanistic insights reported in this paper could guide the rational design of therapeutically useful transition state analog inhibitors of sortases and facilitate protein engineering efforts to expand the utility of sortases as biochemical reagents.

4.5 **Experimental Procedures**

4.5.1 Production, Crystallization, and Structure Determination of SrtB-NPQT* Complex

DNA encoding SrtB (residues 31–244) was amplified by PCR from *S. aureus* genomic DNA, cloned into a pE-SUMO vector (LifeSensors) and transformed into *Escherichia coli* Rosetta (DE3) pLysS cells (Novagen). Protein expression was induced by addition of 1 mM isopropyl β -D-1-thiogalactopyranoside and allowed to continue for 16 h at 16 °C. Protein was purified by affinity purification using HisPur cobalt resin (Thermo) per the manufacturer's instructions. The His₆-SUMO tag was then cleaved by incubating the protein overnight at 4 °C with recombinant ULP1 protease and removed by reapplying the protein mixture to the HisPur cobalt resin. Cbz-NPQT* (where T* is (2*R*,3*S*)-3-amino-4-mercapto-2-butanol, and Cbz is a carbobenzyloxy protecting group) was synthesized as in Ref. 34 and added to purified SrtB in modification buffer (10 mM Tris-HCl, pH 7.0, 20 mM NaCl, 1 M L-proline) at a ratio of 10:1 for a final concentration of 1 mM Cbz-NPQT* to 100 μ M SrtB. The reaction was first reduced with 1 mM DTT for 4 h, then oxidized by addition of 10 mM CuCl2, and allowed to rock gently at room temperature for 7 days. Production of stable complex was confirmed by MALDI mass spectrometry.

Crystals of the SrtB-NPQT* complex were produced from a stock of 150 μ M SrtB-NPQT* in 10 mM Tris-HCl, pH 7.0, 20 mM NaCl. Crystals were grown using the hanging drop, vapor diffusion method in 2.8 M ammonium sulfate, 70 mM sodium citrate, pH 5.0. Data were collected on Beamline 24-ID-C at 100 K at the Advanced Photon Source ($\lambda = 0.964$ Å). Three data sets were scaled, integrated, and merged using XDS and XSCALE (35). Using conventional criteria, the resolution boundary for the data set might have been drawn at 2.9 Å given that I/σ in

this shell (2.98–2.90 Å) is 2.0, and R_{merge} is 72%, with a completeness of 95% and a multiplicity of 5.2. However, recent studies from Karplus and Diederichs (36) have indicated that the CC_{1/2} statistic has superior properties as an indicator of data precision compared with R_{merge} . Moreover, in their study, and in the following work (37), the authors show that high resolution data typically discarded because of high R_{merge} values (i.e., over the conventionally acceptable threshold value of ~60–80%), and low I/σ values (i.e., under the conventionally acceptable threshold value of 2) actually contain information that can improve the quality of the model if used in refinement. Given these results, we thought our model would improve if we used the more generous resolution cutoff (2.5 Å) indicated by the CC_{1/2} statistic (50.4% in the 2.5 Å shell), rather than a conventionally accepted limit (2.9 Å) indicated by the R_{merge} and I/σ statistics. Thus, data extending to 2.5 Å resolution were used for the refinement process, although by conventional standards, the structure should be considered to be resolved at 2.9 Å resolution.

Phases were determined by molecular replacement using the unmodified SrtB structure (28) (Protein Data Bank code 1NG5) as a search model in the program Phaser (38). The NPQT* modifier was modeled into positive density using COOT (39, 40), and the model was prepared through successive iterations of manual adjustment in COOT and refinement in BUSTER (41, 42).

4.5.2 Transpeptidation Assay

Active site mutants were produced using the QuikChange site-directed mutagenesis kit (Stratagene) as per the manufacturer's instructions, confirmed by DNA sequencing, and expressed and purified as described for the wild-type protein. *In vitro* transpeptidation reactions

were performed based on the method developed by Kruger et al. (43). 100 μ M SrtB (wild-type or mutant) was incubated with 2 mM GGGGG and 200 μ M peptide substrate in 100 μ l of assay buffer (300 mM Tris-HCl and 150 mM NaCl) at 37 °C for 24 h. The reactions were quenched by adding 50 μ l of 1 M HCl and injected onto a Waters XSelect HSS C₁₈ reversed phase HPLC column. Peptides were eluted by applying a gradient from 3 to 23% acetonitrile (in 0.1% trifluoroacetic acid) over 25 min at a flow rate of 1 ml/min. Elution of the peptides was monitored by absorbance at 215 nm. Peak fractions were collected, and their identities were confirmed by MALDI-TOF mass spectrometry.

4.5.3 Computational Modeling and Molecular Dynamics

Molecular dynamics simulations were performed with NAMD (44), using the AMBER99SB-ILDN force field (45), a 2-fs time step, and the SHAKE algorithm to constrain all hydrogen containing bonds (46). Nonbonded interactions were truncated at 10 Å, with the use of a smoothing function beginning at 9 Å, and long range electrostatics were handled with the particle mesh Ewald method using a maximum grid spacing of 1 Å and a cubic B spline (47). Parameters for the Cys-Thr linkage were generated with GAFF (48, 49), with the charges derived from a RESP fit (48). Constant temperature was maintained through the use of Langevin dynamics with a damping coefficient of 2 ps⁻¹, whereas the barostat was controlled through a Nosè-Hoover method with a target pressure of 1 atm, a piston period of 100 fs, and a damping time of 50 fs (50, 51).

Models of the thioacyl intermediate were originally constructed from the SrtB-NPQT* structure by replacing the disulfide bond with a thioester in PyMOL (52). The models were solvated in a periodic water box with a solvent distance of 10 Å and parameterized in tLeap (53).

Models were then energy-minimized and equilibrated in NAMD (44) by slowly removing restraints from the initial atom positions over 1 ns with 2-fs steps. For simulations of SrtA, the NMR structure 2KID was utilized (24).

Potential of mean force calculations were performed using two-dimensional replicaexchange umbrella sampling calculations (54). For the first dimension (the x coordinate in Fig. 6, b-d), a vector was defined based on the difference in positions between residues in the SrtA structure 2KID and the SrtB structure presented here for the heavy atoms in the backbones of residues P1, P2, the catalytic cysteine, and the three residues upstream of it in the sortase molecule, along with the heavy atoms in the backbone of residues P1 and the catalytic cysteine. The second coordinate (the y axis in Fig. 6, b–d) was defined as the radius of gyration for the C atoms in residues 98-100 and 142-144 in SrtA (or residues 171-173 and 235-237 in SrtB) along with the heavy atoms of residues P4 and P3. Restraints for umbrella sampling were evenly spaced every 10 Å from -100 to 100 Å in the first coordinate and every 0.5 Å from 5 to 14 Å in the second coordinate. This created a total of 399 simulation "windows," each of which were simulated for 10 ns. Positions were exchanged between adjacent windows every 1 ps based upon a Metropolis criteria with a temperature of 300 K. The weighted histogram analysis method was used for computing the potential of mean force based upon the umbrella sampling calculations (55). Analysis of subsamples from these simulations indicate that the overall free energy profiles require on the order of 5 ns to equilibrate; thus the first 5 ns of each window is discarded in the weighted histogram analysis method analysis presented here.

4.6 Figures



Figure 4.1 Structure of the SrtB-NPQT* complex.

(a) a comparison of the chemical bonds that join the peptide substrate to the SrtB enzyme in the SrtB-NPQT* complex (*left panel*) and the SrtB-NPQT thioacyl catalytic intermediate (*right panel*). Atoms from SrtB are colored *red*. (b) ribbon diagram of the SrtB-NPQT* complex. *Green*, $\beta 2/\beta 3$ loop; *orange*, $\beta 6/\beta 7$ loop; *red*, $\beta 7/\beta 8$ loop. Active site residues and substrate analog are shown as *sticks*. (c) surface representation of SrtB in the complex utilizing the same color scheme as in (b). Active site residue Arg233 is highlighted in *blue*, and Cys223 is in *yellow*.

Crystallographic data collection and refinement statistics Table 4.1

Parameters	SrtB-NPQT*
Data Collection	
Space Group	P2 ₁
Cell Dimensions	
<i>a, b, c</i> (Å)	102.6, 59.12, 72.49
Resolution (Å)	72.49-2.49 (2.58-
	2.49)
$R_{ m merge}$	26.6 (212.4)
$I/\sigma(I)$	3.8 (0.7)
Completeness (%)	92.2 (94.2)
Multiplicity	5.1 (5.0)
CC1/2	98.6 (50.4) ^a
CC*	99.6 (81.8) ^a
Refinement	
Resolution (Å)	72.49-2.49
No. Reflections	28147
$R_{\rm work}/R_{\rm free}$	21.5/27.1
CC _{work}	94.4 (76.4) ^a
CCfree	87.7 (75.5) ^a
No. atoms	
Protein	7248
Ligand/ion	176
Water	27
B factors	
Mean	72.5
Wilson	56.2
Protein	72.3
Ligand/ion	87.6
Water	29.9
RMSD	
Bond lengths (Å)	0.010
Bond angles (°)	1.26

Values in parentheses are for highest-resolution shell. ^a Values reported as a percentage (%)



Figure 4.2 Structure and interactions of the NPQT* modifier in complex with SrtB. (a) $F_0 - F_c$ map contoured at 3 σ . Electron density (*gray mesh*) was generated by removing the NPQT* peptide from the final model and repeating refinement. The map shown is an average of

the density from all four chains in the asymmetric unit. The peptide extends from the active site cysteine. (b) stereo view of the NPQT* peptide (*gray sticks*) and interacting residues (*blue sticks*). Hydrogen bonds are indicated by *dashed green lines*. (c) diagram of the interactions between SrtB (*blue*) and the NPQT* peptide (*gray*). Hydrogen bonds are indicated by *dashed green lines*. SrtB residues that make only hydrophobic contacts are depicted as *blue circles* positioned near their most significant point of contact to the peptide.



Figure 4.3 Expanded view of the active site in the energy-minimized model of the SrtB-NPQT thioacyl intermediate.

Interactions between SrtB and the threonine residue in the sorting signal are shown. The thioacyl carbonyl oxygen atom is positioned to accept hydrogen bonds from the η -nitrogen atom of Arg233 and the backbone nitrogen atom of Glu224. The side chain of Arg233 is held in position by a hydrogen bond between its ϵ -nitrogen atom and the hydroxyl group on the Thr residue located at the P1 position of the sorting signal.



