Class A Sortases: Structures and Alternative Substrate Binding and Cleavage

Brandon Vogel
Western Washington University, brvogel93@gmail.com

Follow this and additional works at: https://cedar.wwu.edu/wwuet

Part of the Chemistry Commons

Recommended Citation
https://cedar.wwu.edu/wwuet/1227

This Masters Thesis is brought to you for free and open access by the WWU Graduate and Undergraduate Scholarship at Western CEDAR. It has been accepted for inclusion in WWU Graduate School Collection by an authorized administrator of Western CEDAR. For more information, please contact westerncedar@wwu.edu.
Class A Sortases: Structures and Alternative Substrate Binding and Cleavage

By

Brandon Vogel

Accepted in Partial Completion
of the Requirements for the Degree
Master of Science

ADVISORY COMMITTEE

Dr. Jeanine Amacher, Chair

Dr. Jay McCarty

Dr. John Antos

GRADUATE SCHOOL

David L. Patrick, Dean
In presenting this thesis in partial fulfillment of the requirements for a master’s degree at Western Washington University, I grant to Western Washington University the non-exclusive royalty-free right to archive, reproduce, distribute, and display the thesis in any and all forms, including electronic format, via any digital library mechanisms maintained by WWU.

I represent and warrant this is my original work, and does not infringe or violate any rights of others. I warrant that I have obtained written permissions from the owner of any third party copyrighted material included in these files.

I acknowledge that I retain ownership rights to the copyright of this work, including but not limited to the right to use all or part of this work in future works, such as articles or books.

Library users are granted permission for individual, research and non-commercial reproduction of this work for educational purposes only. Any further digital posting of this document requires specific permission from the author.

Any copying or publication of this thesis for commercial purposes, or for financial gain, is not allowed without my written permission.

Brandon Vogel

08/03/2023
Class A Sortases: Structures and Alternative Substrate Binding and Cleavage

A Thesis
Presented to
The Faculty of
Western Washington University

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science

by
Brandon Vogel
August 2023
Abstract

Sortases, consisting of classes A-F, are cysteine transpeptidases found in the cell wall of Gram-positive bacteria. They play a crucial role in ligating proteins to the cell wall that are responsible for cell adhesion, immune evasion, host cell invasion, and nutrient acquisition through a transpeptidation reaction. Consequently, they are an attractive therapeutic target. Class A sortases are also utilized in protein engineering applications such as sortase-mediated ligations and sortagging. Despite extensive research in the past two decades, gaps persist in understanding how class A sortases recognize their substrates, primarily due to a lack of structural information on sortases non-covalently bound to substrates. Our work in Chapter 1 presents the first crystal structures of a catalytically inactive Streptococcus pyogenes sortase A (spySrtA) bound to peptide substrates LPATA, LPATS, and a mimic of the ligation product LPAT-Lipid II. Then, 900 ns molecular dynamic simulations were performed to explore ligand binding dynamics. Additionally, our peptide-bound structures were used to model the acyl-enzyme intermediate of spySrtA-LPAT. Chapter 2 builds on our collaborators' discovery in the Antos lab that certain class A sortases participate in an alternative cleavage motif. Our work involved 18-hour endpoint mass spectrometry assays with various point mutations and chimeric enzymes, revealing that an H143A mutation reduced this alternative cleavage while maintaining near wild type activity. These findings provide insights into target recognition and binding, potentially benefiting therapeutic design and enzyme tuning for protein engineering applications.
Acknowledgements

I would like to convey my profound thanks to Dr. Jeanine Amacher who brought me into their lab during the pandemic and offered me the opportunity to learn how to conduct scientific research. She supported me throughout this entire process, and I will be forever grateful for her tutelage, patience, and mentorship. Without her guidance through my undergraduate degree and in my graduate studies, I would not be where I am now or where I may possibly go in the future. Many doors are now open thanks to her efforts, care, and instruction.

I’d also like to acknowledge and thank the members of the Amacher lab who have had my back during each and every quarter, from ending FPLC runs to helping with peptide synthesis and protein purifications. This work would not have been possible without the assistance and comradery shown by them.