Figure 4.4 Transpeptidation activity of wild-type and mutant SrtB.

(a) representative HPLC chromatograms showing the reaction products that are produced when SrtB was incubated with SNKDKVENPQTNAGT (sorting signal in bold type) and GGGGG peptides that mimic its sorting signal and secondary lipid II substrates, respectively. Reactions performed in the presence (*left panel*) and absence (*right panel*) of SrtB are shown. Only when SrtB is present (*left panel*) is the appropriate transpeptidation peptide product produced SNKDKVENPQTGGGGG). (b) transpeptidation activity of SrtB mutants. The indicated SrtB mutant was incubated with 200 μ M peptide containing an NPQTN sorting signal and pentaglycine and monitored by HPLC as described above. The *dark shaded bars* indicate the amount of full-length peptide remaining after reaction with the enzyme. The *lightly shaded barr* indicates the amount of transpeptidation product that was formed. An *asterisk* indicates that no transpeptidation product could be detected after 24 h. The *error bars* represent the standard deviation of three reactions. (c) transpeptidation activity of sorting signal amino acid mutants. Sorting signal peptides containing select alanine substitutions were assayed for their ability to be utilized by SrtB as a substrate and monitored by HPLC as described above. The amount of transpeptidation product formed for each mutant peptide is expressed as a percentage of the amount formed from reaction of SrtB with the native NPQTN sorting signal.



Figure 4.5 Alignment of SrtB-NPQT* and SrtA-LPAT* (24). SrtB is shown as *blue ribbons*, and SrtA is shown as *green ribbons*.



Figure 4.6 Structures and free energy profiles of the SrtA and SrtB thioacyl complexes. (a) selected structures obtained from MD simulations of the SrtA and SrtB thioacyl complexes. The left and middle panels show structures of the SrtA thioacyl complex in which the threonine side chain in the sorting signal either interacts with the active site arginine (*left panel*, SrtA-LPAT Thr-"In") or projects away from the active site (*middle panel*, SrtA-LPAT Thr-"Out"). The *right panel* shows the structure of the SrtB-NPQT thioacyl complex in which the threonine side chain interacts with the arginine residue (Thr-in). The structures are displayed with the SrtA and SrtB surfaces colored *green* and *blue*, respectively. Surface representations of labeled active site residues were calculated independently of the remaining protein surface, which allows for the visualization of SrtB His130 (*cyan*) behind Leu96 (*dark gray*) in the right panel, even though this residue would not normally be considered solvent-accessible in this conformation. (**b**–**d**) free

energies were calculated from positions sampled during a Hamiltonian replica exchange simulation for the SrtA-LPAT (b), SrtB-NPQT (c), and SrtB-NPAT (d) thioacyl complexes. The *x* axis records the position of P1 and P2 residues in the bound sorting signal relative to the active site cysteine residue. The *y* axis describes the positioning of residues P3 and P4 in the sorting signal relative to the body of the protein.



Figure 4.7 The SrtB transpeptidation mechanism showing how the sorting signal may stabilize the oxyanion hole.

(a) the SrtB active site with His130 in its imidazolium form and Cys223 in its thiolate form. (b) the NPQTN substrate binds with its P1 Thr residue in the in position within the active site. The Cys223 thiolate performs a nucleophilic attack on the P1 Thr carbonyl carbon. (c) the first tetrahedral intermediate is formed. The P1 Thr residue interacts with Arg233 to construct a substrate-stabilized oxyanion hole in which the side chain of Arg233 and the backbone amide of Glu224 hydrogen bond to the oxyanion. The His130 imidazolium group donates a proton to the leaving group to complete breakage of scissile bond. (d) the P1 Thr residue maintains hydrogen bonds with Arg233 to stabilize its interaction with the thioacyl intermediate. (e) the incoming GGGGG peptide from lipid II acts as a nucleophile that attacks the carbonyl carbon of the thioacyl bond. (f) the second tetrahedral intermediate is formed and is again stabilized by the substrate-stabilized oxyanion hole. This intermediate then collapses, releasing the NPQTGGGGG transpeptidation product and returning the enzyme to its active form (a).

4.7 References

- Spirig, T., Weiner, E. M., and Clubb, R. T. (2011) Sortase enzymes in Gram-positive bacteria. *Mol. Microbiol.* 10.1111/j.1365-2958.2011.07887.x
- Clancy, K. W., Melvin, J. A., and McCafferty, D. G. (2010) Sortase transpeptidases: insights into mechanism, substrate specificity, and inhibition. *Biopolymers* 94, 385–396
- Navarre, W. W., and Schneewind, O. (1999) Surface Proteins of Gram-Positive Bacteria and Mechanisms of Their Targeting to the Cell Wall Envelope. *Microbiol. Mol. Biol. Rev.* 63, 174–229
- Proft, T. (2010) Sortase-mediated protein ligation: an emerging biotechnology tool for protein modification and immobilisation. *Biotechnol. Lett.* 32, 1–10
- Popp, M. W.-L., and Ploegh, H. L. (2011) Making and Breaking Peptide Bonds: Protein Engineering Using Sortase. *Angew. Chem. Int. Ed.* 50, 5024–5032
- Tsukiji, S., and Nagamune, T. (2009) Sortase-Mediated Ligation: A Gift from Gram-Positive Bacteria to Protein Engineering. *ChemBioChem* 10, 787–798
- Mazmanian, S. K., Liu, G., Jensen, E. R., Lenoy, E., and Schneewind, O. (2000) Staphylococcus aureus sortase mutants defective in the display of surface proteins and in the pathogenesis of animal infections. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5510–5515
- Mazmanian, S. K., Ton-That, H., and Schneewind, O. (2001) Sortase-catalysed anchoring of surface proteins to the cell wall of Staphylococcus aureus. *Mol. Microbiol.* 40, 1049– 1057

- Weiss, W. J., Lenoy, E., Murphy, T., Tardio, L., Burgio, P., Projan, S. J., Schneewind,
 O., and Alksne, L. (2004) Effect of srtA and srtB gene expression on the virulence of
 Staphylococcus aureus in animal models of infection. *J. Antimicrob. Chemother.* 53, 480–486
- Maresso, A. W., and Schneewind, O. (2006) Iron acquisition and transport in Staphylococcus aureus. *Biometals Int. J. Role Met. Ions Biol. Biochem. Med.* 19, 193–203
- Maresso, A. W., Chapa, T. J., and Schneewind, O. (2006) Surface protein IsdC and sortase B are required for heme-iron scavenging of Bacillus anthracis. *J. Bacteriol.* 188, 8145–8152
- Mazmanian, S. K., Ton-That, H., Su, K., and Schneewind, O. (2002) An Iron-Regulated Sortase Anchors a Class of Surface Protein during Staphylococcus aureus Pathogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 99, 2293–2298
- Schneewind, O., Model, P., and Fischetti, V. A. (1992) Sorting of protein A to the staphylococcal cell wall. *Cell* 70, 267–281
- 14. Frankel BA, Kruger RG, Robinson DE, Kelleher NL, McCafferty DG (2005)
 Staphylococcus aureus sortase transpeptidase SrtA: insight into the kinetic mechanism and evidence for a reverse protonation catalytic mechanism. *Biochemistry* 44:11188– 11200.
- Huang, X., Aulabaugh, A., Ding, W., Kapoor, B., Alksne, L., Tabei, K., and Ellestad, G.
 (2003) Kinetic mechanism of Staphylococcus aureus sortase SrtA. *Biochemistry* 42, 11307–11315

- Perry, A. M., Ton-That, H., Mazmanian, S. K., and Schneewind, O. (2002) Anchoring of surface proteins to the cell wall of Staphylococcus aureus. III. Lipid II is an *in vivo* peptidoglycan substrate for sortase-catalyzed surface protein anchoring. *J. Biol. Chem.* 277, 16241–16248
- Ruzin, A., Severin, A., Ritacco, F., Tabei, K., Singh, G., Bradford, P. A., Siegel, M. M.,
 Projan, S. J., and Shlaes, D. M. (2002) Further evidence that a cell wall precursor [C(55)MurNAc-(peptide)-GlcNAc] serves as an acceptor in a sorting reaction. *J. Bacteriol.* 184, 2141–2147
- Schneewind, O., Fowler, A., and Faull, K. F. (1995) Structure of the cell wall anchor of surface proteins in Staphylococcus aureus. *Science* 268, 103–106
- Marraffini, L. A., and Schneewind, O. (2005) Anchor Structure of Staphylococcal Surface Proteins V. ANCHOR STRUCTURE OF THE SORTASE B SUBSTRATE IsdC. J. Biol. Chem. 280, 16263–16271
- Mazmanian, S. K., Skaar, E. P., Gaspar, A. H., Humayun, M., Gornicki, P., Jelenska, J., Joachmiak, A., Missiakas, D. M., and Schneewind, O. (2003) Passage of Heme-Iron Across the Envelope of Staphylococcus aureus. *Science* 299, 906–909
- Ton-That, H., Mazmanian, S. K., Alksne, L., and Schneewind, O. (2002) Anchoring of surface proteins to the cell wall of Staphylococcus aureus. Cysteine 184 and histidine 120 of sortase form a thiolate-imidazolium ion pair for catalysis. *J. Biol. Chem.* 277, 7447–7452

- Frankel BA, Tong Y, Bentley ML, Fitzgerald MC, McCafferty DG (2007) Mutational analysis of active site residues in the Staphylococcus aureus transpeptidase SrtA.
 Biochemistry 46:7269–7278.
- Bentley, M. L., Lamb, E. C., and McCafferty, D. G. (2008) Mutagenesis Studies of Substrate Recognition and Catalysis in the Sortase A Transpeptidase from Staphylococcus aureus. *J. Biol. Chem.* 283, 14762–14771
- Suree, N., Liew, C. K., Villareal, V. A., Thieu, W., Fadeev, E. A., Clemens, J. J., Jung, M. E., and Clubb, R. T. (2009) The Structure of the Staphylococcus aureus Sortase-Substrate Complex Reveals How the Universally Conserved LPXTG Sorting Signal Is Recognized. *J. Biol. Chem.* 284, 24465–24477
- Tian, B.-X., and Eriksson, L. A. (2011) Catalytic Mechanism and Roles of Arg197 and Thr183 in the Staphylococcus aureus Sortase A Enzyme. *J. Phys. Chem. B* 115, 13003– 13011
- Liew, C. K., Smith, B. T., Pilpa, R., Suree, N., Ilangovan, U., Connolly, K. M., Jung, M. E., and Clubb, R. T. (2004) Localization and mutagenesis of the sorting signal binding site on sortase A from Staphylococcus aureus. *FEBS Lett.* 571, 221–226
- 27. Marraffini, L. A., Ton-That, H., Zong, Y., Narayana, S. V. L., and Schneewind, O. (2004) Anchoring of surface proteins to the cell wall of Staphylococcus aureus. A conserved arginine residue is required for efficient catalysis of sortase A. *J. Biol. Chem.* 279, 37763–37770

- Zhang, R., Wu, R., Joachimiak, G., Mazmanian, S. K., Missiakas, D. M., Gornicki, P., Schneewind, O., and Joachimiak, A. (2004) Structures of Sortase B from Staphylococcus aureus and Bacillus anthracis Reveal Catalytic Amino Acid Triad in the Active Site. *Structure* 12, 1147–1156
- 29. Weiner, E. M., Robson, S., Marohn, M., and Clubb, R. T. (2010) The Sortase A enzyme that attaches proteins to the cell wall of Bacillus anthracis contains an unusual active site architecture. *J. Biol. Chem.* **285**, 23433–23443
- Race, P. R., Bentley, M. L., Melvin, J. A., Crow, A., Hughes, R. K., Smith, W. D., Sessions, R. B., Kehoe, M. A., McCafferty, D. G., and Banfield, M. J. (2009) Crystal Structure of Streptococcus pyogenes Sortase A. J. Biol. Chem. 284, 6924–6933
- Ilangovan, U., Ton-That, H., Iwahara, J., Schneewind, O., and Clubb, R. T. (2001) Structure of sortase, the transpeptidase that anchors proteins to the cell wall of Staphylococcus aureus. *Proc. Natl. Acad. Sci.* 98, 6056–6061
- 32. Khare, B., Krishnan, V., Rajashankar, K. R., I-Hsiu, H., Xin, M., Ton-That, H., and Narayana, S. V. (2011) Structural Differences between the Streptococcus agalactiae Housekeeping and Pilus-Specific Sortases: SrtA and SrtC1. *PLoS ONE* 6, e22995
- 33. Zong, Y., Mazmanian, S. K., Schneewind, O., and Narayana, S. V. L. (2004) The Structure of Sortase B, a Cysteine Transpeptidase that Tethers Surface Protein to the Staphylococcus aureus Cell Wall. *Structure* 12, 105–112

- Jung, M. E., Clemens, J. J., Suree, N., Liew, C. K., Pilpa, R., Campbell, D. O., and Clubb, R. T. (2005) Synthesis of (2R,3S) 3-amino-4-mercapto-2-butanol, a threonine analogue for covalent inhibition of sortases. *Bioorg. Med. Chem. Lett.* 15, 5076–5079
- 35. Kabsch, W. (2010) XDS. Acta Crystallogr. D Biol. Crystallogr. 66, 125–132
- Karplus, P. A., and Diederichs, K. (2012) Linking Crystallographic Model and Data Quality. *Science* 336, 1030–1033
- Diederichs, K., and Karplus, P. A. (2013) Better models by discarding data? *Acta Crystallogr. D Biol. Crystallogr.* 69, 1215–1222
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674
- Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501
- Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics.
 Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132
- Smart, O. S., Womack, T. O., Flensburg, C., Keller, P., Paciorek, W., Sharff, A.,
 Vonrhein, C., and Bricogne, G. (2012) Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D Biol. Crystallogr.* 68, 368–380
- 42. Bricogne, G. (1993) Direct phase determination by entropy maximization and likelihood ranking: status report and perspectives. *Acta Crystallogr. D Biol. Crystallogr.* **49**, 37–60

- 43. Kruger, R. G., Dostal, P., and McCafferty, D. G. (2004) Development of a highperformance liquid chromatography assay and revision of kinetic parameters for the Staphylococcus aureus sortase transpeptidase SrtA. *Anal. Biochem.* **326**, 42–48
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26, 1781–1802
- Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78, 1950–1958
- 46. Kräutler, V., van Gunsteren, W. F., and Hünenberger, P. H. (2001) A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem.* **22**, 501–508
- Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J. Chem. Phys. 98, 10089–10092
- Wang, J., Wang, W., Kollman, P. A., and Case, D. A. (2006) Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* 25, 247–260
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004)
 Development and testing of a general amber force field. *J. Comput. Chem.* 25, 1157–1174

- Feller, S. E., Zhang, Y., Pastor, R. W., and Brooks, B. R. (1995) Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* 103, 4613– 4621
- Martyna, G. J., Tobias, D. J., and Klein, M. L. (1994) Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* 101, 4177–4189
- 52. Schrödinger, L. (2010) The PyMOL Molecular Graphics System, Version 1.3r1.
- D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Goetz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman (2012) *AMBER 12*, University of California, San Francisco
- Jiang, W., Luo, Y., Maragliano, L., and Roux, B. (2012) Calculation of Free Energy Landscape in Multi-Dimensions with Hamiltonian-Exchange Umbrella Sampling on Petascale Supercomputer. J. Chem. Theory Comput. 8, 4672–4680
- 55. Souaille, M., and Roux, B. (2001) Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.*135, 40–57
- Schechter, I., and Berger, A. (1967) On the size of the active site in proteases. I. Papain.
 Biochem. Biophys. Res. Commun. 27, 157–162
- 57. Bentley, M. L., Gaweska, H., Kielec, J. M., and McCafferty, D. G. (2007) Engineering the substrate specificity of Staphylococcus aureus Sortase A. The beta6/beta7 loop from SrtB confers NPQTN recognition to SrtA. J. Biol. Chem. 282, 6571–6581
- Piotukh, K., Geltinger, B., Heinrich, N., Gerth, F., Beyermann, M., Freund, C., and Schwarzer, D. (2011) Directed Evolution of Sortase A Mutants with Altered Substrate Selectivity Profiles. J. Am. Chem. Soc. 133, 17536–17539
- 59. Mariscotti, J. F., García-del Portillo, F., and Pucciarelli, M. G. (2009) The Listeria monocytogenes sortase-B recognizes varied amino acids at position 2 of the sorting motif. *J. Biol. Chem.* **284**, 6140–6146
- Comfort, D., and Clubb, R. T. (2004) A Comparative Genome Analysis Identifies
 Distinct Sorting Pathways in Gram-Positive Bacteria. *Infect. Immun.* 72, 2710–2722
- 61. Kruger RG et al. (2004) Analysis of the Substrate Specificity of the Staphylococcus aureus Sortase Transpeptidase SrtA. *Biochemistry* 43:1541–1551.
- Nicola, G., Peddi, S., Stefanova, M., Nicholas, R. A., Gutheil, W. G., and Davies, C.
 (2005) Crystal Structure of Escherichia coli Penicillin-Binding Protein 5 Bound to a Tripeptide Boronic Acid Inhibitor: A Role for Ser-110 in Deacylation. *Biochemistry* 44, 8207–8217
- 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) Staphylococcus aureus Sortase A transpeptidase. Calcium promotes sorting signal binding by altering the mobility and structure of an active site loop. *J. Biol. Chem.* 281, 1817–1826

162

64. Connolly, K. M., Smith, B. T., Pilpa, R., Ilangovan, U., Jung, M. E., and Clubb, R. T.
(2003) Sortase from Staphylococcus aureus Does Not Contain a Thiolate-Imidazolium Ion Pair in Its Active Site. *J. Biol. Chem.* 278, 34061–34065 Chapter 5

The "lid" of S. pneumoniae SrtC1 is not dynamic in solution and

regulates substrate access to the active site

The work described in this chapter is a version of a manuscript to be submitted for publication. Jacobitz A. W., Yi, S. W., McConnell, S., Peterson, R., Jung, M. E., and Clubb, R. T. 2015.

I was the primary author of this work. I designed experiments, performed biochemical assays and NMR experiments, and analyzed data.

5.1 Overview

Sortase enzymes perform a transpeptidation reaction that covalently links their two substrates via a new peptide bond. While most sortase enzymes are known for catalyzing this transpeptidation between their primary protein substrate and the cell wall peptidoglycan, a subset of these enzymes instead covalently links proteins to each other via isopeptide bonds to produce elongated protein polymers known as pili. These protein-polymerizing class C sortases are structurally distinguished by the presence of an N-terminal extension that occludes the active site. Numerous publications have hypothesized that the lid is mobile in solution as this would be necessary for the binding and catalysis of the enzyme's substrate. Here we show using NMR dynamics measurements and in vitro assays that the lid of wild-type Streptococcus pneumoniae SrtC1 is rigid in solution and does not experience conformational dynamics on a mechanistically relevant timescale. Additionally, we show that point mutations to the lid induce dynamic behavior that correlates with an increase in both the hydrolysis of the primary substrate and general access to the active site cysteine residue as evidenced by increased oxidation. These results suggest that the lid of the S. pneumoniae SrtC1 enzyme performs a regulatory function, and they imply that interaction with other regions of the full length protein substrate, or other as yet unnamed surface factors, are likely required to activate the enzyme in vivo.

5.2 Introduction

Gram-positive bacteria commonly utilize their thick cell walls as a scaffold for anchoring proteins necessary for interacting with their environment. Many of these proteins are covalently attached to the cell wall through a reaction catalyzed by sortase enzymes. A particular subtype of sortases, the class C sortases, are not only implicated in the attachment of surface proteins to the cell wall of Gram-positive organisms but have also been shown to polymerize pili: long filamentous protein fibers that are used by some bacteria to make initial, long-range contacts to host cells^{1–3}. These pili are thought to provide a preliminary point of attachment to the host cell which acts like a grappling hook, allowing the bacterium to swing closer to the cell, thereby facilitating the formation of additional close-range adhesive interactions via auxiliary surface anchored adhesive proteins³. Pili have subsequently been shown to increase the virulence of certain bacterial species^{3–5}, participate in immune system evasion⁶, are implicated in biofilm formation^{7,8} and have also been heavily researched for their potential as vaccine components due to their ability to elicit an immune response^{9–11}.

Sortase enzymes are ubiquitous in Gram-positive bacteria and can be classified into 5 distinct classes based on their sequence and structural similarities^{12,13}. Classes A, B, D, E, and F are generally considered to be cell wall anchoring sortases, while the class C sortases stand alone in their ability to polymerize their protein substrates (it should be noted that there have recently been two sortases which have been structurally defined as class B that are in fact pilus associated sortases)^{14,15}. The cell wall anchoring sortases all recognize lipid II, a precursor to cell wall synthesis, as their secondary substrate. Class A enzymes are often referred to as "housekeeping" enzymes; they are constitutively expressed, and generally anchor a variety of proteins to the cell wall. Class B, D, E, and F enzymes on the other hand are generally utilized for attaching a

particular subclass of proteins to the cell wall and are typically expressed along with their substrate(s) only under specific conditions. The class C enzymes are similar to the class B, D, E and F enzymes in that they are also sequestered in specific genomic islands along with their pilus protein substrates, but they are unique in that most class C enzymes are known to recognize two distinct positions on a single protein to polymerize it into a pilus instead of utilizing lipid II as their secondary substrate. When fully polymerized, these pili can contain hundreds of individual pilin protein monomers and be over 1 micron in length.

A large number of sortase structures have been solved and they all share a common $8-\beta$ barrel sortase fold and a His-Cys-Arg catalytic triad^{13,16}. Each of these enzymes recognizes a Cterminal cell wall sorting signal (CWSS) composed of a 5 residue sorting signal motif, followed by a transmembrane helix and positively charged cytoplasmic anchoring domain. The 5 residue sorting signal is generally of the form LPXTG, where the Pro and Thr residues are most highly conserved, X can be any amino acid, and the first and last residue often confer specificity to the individual sortase^{12,13,17}. The reaction begins when the sortase binds the sorting signal and the active site cysteine performs a nucleophilic attack on the threonine's carbonyl carbon. This produces a negatively charged tetrahedral intermediate which is thought to be stabilized by the active site arginine¹⁷, and the active site histidine is thought to act as an acid to protonate the terminal amine on the glycine of the leaving group^{18,19}. The tetrahedral intermediate then quickly collapses to form a semi-stable thioacyl intermediate between the active site cysteine and the protein substrate. For the more heavily studied cell wall anchoring sortases, the next step of the reaction requires the binding of cell wall precursor lipid II. The cross-bridge peptide of the lipid II molecule varies from organism to organism, but will invariably contain a terminal amine group which is deprotonated by the active site histidine before performing a nucleophilic attack on the Thr carbonyl carbon, leading to a second tetrahedral intermediate which collapses to form a new peptide bond between the sorting signal threonine and the cross-bridge peptide¹³. Pilin polymerizing sortases follow the same basic mechanism but utilize the amino-nucleophile of a lysine sidechain within their pilin protein substrate to resolve the thioacyl intermediate and generate an isopeptide bond between the sorting signal of one monomer and the lysine of another^{2,3,10}. That is, the first of these two recognition motifs is the same CWSS at the Cterminus of the protein substrate, but the other is typically a 4 residue lysine-containing pilin motif located in the N-terminal domain of the protein³.

Several species of bacteria are known to harbor pilus gene clusters typically containing genes for 2 or 3 pilin proteins and 1 to 3 sortases. Interestingly, these class C sortases are the only known class of sortase to harbor an N-terminal extension known as the "lid" that completely occludes the active site^{13,20–22}(**Fig. 1**). The lid region has been found in all class C sortases studied to date and invariably maintains a conserved DP(F/W/Y) anchor motif wherein the aspartate residue forms a hydrogen bond with the active site arginine, and the aromatic at the end of the motif is wedged into the active site where it forms a sulfur-aromatic interaction with the active site cysteine²². Based on the fact that the lid completely blocks access to the sorting signal binding site seen in all sortase-substrate complex structures reported to date^{17,23,24} along with the observations that several class C sortase crystal structures are missing electron density for regions flanking the lid anchor residues^{25–28} and many display elevated B-factors throughout this region^{20–22}, it has been proposed that the lid region of these sortases must be highly flexible^{13,20–22,27}.

Streptococcus pneumoniae is one of the most common causative agents of pneumonia, and is also known to cause meningitis, and septicemia, and the ability of *S. pneumoniae* to cause invasive disease is enhanced by the presence of pili⁴. *S. pneumoniae* can express two distinct types of pili carried on separate genomic islands, known as pilus island 1 (PI-1) or pilus island 2 (PI-2)²⁹. PI-1 is present in ~30% of *S. pneumoniae* strains and these pili have been more thoroughly characterized. They consist of three pilus proteins, RrgA, RrgB, and RrgC, which are polymerized by three sortases, SpnSrtC1, SpnSrtC2, and SpnSrtC3. Structures for each of these sortases have been solved and they all display nearly identical structures, each with the conserved lid occluding the active site^{20,25}. Mutations to the anchor residues in the lid of SpnSrtC1 have been shown to decrease the thermal stability of the enzyme *in vitro*²¹. These same mutations were shown to alter the pattern of, but not abrogate, the production of polymerized RrgB species *in vitro*²⁰, but similar mutations to either *Streptococcus agalactiae* sortase C-1 (SagSrtC1) or *Actinomyces oris* sortase C-1 (AoSrtC1) had no visible effect on pilus polymerization *in vivo*^{30,31}.

Here we show that despite a number of papers previously referring to the lid region of SrtC enzymes as flexible, the lid of SpnSrtC1 is not in fact dynamic in solution, but instead has two dynamic regions flanking the lid which may function as hinges, as has been proposed previously by Manzano and coworkers²⁰. Mutations to the anchor residues of the lid make the region highly dynamic and dramatically increase hydrolysis of a sorting signal peptide *in vitro*. We also show that the lid protects the active site cysteine from oxidation, and mutations to the lid abolish this protective effect. Based on these findings, we propose that the lid performs a regulatory function to prevent access to the active site in the absence of additional factors, likely

found on the cell surface, which prepare the protein for activity by mobilizing or altering the conformation of the lid.

5.3 Results

5.3.1 Initial NMR analysis indicates SpnSrtC1 forms dimers in solution.

Soluble SpnSrtC1, residues 17-228 was uniformly labeled with ¹⁵N and ¹³C and purified from E. coli. 94% of backbone resonances could be definitively assigned using standard triple resonance assignment methods, leaving a few 1-2 residue unassigned segments, and a single 7 residue stretch of amino acids (residues Q167-E172) which could not be assigned. Interestingly, this 7 residue stretch of amino acids forms the majority of the interface seen between the two SpnSrtC1 monomers in the asymmetric unit of the 2008 crystal structure solved by Manzano and coworkers²⁰. Since residues participating in binding interactions on the intermediate exchange timescale (i.e. binding and dissociation occurs roughly on the order of 1 ms) are frequently broadened beyond detection, we hypothesized that these resonances could be missing due to the existence of a transient dimerization occurring between SrtC1 monomers in solution. To determine whether this crystallographic dimer was indeed present in solution, we performed a series of dilutions from 2 mM to 100 μ M and recorded ¹H-¹⁵N HSOC spectra. As the concentration of the protein was lowered, significant changes in the spectrum were seen in isolated resonances, along with the appearance of several additional peaks which could only be detected in the lowest concentration sample (Fig. 2). Although new resonances which appear in the 100 μ M spectrum could not be assigned, as this concentration is too low for the use of significantly less sensitive triple resonance experiments that are required for backbone assignments, the existing assigned resonances which either experienced an increase in intensity

on dilution, or a change in chemical shift, could be identified and plotted onto the crystal structure (**Fig 2B**). These residues are largely clustered in the binding interface observed in the crystal structure, indicating that this dimer is also present in solution.

Additionally, we utilized T_1 and T_2 relaxation measurements to calculate the overall tumbling time for the protein, τ_c , which is defined as the amount of time it takes for the protein to rotate through 1 radian in space. τ_c values of known globular proteins consistently yield values of ~0.6 ns times the molecular weight (MW) of the protein in kilodaltons. The MW of soluble SpnSrtC1 is 23.813 kDa yielding a theoretical τ_c of 14.3 ns. Based on measurements of the T₁ and T₂ relaxation constants, we calculated the τ_c of SpnSrtC1 in solution to be 17.8 ns at a concentration of 1 mM. This τ_c is significantly higher than would be expected for a monomer of this size. If the protein formed a stable dimer, the τ_c would be expected to be around 28.6 ns, and a protein of this size would experience extremely fast T₂ relaxation preventing the detection of good quality triple resonance spectra, an effect we did not observe. The slightly elevated observed τ_c value is instead likely a consequence of the intermediate exchange kinetics of dimerization for this protein that had been suggested by the broadening of interfacial residues discussed above. This would indicate that at any given time, only a small percentage of the available proteins in solution would be dimeric and tumble at the slower ~28.6 ns rate, but that these dimers are still detectable as an increase in the average τ_c of all the proteins in the sample, leading to a slightly higher measured τ_c than expected.

5.3.2 NMR dynamics data indicate that the SpnSrtC1 lid is not dynamic *in vitro*.

In order to determine whether the lid of SpnSrtC1 is mobile in solution, we performed dynamics experiments using solution NMR spectroscopy. NMR spectroscopy is a powerful technique for dissecting the motions of individual backbone amides across a wide range of biologically relevant timescales. First, T₁, T₂, and HetNOE experiments were performed to detect the presence of fast-motions on the ps-ns timescale. The results of T_1 and T_2 experiments were used to determine the rotational correlation time, τ_e , for each backbone amide residue in the protein (Fig. 3). τ_e represents the amount of time it takes for a single amide bond vector to rotate through 1 radian of space. By comparing the τ_e to the overall rotational correlation time for the entire protein, τ_c , (Fig.3A, red line) residues whose bond vectors are moving more rapidly than the body of the protein can be detected. Any residue that rapidly samples disordered states in solution would be expected to have a τ_e that is much lower than the protein's τ_c , indicating that the residue can move through space more quickly than the body of the protein. This is comparable to a flag rapidly flapping in the wind (the mobile residues) while affixed to the bow of a slowly rocking ship (the body of the protein). The flag moves on the same slow timescale of the ship, but will experience faster overall motion because of additional dynamic behavior. Unexpectedly, lid residues contacting the active site (D58 and W60) have similar τ_e values as the entire structured core domain of the protein, 16.4 and 16.9 ns respectively, compared to a τ_c of 17.8 ± 1.1 ns for the structured portion of the protein. For comparison, the unstructured Cterminus of the protein has τ_e values that drop as low as 10 ns. This is confirmed by HetNOE values above 0.6 which indicates that these residues are structured in solution (Fig. 3B), as residues are typically only considered disordered if their HetNOE is below 0.6 (Fig. 3B, red line). Notably, residues 56-57 before and 65-70 after the lid show increased dynamics on this timescale indicating that they are highly mobile, even though the lid residues from this Nterminal extension that are actually engaged with the active site are not. This supports the hypothesis that the N-terminal extension contains "hinges," originally proposed by Manzano and

coworkers to be present in SpnSrtC3 based on elevated B-factors for this protein being highest in two discrete regions flanking the lid²⁰. This hypothesis is further supported by the significant number of SrtC structures solved to date wherein these regions either showcase the highest B-factors in the structure, or lack density altogether^{20–22,25–28}. Together these NMR experiments indicate that the lid residues that contact the active site do not experience motions over the ps-ns timescale, indicating that they are predominantly in a rigid, ordered conformation, likely akin to that visualized in the crystal structure, whereas regions of sequence adjacent to the lid are in fact highly mobile, and likely represent hinges for the opening of the lid and subsequent unmasking of the active site.

As both of these experiments detect motions on the ps-ns timescale, we conducted additional experiments to determine if the lid was instead mobile over longer μ s-ms timescales. To measure this phenomenon, we first conducted CPMG experiments³² to directly detect the R_{ex} component of relaxation on a per residue basis (**Fig. 3C**). R_{ex} values are generally considered to be an indicator of this slower timescale motions if they are greater than ~30% of the measured T₂ for an individual residue which typically means values above 10 s⁻¹ are considered significant. Residues in and around the lid region do not show significantly elevated R_{ex} values, indicating that they are in fact not mobile on the μ s-ms timescale. Additional experiments to sample within this timescale (CEST), and out to the ms-s timescale (NZ-exchange) were also conducted, and showed no spectral perturbations which would be indicative of longer timescale motions (data not shown). Together, these experiments indicate that the lid of SpnSrtC1 is not dynamic in solution as has been suggested by previous reports and is instead largely structured in the absence of additional factors, even out to very long timescales.

5.3.3 Mutations to the "lid" of SpnSrtC1 make the region highly dynamic in solution.

Given that the "lid" of wild-type SpnSrtC1 is not dynamic in vitro, we set out to determine the effects of point mutations in the lid on its mobility in solution. Several previous studies have shown that point mutations to the conserved Asp and Trp residues in the DP(F/W/Y) motif of SpnSrtC1 lowered its thermostability²¹. Similar mutations were also shown to increase the rate of hydrolysis and were even necessary to induce *in vitro* pilin polymerization by SagSrtC1^{30,33}. To determine if these mutations modified the mobility of the lid, we generated SpnSrtC1-D58A and W60A point mutants and characterized their effects on the protein's dynamics using the same fast-timescale T_1 , T_2 , and HetNOE experiments discussed above for the WT enzyme. SpnSrtC1-W60A shows significantly increased mobility in the lid, and throughout the entire region between the two hinges identified for the WT enzyme (residues 56-70). T_1 and T_2 experiment derived τ_e values indicate that the region tumbles at 10.7 ns, 5.2 ns faster than the average for the structured portion of this enzyme (Fig. 4). The HetNOE experiment additionally indicates that these residues are largely disordered, with every residue that could be assigned and characterized between the two hinges, residues 58 to 64, having reduced HetNOE values compared to WT, with many of these being less than 0.6 indicating that they are disordered. SpnSrtC1-D58A also shows significantly increased mobility in the lid, and especially in flanking regions based on an analysis of the τ_e and HetNOE measurements (Fig. 4). Interestingly, these experiments show decreased dynamics in a small region immediately after the conserved lid residue, W60. To further investigate whether this less dynamic region indicates that the lid tryptophan might be sampling a closed state similar to the WT enzyme, we performed additional NOESY experiments and compared the spectra for the Trp indole proton. As the NOESY experiment is able to detect the through-space coupling of nuclei within 5 Å of each other, the

pattern of numerous peaks seen for the WT enzyme indicates that the side chain of the residue is folded and in close proximity to several other residues (**Fig. 4E**). The W60 indole in the D58A mutant on the other hand is split into 3 peaks in the HSQC, each of which shares only 1 or 2 NOESY peaks with the WT spectrum that may represent intraresidue NOEs to adjacent protons in the sidechain. This large difference in NOE pattern suggests that the tryptophan sidechain in the D58A mutant no longer participates in native contacts to the enzyme active site, and has likely become dislodged from the position it maintains in the WT enzyme. This suggests that even though several of the residues following W60 do not appear to be dynamic based on HetNOE and τ_e values, the conformation of this entire region has been altered significantly, and the native, closed state of the lid has been disrupted.

5.3.4 Mutation of the lid is required to induce *in vitro* hydrolysis.

Knowing that the SpnSrtC1 lid is rigid in solution and becomes mobile upon mutation, we set out to determine whether this increased mobility would affect the ability of the enzyme to function *in vitro*. To test this, we performed *in vitro* hydrolysis experiments by incubating the sortase with a 15 amino acid peptide derived from RrgB and containing the IPQTG sorting signal sequence for 24 hrs at 37°C and then separating the mixture via HPLC. We then utilized MALDI mass spectrometry to confirm the identities of the resultant peaks in the chromatogram. By separating the substrate peptide from the cleaved product using HPLC and measuring the absorbance at 215 nm we could monitor both the amount of cleaved product peptide formed, and substrate peptide remaining, allowing for a thorough disambiguation of the reaction progress which is not possible with the more commonly used FRET based assay which only reports on the breakage of the peptide bond and not the subsequent release of the thioacyl intermediate from the

enzyme. We conducted the experiments with WT SpnSrtC1, SpnSrtC1-D58A and W60A point mutants used in our NMR studies, as well as an SpnSrtC1-D58A-W60A double mutant and an additional mutant wherein the entire lid containing region, residues 55-69 inclusive, was replaced with a 15 residue glycine-serine linker (SpnSrtC1-GSlid). We used the GSlid construct as an approximation of a completely open state of the lid, since a construct where we completely deleted the N-terminal extension containing the lid proved to be unstable over the timecourse of the reaction. Interestingly, though the WT enzyme has been shown to perform a transpeptidation reaction in vitro with a full length RrgB protein which contains both the primary IPQTG sorting signal substrate, and secondary YPKN pilin motif substrate²⁰, the enzyme does not catalyze the initial cleavage step of the reaction *in vitro* in the absence of its secondary substrate (Fig. 5). It is important to note that not only is the WT enzyme incapable of hydrolyzing the substrate, as evidenced by the lack of detectable product peptide, but it is also incapable of beginning its reaction by performing a nucleophilic attack on the IPQTG motif's Thr carbonyl to catalyze the creation of the thioacyl intermediate. This latter reaction could be visualized in our assay as a reduction in the amount of substrate peptide remaining at the end of the reaction without the appearance of a new product peak, but this was not observed. Instead, the WT SpnSrtC1 reaction is indistinguishable from peptide alone or from a reaction with the C193A active site mutant. The lid mutants on the other hand, which we showed by NMR to have much more mobile lid regions, all catalyze a hydrolysis reaction to a similar extent within the error of the measurement except for the D58A mutant which was slightly slower, perhaps due to partial occlusion of the active site by one or more of the multiple semi-stable positions we detected for its lid Trp residue based on NOESY data. While even the mutant enzymes only process about 15% of their substrate in 24 hrs under the conditions tested, this rate of cleavage is similar to that seen for the WT SrtB

enzyme from *Staphylococcus aureus* (Jacobitz unpublished data) indicating that this could be a fully active form of the enzyme.

5.3.5 A closed lid is important for protecting the active site cysteine from oxidation.

To further test the importance of the closed lid, we performed disulfide bonding experiments utilizing a modified Cbz-IPQT* sorting signal substrate, where Cbz is a carbobenzyloxy protecting group and T* is a modified threonine residue with a thiol group in place of its C-terminus ((2R,3S)-3-amino-4-mercapto-2-butanol) which should allow it to form a disulfide bond with the active site cysteine. This type of modified substrate has been used previously to successfully produce stable enzyme substrate complexes for 3 other sortases^{17,23,24}. indicating the utility of this method. When incubated with the IPQT* substrate at room temperature for 24 hrs, the WT SpnSrtC1 showed no detectable disulfide formation by MALDI-MS (Fig. 6A). When the same experiment was repeated with the SpnSrtC1-W60A mutant, a mass shift of 595 Da (theoretical mass of Cbz-IPQT* = 593.4 Da) could be detected indicating the formation of a disulfide-bonded complex (Fig. 6B). The difference in cysteine availability highlighted by this experiment further indicates that the lid of WT SpnSrtC1 not only prevents unwanted catalytic activity, but actually physically limits access to the active site cysteine, which would be able to form a disulfide bond whether or not the enzyme was in a fully catalytically active state.

5.4 Discussion

Class C sortases are unique among sortases in that they catalyze the covalent linkage of two protein substrates instead of attaching proteins to the cell wall peptidoglycan¹³. This reaction

leads to the polymerization of long filamentous pili which are necessary for adhesion to host cells and biofilm formation. Although the class C sortase enzymes utilize a lysine side chain from their protein substrate as the nucleophile to conclude their reaction instead of the Nterminus of a lipid II cross bridge peptide that other classes employ, the basic reaction is highly similar, perhaps owing to the high degree of conservation within the core β -barrel of the sortase motif. All sortases studied to date utilize the same conserved His-Cys-Arg catalytic triad, and are thought to recognize their substrates using the same binding groove beginning at the $\beta 6/\beta 7$ loop and moving towards the active site. Given the number of similarities shared between members of this family it is perhaps unsurprising that we observed the dimerization of SpnSrtC1 in solution (Fig. 2), as dimerization in solution has already been detected for several other members of the family^{34–36}. Interestingly, while the details of this dimerization event are not fully understood, and each of the proposed binding interfaces is somewhat variable, all of these proteins have utilized residues from the β 4- β 5 loop, on the "back" face of the enzyme, opposite the active site. This dimerization was recently shown to alter the rate of catalysis for S. aureus SrtA in vivo³⁶. It is interesting to note that the vast majority of sequences deposited for SpnSrtC1 that can be identified from a BLAST search with the 279 residue UniProt sequence, Q97SB9, referenced in previous studies contain exactly 21 additional N-terminal residues, and analysis of this full 300 residue sequence with the TMHMM server³⁷ predicts an additional N-terminal transmembrane (TM) helix with >99.9% probability that could not have been predicted from the shortened sequence referenced previously. This predicted N-terminal TM-helix has been shown to be commonplace among other class C sortases and is known to be necessary for determining appropriate membrane localization³⁰. Surprisingly, both N and C-termini from both structures in the proposed dimer visualized in the SpnSrtC1 crystal structure all localize to a single face of the

dimeric assembly which puts the two active sites on opposite ends, each facing solvent. It is possible that this dimer is significant *in vivo* when faced with a full length RrgB substrate that contains both the primary and secondary recognition motifs as these two motifs on the RrgB protein are separated from each other by 95 Å which is significantly longer than the 35 Å that separate the two sortase active sites in the dimer. This indicates that a single tri-domain RrgB protein could feasibly interact with both active sites of the dimer simultaneously, potentially allowing a single sortase dimer to polymerize an entire pilus by simply adding monomers iteratively at one active site then the next without the need for sortases to "line up" to carry out a concerted pilus polymerization as has been suggested previously^{2,10}.

In addition to their functional distinction, class C sortases are also structurally unique in that they exhibit an N-terminal extension which occludes the active site and has thus been termed the lid^{13,20–22}. Numerous publications have referred to this lid as "flexible" based on elevated B-factors or missing density in adjacent regions of crystal structures, or based on the fact that removal of the lid would be necessary for substrates to access the active site^{13,20–22,27}. NMR dynamics data indicates to the contrary that in solution the key lid residues that actually contact the active site are not dynamic, but are instead flanked by dynamic regions which potentially act as hinges for the removal of the lid and unmasking of the active site under the appropriate circumstances (**Fig. 3**). Additional analysis of the ability of the WT enzyme to perform the initial cleavage step of its reaction *in vitro* with purified peptide containing its primary IPQTG sorting signal substrate indicates that this closed conformation of the lid actually prevents catalysis. Only when mutations to the lid are introduced which increase the flexibility of this region, as indicated by reduction in both the τ_e and HetNOE values for these residues (**Fig. 4**), does the enzyme catalyze the initial step of its reaction *in vitro* (**Fig. 5**). The same motion-inducing mutations are

also required to grant access to the cysteine even for the simple production of a disulfide bond, the formation of which does not require access to the rest of the binding site or participation from any of the other active site residues (**Fig. 6**). This indicates that not only does the rigidity of the lid in the WT enzyme prevent the hydrolysis of the enzyme's native substrate in solution, but that it does so by completely preventing access to the active site cysteine instead of simply inactivating the enzyme via some form of conformational manipulation, or obstruction of individual recognition pockets in the binding site.

Numerous publications have confirmed the enzyme's ability to perform this reaction under native conditions in vivo^{25,38,39}, and it has even been shown to perform the native transpeptidation reaction in vitro when presented with full length substrates containing both its primary, IPQTG sorting signal, and secondary, YPKN pilin motif substrates^{20,21}. All other classes of sortase enzymes studied to date have been shown to catalyze the hydrolysis of their primary substrate in vitro when deprived of a secondary substrate with which to perform their native transpeptidation reaction^{28,30,35,40-43}. Since SpnSrtC1 cannot perform the initial cleavage of its primary binding motif in vitro in the absence of secondary substrate as numerous other sortases have been shown to do (without the artificial mobilization of its lid through the incorporation of destabilizing mutations) it follows that the recognition of an additional factor must be required to remove the lid and allow for catalysis. Interestingly, our in vitro data also suggests that in the absence of a secondary substrate or other additional factors, WT SpnSrtC1 does not even form a thioacyl intermediate, as this would be detectable as a loss of substrate, without an accompanying formation of hydrolysis product (Fig. 5). This indicates that the thioacyl intermediate reported previously from in vitro reactions with full length protein substrates can only form when the lid is opened by additional interactions with the substrate that

are outside of the sorting signal²¹. It is tempting to assume that YPKN pilin motif would fulfill the role of this additional factor that unhinges the lid, but simply based on the fact that this enzyme has been shown to catalyze the transpeptidation *in vitro* when provided with a full length substrate, it is only possible to conclude that there is another region of the RrgB substrate protein outside of the IPQTG sorting signal that activates the sortase, presumably by opening the lid. Unfortunately even this conclusion is not infallible, as the data for the *in vitro* reaction is not entirely convincing given that we, and others, have observed soluble pilin proteins to associate into stable higher molecular weight species which closely mimic product of the sortase catalyzed polymerization reaction (unpublished data, and ¹⁵). Additionally, the WT S. agalactiae SrtC1 protein was shown to cleave its primary substrate in vitro, but still does not perform the transpeptidation reaction in the presence of a second pilin motif containing peptide. Mutations to the lid of this protein were shown to increase the rate of peptide hydrolysis in vitro by ~2.5 fold, and had the additional activating effect of permitting transpeptidation in solution^{30,33}. The fact that both the S. pneumoniae system studied in this work and the S. agalactiae system studied previously are not fully active *in vitro* with an intact lid reiterates the necessity for additional factors, outside of the two well-known substrate motifs, to open the lid of the enzyme and allow catalysis to proceed as it does in vivo. Sortase enzymes have been observed to colocalize with a number of factors on the cell surface, including the SEC translocon, cell wall synthesis machinery, and their substrates^{13,44-46}, and we postulate that either additional regions within the pilin proteins themselves, or one of these additional surface factors must be necessary for removing the lid and activating the enzyme *in vivo*.

From this evidence it is clear that the lid present in class C enzymes is competent in preventing the reaction from taking place. But why would an entire class of enzymes have

evolved a conserved feature to prevent or retard their primary function? First and foremost, the hydrolysis and subsequent loss of protein substrates to solvent, which are intended to instead be polymerized and attached to the cell surface, is inherently wasteful for the cell. While accidental cleavage and loss of a few proteins here and there (as may be common for cell wall anchoring sortases) might be acceptable to the cell, the energetic cost of building a 100 protein, 1 μ m long pilus, and then subsequently cleaving its C-terminus and releasing it into solvent could be crippling especially if this was allowed to happen at regular intervals. It's also possible that preventing the functionality of the enzyme is simply a byproduct of protecting the active site from inactivation by oxidation of its cysteine. It has been shown previously that the cysteines of several other sortase enzymes are susceptible to oxidation as evidenced by both an increase in catalytic activity reported in the presence of reducing agents, and structures of Streptococcus progenes SrtA solved with its active site cysteine in multiple oxidation states^{47,48}. It has even been proposed that certain host organisms may create an oxidizing environment surrounding a site of infection to deactivate sortase enzymes at the cell surface and render those cells avirulent^{48,49}. The fact that we observed rapid formation of a disulfide linked complex between an IPQT* sorting signal mimic and the W60A lid mutant of SpnSrtC1, but not between the same mimic and the WT protein suggests that the lid indeed functions to prevent oxidation of the active site cysteine (Fig. 6). When tested in vivo, SrtC1 lid mutants showed no discernable effect on transpeptidation^{30,31}. This is in agreement with our proposal that the lid is necessary largely to avoid the energetic penalties associated with cleavage and loss of large numbers of pilus proteins or oxidative inactivation of the sortase itself, as these energetic penalties would likely only provide a fitness advantage in a host setting where slight improvements in efficiency could be pivotal in determining the cell's ability to maintain a successful infection. We thus propose that the lid functions as a regulatory mechanism for the class C sortase that prevents the unwanted cleavage of substrates, and protects the active site from oxidative inactivation.

5.5 **Experimental Procedures**

5.5.1 Cloning, protein expression and purification.

DNA encoding codon optimized soluble SpnSrtC1, residues 17-228 based on uniprot sequence Q97SB9, was generated by recursive PCR from overlapping primers⁵⁰. Codon optimization was performed with OPTIMIZER⁵¹. DNA encoding SpnSrtC1 was then cloned into the pESUMO vector (LifeSensors) and transformed into Escherichia coli BL21 (DE3) cells (Stratagene). Expression was induced at an OD₆₀₀ of 0.5 by addition of 1 mM isopropyl β -D-1thiogalactopyranoside (IPTG) for 5 hrs at 37 °C. Protein was purified by affinity purification using HisPur Co²⁺ resin (Thermo) as per the manufacturer's instructions. The His₆-SUMO tag was cleaved by incubating the protein overnight at 4 °C with recombinant ULP1 protease and subsequently removed by reapplying the protein mixture to HisPur Co^{2+} resin. D58A, W60A, D58A-W60A double mutant, and GSlid mutants of SpnSrtC1 were generated by site directed mutagenesis using standard procedures. ¹⁵N and ¹³C labeled variants were produced by expressing the proteins in M9-minimal media supplemented with ¹⁵N NH₄Cl and/or ¹³C glucose. For these expressions, cells were grown to an OD_{600} of 0.5 at 37 °C, then transferred to an 18 °C incubator and induced with 1 mM IPTG for 16 hrs. Labeled proteins were purified as specified above.

5.5.2 NMR spectroscopy.

All protein samples used for NMR experiments were concentrated and dialyzed into NMR buffer (50 mM sodium phosphate pH 6.5, 50 mM NaCl, 10% D₂O). NMR spectra were recorded on Bruker 500, 600, and 800 MHZ spectrometers equipped with triple resonance cryogenic probes and processed using NMRPipe⁵². Chemical shift assignments were obtained through the analysis of HNCO, HN(CA)CO, HNCA, HNCACB, CBCA(CO)NH, CC(CO)NH, and HBHANH experiments in CARA 1.9.0⁵³.

Dilution experiments to investigate the potential for SpnSrtC1 to dimerize were performed by collecting 15 N-HSQC spectra beginning at 2 mM, and diluting the protein serially with NMR buffer + 10% D₂O to 1.5, 1, 1.5, 0.5, 0.25, 0.2, and 0.1 mM. The number of scans was increased to account for a loss of signal from a reduction of concentration based on the following relationship:

$$NS = NS_0(C_1/C_2)^2$$

Analysis of signal intensities and chemical shift changes was performed in SPARKY⁵⁴, and structure figures were generated in PyMOL⁵⁵.

Dynamics experiments were conducted on 1 mM ¹⁵N labeled samples in NMR buffer described above. Heteronuclear NOE, T_1 and T_2 experiments were conducted at 298 K and analyzed in SPARKY⁵⁴. The residue specific tumbling time, τ_e was calculated using a spherical approximation based on the following equation:

$$\tau_e = \frac{1}{4\pi\nu_N} \sqrt{6\frac{T_1}{T_2} - 7}$$

where v_N is the ¹⁵N resonance frequency in Hz. The molecular tumbling time, τ_c , was calculated as the average of all τ_e values for the protein. CPMG experiments were conducted using a modified approach based on the original method developed by Palmer et. al.⁵⁶. Here, full T_2 datasets were collected for 3 different CPMG pulse delays, 0.0023 s, 0.0048 s, and 0.000514 s. R_2 values (1/ T_2) from these measurements were used to calculate R_{ex} by subtracting the maximum calculated R_2 (either the 0.0023 s or 0.0048 s delay, as R_2 values calculated at these longer delays are known to oscillate slightly), from the minimum R_2 determined at the shortest delay (0.00514 s).

5.5.3 *In vitro* cleavage assay.

SpnSrtC1 and mutants were dialyzed into assay buffer (25 mM TRIS-HCl pH 7.5, 100 mM NaCl). Sorting signal containing peptide: NH₂-VNKKIT**IPQTG**GIGT-CO₂ was synthesized by LifeTein and used without further purification. Lyophilized peptide was dissolved in assay buffer to 2 mM and final concentration was verified using a BCA assay kit (Thermo) per manufacturer's instructions. Reactions were conducted with 100 μ M enzyme and 200 μ M peptide in a total volume of 100 μ L at 37 °C. After 24 hrs, reactions were quenched by addition of 50 μ L 1 M HCl and 100 μ L of the reaction mixture was injected onto a Waters XSelect HSS C₁₈ reversed phase HPLC column. Peptides were eluted with a gradient from 5-30% acetonitrile, 0.1% trifluoroacetic acid over 20 mins. Elution and quantification of peptides was conducted by monitoring absorbance of the peptide bond at 215 nm. Fractions from peaks corresponding to substrate and product were collected and their identities confirmed by matrix assisted laser desorption ionization-mass spectrometry (MALDI-MS).

5.5.4 *In vitro* oxidation assay.

The IPQT* (where T* is (2*R*,3*S*)-3-amino-4-mercapto-2-butanol) substrate mimic was synthesized as in ⁵⁷. The compound was mixed with SpnSrtC1 WT or SpnSrtC1-W60A mutant at a 10:1 molar ratio of mimic to protein and allowed to mix on a rotisserie at room temperature for 24 hrs. Reactions were then applied to zip-tip C₁₈ pipette tips (Millipore) and eluted as per manufacturer's instructions. 1 μ L eluent was then spotted onto a MALDI plate in triplicate and crystallized by the addition of 0.5 μ L 2,5-dihydroxybenzoic acid (DHB) matrix dissolved in 50% ethanol, 0.1% TFA. Formation of a disulfide bond with the mimic was detected via MALDI-MS by a mass increase of 592 Da.

5.6 Figures



Figure 5.1 S. pneumoniae SrtC1 structure and lid motif.

(A) Cartoon representation of SpnSrtC1 from PDB 2W1J²⁰. Active site His, Cys, Arg, and lid Trp residue shown as sticks. N-terminal extension preceding the sortase β -barrel core is colored *red*. (B) Transparent surface representation of SpnSrtC1, showing the lid occluding the active site. (C) Zoom in of SpnSrtC1 active site, showing conserved DP(F/W/Y) lid motif in active site as sticks.



Figure 5.2 NMR analysis shows SpnSrtC1 is dimeric *in vitro*.

(A) HSQC of 300 μ M SpnSrtC1 (*green*) overlayed over an HSQC of 100 μ M SpnSrtC1 (*red*), showing the appearance of several peaks along with the movement or increase in intensity of several others. (B) The two monomers in the asymmetric unit of the crystal structure of SpnSrtC1, PDB ID 2W1J²⁰, are shown as a cartoon representation. Residues with increases in peak height or changes in chemical shift of greater than 1 stdev from the mean are displayed on the structure in *red*, unassigned residues are shown in *beige*, and both are displayed as sticks.



Figure 5.3 NMR dynamics of SpnSrtC1.

A diagram of secondary structure elements is shown at the top of the figure where helices are represented as cylinders and β -strands as arrows. (A) Residue specific tumbling time (τ_e) for each assigned residue in SpnSrtC1. Average for the structured region of the protein plotted as a blue line. (B) Heteronuclear NOE for each residue in SpnSrtC1. A *red* line is shown at a HetNOE value of 0.6. A HetNOE value below this line indicates that the residue can be considered

dynamic. (C) R_{ex} component of relaxation as determined by direct CPMG detection, plotted for each residue.



Figure 5.4 NMR shows SpnSrtC1 mutants have more dynamic lids.

A diagram of secondary structure elements is shown at the top where helices are represented as cylinders and β -strands as arrows. (A) Residue specific tumbling time (τ_e) for each assigned

residue in SrtC1-W60A. (**B**) Heteronuclear NOE for each residue in SpnSrtC1-W60A. *Red* line at 0.6 indicates the cutoff below which residues are considered disordered (**C**) Residue specific tumbling time (τ_e) for each assigned residue in SrtC1-D58A. (**D**) Heteronuclear NOE for each residue in SpnSrtC1-D58A. *Red* line at 0.6 indicates the cutoff below which residues are considered disordered. (**E**) Comparison of NOESY spectra for the W60 indole proton for the WT SpnSrtC1 protein, *left*, or three possible indole configurations seen in the D58A mutant, represented by the *right* 3 strips.



Figure 5.5 SpnSrtC1 lid mutants are more active *in vitro*.

(A) Product formed after 24 hr reactions with IPQTG containing peptide is shown for WT, and various SpnSrtC1 mutant constructs. (B) The amount of substrate remaining after 24 hr incubation for the same reaction detailed above.



Figure 5.6 The lid of SpnSrtC1 protects the active site cysteine from modification by a sulfhydryl modifier.

(A) SrtC1 before (*top*) and after (*bottom*) 24 hr incubation with IPQT* disulfide bond forming substrate mimic showing no significant formation of the disulfide complex. Theoretical mass of SrtC1-WT = 23,813 Da, SrtC1-IPQT* complex = 24,405 Da (**B**) SrtC1-W60A mutant before (*top*) and after (*bottom*) 24 hr incubation with IPQT* disulfide forming peptide showing conversion to the disulfide bonded complex. Theoretical mass of SrtC1-W60A = 23,698 Da, SrtC1-W60A-IPQT* complex = 24,290 Da.

5.7 References

- Ton-That H, Schneewind O. Assembly of pili in Gram-positive bacteria. *Trends Microbiol.* 2004;12(5):228-234. doi:10.1016/j.tim.2004.03.004.
- Telford JL, Barocchi MA, Margarit I, Rappuoli R, Grandi G. Pili in Gram-positive pathogens. *Nat Rev Microbiol*. 2006;4(7):509-519. doi:10.1038/nrmicro1443.
- Mandlik A, Swierczynski A, Das A, Ton-That H. Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. *Trends Microbiol*. 2008;16(1):33-40. doi:10.1016/j.tim.2007.10.010.
- Barocchi MA. A pneumococcal pilus influences virulence and host inflammatory responses. *Proc Natl Acad Sci.* 2006;103(8):2857-2862. doi:10.1073/pnas.0511017103.
- Mandlik A, Swierczynski A, Das A, Ton-That H. Corynebacterium diphtheriae employs specific minor pilins to target human pharyngeal epithelial cells. *Mol Microbiol*. 2007;64(1):111-124. doi:10.1111/j.1365-2958.2007.05630.x.
- Rintahaka J, Yu X, Kant R, Palva A, von Ossowski I. Phenotypical analysis of the Lactobacillus rhamnosus GG fimbrial spaFED operon: surface expression and functional characterization of recombinant SpaFED pili in Lactococcus lactis. *PLoS One*. 2014;9(11):e113922. doi:10.1371/journal.pone.0113922.
- Mishra A, Wu C, Yang J, Cisar JO, Das A, Ton-That H. The Actinomyces oris type 2 fimbrial shaft FimA mediates co-aggregation with oral streptococci, adherence to red

blood cells and biofilm development. *Mol Microbiol*. June 2010. doi:10.1111/j.1365-2958.2010.07252.x.

- Mishra A, Devarajan B, Reardon ME, et al. Two autonomous structural modules in the fimbrial shaft adhesin FimA mediate Actinomyces interactions with streptococci and host cells during oral biofilm development. *Mol Microbiol*. 2011;81(5):1205-1220. doi:10.1111/j.1365-2958.2011.07745.x.
- Margarit I, Rinaudo CD, Galeotti CL, et al. Preventing bacterial infections with pilusbased vaccines: the group B streptococcus paradigm. *J Infect Dis*. 2009;199(1):108-115. doi:10.1086/595564.
- Hendrickx APA, Budzik JM, Oh S-Y, Schneewind O. Architects at the bacterial surface sortases and the assembly of pili with isopeptide bonds. *Nat Rev Microbiol*. 2011;9(3):166-176. doi:10.1038/nrmicro2520.
- Young PG, Moreland NJ, Loh JM, et al. Structural conservation, variability, and immunogenicity of the T6 backbone pilin of serotype M6 Streptococcus pyogenes. *Infect Immun.* 2014;82(7):2949-2957. doi:10.1128/IAI.01706-14.
- Comfort D, Clubb RT. A Comparative Genome Analysis Identifies Distinct Sorting Pathways in Gram-Positive Bacteria. *Infect Immun.* 2004;72(5):2710-2722. doi:10.1128/IAI.72.5.2710-2722.2004.
- Spirig T, Weiner EM, Clubb RT. Sortase enzymes in Gram-positive bacteria. *Mol Microbiol.* 2011;82(5):1044-1059. doi:10.1111/j.1365-2958.2011.07887.x.
- Kang HJ, Coulibaly F, Proft T, Baker EN. Crystal Structure of Spy0129, a Streptococcus pyogenes Class B Sortase Involved in Pilus Assembly. *PLoS One*. 2011;6(1):e15969. doi:10.1371/journal.pone.0015969.
- Shaik MM, Lombardi C, Maragno Trindade D, et al. A structural snapshot of type II pilus formation in Streptococcus pneumoniae. *J Biol Chem.* July 2015:M115.647834 - . doi:10.1074/jbc.M115.647834.
- Clancy KW, Melvin JA, McCafferty DG. Sortase transpeptidases: insights into mechanism, substrate specificity, and inhibition. *Biopolymers*. 2010;94(4):385-396. doi:10.1002/bip.21472.
- Jacobitz AW, Wereszczynski J, Yi SW, et al. Structural and computational studies of the Staphylococcus aureus Sortase B-substrate complex reveal a substrate-stabilized oxyanion hole. *J Biol Chem.* February 2014:jbc.M113.509273. doi:10.1074/jbc.M113.509273.
- Frankel BA, Kruger RG, Robinson DE, Kelleher NL, McCafferty DG. Staphylococcus aureus sortase transpeptidase SrtA: insight into the kinetic mechanism and evidence for a reverse protonation catalytic mechanism. *Biochemistry*. 2005;44(33):11188-11200. doi:10.1021/bi050141j.
- Frankel BA, Tong Y, Bentley ML, Fitzgerald MC, McCafferty DG. Mutational analysis of active site residues in the Staphylococcus aureus transpeptidase SrtA. *Biochemistry*. 2007;46(24):7269-7278. doi:10.1021/bi700448e.

- Manzano C, Contreras-Martel C, El Mortaji L, et al. Sortase-mediated pilus fiber biogenesis in Streptococcus pneumoniae. *Struct (London, Engl 1993)*. 2008;16(12):1838-1848. doi:10.1016/j.str.2008.10.007.
- Manzano C, Izoré T, Job V, Di Guilmi AM, Dessen A. Sortase activity is controlled by a flexible lid in the pilus biogenesis mechanism of gram-positive pathogens. *Biochemistry*. 2009;48(44):10549-10557. doi:10.1021/bi901261y.
- Persson K. Structure of the sortase AcSrtC-1 from Actinomyces oris. *Acta Crystallogr D Biol Crystallogr*. 2011;67(Pt 3):212-217. doi:10.1107/S0907444911004215.
- Suree N, Liew CK, Villareal VA, et al. The Structure of the Staphylococcus aureus Sortase-Substrate Complex Reveals How the Universally Conserved LPXTG Sorting Signal Is Recognized. *J Biol Chem*. 2009;284(36):24465-24477. doi:10.1074/jbc.M109.022624.
- 24. Chan AH, Yi SW, Terwilliger AL, Maresso AW, Jung ME, Clubb RT. Structure of the Bacillus anthracis Sortase A Enzyme Bound to Its Sorting Signal: a Flexible Aminoterminal Appendage Modulates Substrate Access. *unpublished*. 2015.
- Neiers F, Madhurantakam C, Fälker S, et al. Two crystal structures of pneumococcal pilus sortase C provide novel insights into catalysis and substrate specificity. *J Mol Biol.* 2009;393(3):704-716. doi:10.1016/j.jmb.2009.08.058.

- Khare B, Krishnan V, Rajashankar KR, et al. Structural differences between the Streptococcus agalactiae housekeeping and pilus-specific sortases: SrtA and SrtC1. *PLoS One*. 2011;6(8):e22995. doi:10.1371/journal.pone.0022995.
- 27. Khare B, Fu Z-Q, Huang I-H, Ton-That H, Narayana SVL. The crystal structure analysis of group B Streptococcus sortase C1: a model for the "lid" movement upon substrate binding. *J Mol Biol.* 2011;414(4):563-577. doi:10.1016/j.jmb.2011.10.017.
- Cozzi R, Prigozhin D, Rosini R, et al. Structural Basis for Group B Streptococcus Pilus 1 Sortases C Regulation and Specificity. *PLoS One*. 2012;7(11):e49048. doi:10.1371/journal.pone.0049048.
- Danne C, Dramsi S. Pili of gram-positive bacteria: roles in host colonization. *Res Microbiol.* 163(9-10):645-658. doi:10.1016/j.resmic.2012.10.012.
- Cozzi R, Malito E, Nuccitelli A, et al. Structure analysis and site-directed mutagenesis of defined key residues and motives for pilus-related sortase C1 in group B Streptococcus. *FASEB J.* 2011;25(6):1874-1886. doi:10.1096/fj.10-174797.
- Wu C, Mishra A, Reardon ME, et al. Structural determinants of Actinomyces sortase SrtC2 required for membrane localization and assembly of type 2 fimbriae for interbacterial coaggregation and oral biofilm formation. *J Bacteriol.* 2012;194(10):2531-2539. doi:10.1128/JB.00093-12.
- Mittermaier A, Kay LE. New Tools Provide New Insights in NMR Studies of Protein Dynamics. *Science (80-)*. 2006;312(5771):224-228. doi:10.1126/science.1124964.

- Cozzi R, Zerbini F, Assfalg M, et al. Group B Streptococcus pilus sortase regulation: a single mutation in the lid region induces pilin protein polymerization in vitro. *FASEB J*. 2013;27(8):3144-3154. doi:10.1096/fj.13-227793.
- Lu C, Zhu J, Wang Y, et al. Staphylococcus aureus sortase A exists as a dimeric protein in vitro. *Biochemistry*. 2007;46(32):9346-9354. doi:10.1021/bi700519w.
- Robson SA, Jacobitz AW, Phillips ML, Clubb RT. Solution Structure of the Sortase Required for Efficient Production of Infectious Bacillus anthracis Spores. *Biochemistry*. 2012;51(40):7953-7963. doi:10.1021/bi300867t.
- Zhu J, Xiang L, Jiang F, Zhang ZJ. Equilibrium of sortase A dimerization on Staphylococcus aureus cell surface mediates its cell wall sorting activity. *Exp Biol Med* (Maywood). June 2015. doi:10.1177/1535370215592122.
- Sonnhammer EL, von Heijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol.* 1998;6:175-182. http://www.ncbi.nlm.nih.gov/pubmed/9783223. Accessed June 29, 2015.
- Fälker S, Nelson AL, Morfeldt E, et al. Sortase-mediated assembly and surface topology of adhesive pneumococcal pili. *Mol Microbiol*. 2008;70(3):595-607. doi:10.1111/j.1365-2958.2008.06396.x.
- Shaik MM, Maccagni A, Tourcier G, Di Guilmi AM, Dessen A. Structural basis of pilus anchoring by the ancillary pilin RrgC of Streptococcus pneumoniae. *J Biol Chem*. 2014;289(24):16988-16997. doi:10.1074/jbc.M114.555854.

- Weiner EM, Robson S, Marohn M, Clubb RT. The Sortase A enzyme that attaches proteins to the cell wall of Bacillus anthracis contains an unusual active site architecture. J Biol Chem. 2010;285(30):23433-23443. doi:10.1074/jbc.M110.135434.
- Wallock-Richards DJ, Marles-Wright J, Clarke D, et al. Molecular Basis of Streptococcus mutans Sortase A Inhibition by the Flavonoid Natural Product trans-Chalcone. *Chem Commun.* June 2015. doi:10.1039/C5CC01816A.
- Maresso AW, Chapa TJ, Schneewind O. Surface protein IsdC and sortase B are required for heme-iron scavenging of Bacillus anthracis. *J Bacteriol*. 2006;188(23):8145-8152. doi:10.1128/JB.01011-06.
- 43. Ton-That, H., Mazmanian, S. K., Alksne, L., and Schneewind O. Anchoring of Surface Proteins to the Cell Wall of Staphylococcus aureus. SORTASE CATALYZED IN VITRO TRANSPEPTIDATION REACTION USING LPXTG PEPTIDE AND NH2-GLY3 SUBSTRATES. *J Biol Chem.* 2000;275(13):9876-9881. doi:10.1074/jbc.275.13.9876.
- Hu P, Bian Z, Fan M, Huang M, Zhang P. Sec translocase and sortase A are colocalised in a locus in the cytoplasmic membrane of Streptococcus mutans. *Arch Oral Biol.* 2008;53(2):150-154. doi:10.1016/j.archoralbio.2007.08.008.
- 45. Kline KA, Kau AL, Chen SL, et al. Mechanism for sortase localization and the role of sortase localization in efficient pilus assembly in Enterococcus faecalis. *J Bacteriol*. 2009;191(10):3237-3247. doi:10.1128/JB.01837-08.

- 46. Raz A, Fischetti VA. Sortase A localizes to distinct foci on the Streptococcus pyogenes membrane. *Proc Natl Acad Sci U S A*. 2008;105(47):18549-18554. doi:10.1073/pnas.0808301105.
- Mazmanian SK, Ton-That H, Su K, Schneewind O. An iron-regulated sortase anchors a class of surface protein during Staphylococcus aureus pathogenesis. *Proc Natl Acad Sci U* SA. 2002;99(4):2293-2298. doi:10.2307/3057958.
- Race PR, Bentley ML, Melvin JA, et al. Crystal Structure of Streptococcus pyogenes Sortase A. *J Biol Chem.* 2009;284(11):6924-6933. doi:10.1074/jbc.M805406200.
- 49. Melvin JA, Murphy CF, Dubois LG, Thompson JW, Moseley MA, McCafferty DG. Staphylococcus aureus sortase A contributes to the Trojan horse mechanism of immune defense evasion with its intrinsic resistance to Cys184 oxidation. *Biochemistry*. 2011;50(35):7591-7599. doi:10.1021/bi200844h.
- Prodromou C, Pearl LH. Recursive PCR: a novel technique for total gene synthesis. *Protein Eng.* 1992;5(8):827-829. http://www.ncbi.nlm.nih.gov/pubmed/1287665. Accessed August 13, 2015.
- Puigbò P, Guzmán E, Romeu A, Garcia-Vallvé S. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Res*. 2007;35(Web Server issue):W126-W131. doi:10.1093/nar/gkm219.
- 52. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*.

1995;6(3):277-293. http://www.ncbi.nlm.nih.gov/pubmed/8520220. Accessed February 13, 2015.

- 53. Keller R. {CARA: computer aided resonance assignment. http://cara.nmr.ch/}. 2004.
- 54. Goddard TD, Kneller DG. Sparky 3. Univ California, San Fr. 2006.
- 55. Schrödinger L. The PyMOL Molecular Graphics System, Version 1.3r1. August 2010.
- Palmer AG, Kroenke CD, Loria JP. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol*. 2001;339:204-238. http://www.ncbi.nlm.nih.gov/pubmed/11462813. Accessed August 13, 2015.
- 57. Jung ME, Clemens JJ, Suree N, et al. Synthesis of (2R,3S) 3-amino-4-mercapto-2-butanol, a threonine analogue for covalent inhibition of sortases. *Bioorg Med Chem Lett*. 2005;15(22):5076-5079. doi:10.1016/j.bmcl.2005.07.073.