My deepest appreciation to my other thesis committee members, Drs. John Antos and Jay McCarty, who made this work possible at all. Prof. John Antos was instrumental in the development of the assays that led to the completion of this work. His patience was greatly appreciated while instructing me on the intricacies of HPLC and mass spectroscopy data. Prof. Jay McCarty was an amazing bioinformatics and molecular dynamics tutor. I would know next to nothing about molecular dynamics were it not for his incredible instruction and depth of patience.

Another thanks to the professors at Western that offered excellent instruction, teaching me everything I know and some of what I have yet to understand. Special thanks to Spencer Anthony-Cahill, James Vyvyan, Serge Smirnov, and Clint Speigel for adding guidance during my Undergraduate studies and lending support during my Graduate studies.

I must also extend the warmest of thanks to the people in the Scientific and Technical Services department at Western Washington University. Sarina Kiesser, Cassi King, and Alyssa Tsukada were absolutely instrumental in the completion of the research conducted here.

Thanks to all the past and present members of the Spiegel, Antos, Anthony-Cahill, and Smirnov labs for their advice, support, and previous work that led me here though my studies.

I would like to gratefully acknowledge the funding I received from the NSF and Western Washington University for my research and teaching assistantships.

And finally, thank you to the family and friends that always supported my endeavors throughout my studies both as an undergraduate and graduate student.

None of this would have been remotely possible without any of you. My life, and that of everyone around me, will be forever altered for the better. Thank you all so very much.
Table of Contents

Abstract .............................................................................................................................. iv

Acknowledgements ......................................................................................................... v

List of Tables and Figures .............................................................................................. vii

Introduction ......................................................................................................................1

Chapter 1 .......................................................................................................................... 8

Chapter 2 ..........................................................................................................................40

Chapter 3 ..........................................................................................................................56

Works Cited ......................................................................................................................59

Appendix A ......................................................................................................................64

Appendix B ......................................................................................................................77

Sortase Sequences ..........................................................................................................85
List of Tables and Figures

Introduction:

Figure 1. Sortase mediated ligation scheme ......................................................... 1
Figure 2. Structure of Streptococcus Pyogenes Sortase A ....................................... 2
Figure 3. Sortase Binding regions .......................................................................... 3
Figure 4. Sortase catalytic mechanism ................................................................. 5

Chapter 1:

Figure 5. Structural model of the spySrtA catalytic mechanism............................... 12
Figure 6. The spySrtA complex structures with LPATS and LPATA peptides .......... 14
Table 1. Data Collection and Refinement Statistics .................................................. 16
Figure 7. Stereochemistry of the spySrtA-LPATS interaction .................................. 20
Figure 8. A model of the acyl-enzyme intermediate of spySrtA-LPAT ..................... 23
Figure 9. The structure of spySrtA bound to a peptide model of the LPAT-lipid II ligation product .................................................................................................................. 28
Figure 10. Molecular dynamics simulations of spySrtA structures ............................ 31

Chapter 2:

Figure 11. Abz-LPATGGK-Dnp peptide .................................................................. 42
Figure 12. Synthesized peptides. ............................................................................ 43
Figure 13. MSA of class sortases ........................................................................... 44
Table 2. Percent conversion table .......................................................................... 45
Figure 14. Measurements from Molecular dynamics simulations ............................ 49

Appendix A:

Figure S1. Sequences and LC-ESI-MS characterization of wild-type and C208A spySrtA .. 68
Figure S2. Model transacylation reactions of spySrtA with LPATA and LPATS substrates…. 69
Figure S3. Structural alignments of spySrtA complex structures ...............................70
Figure S4. The spySrtA LPATA- and LPATS-complex structures ............................71
Figure S5. Comparison of transacylation reactions with LPAT-LII and LPATA ..........72
Figure S6. Structure of spySrtA bound to LPAT-LII peptide ..................................73
Figure S7. Molecular dynamics simulations of spySrtA ...........................................74
Figure S8. Synthesis and characterization of LPAT-LII ...........................................75
Table S1. Details of the molecular dynamics simulation size ....................................76

Appendix B:

Figure S9. UV-vis trace of WTSpySrtA ................................................................77
Figure S10. UV-vis trace of WTlmSrtA ...............................................................78
Figure S11. UV-vis trace of WTSagSrtA .............................................................79
Figure S12. UV-vis trace of H143A SpySrtA .......................................................80
Figure S13. UV-vis trace of I211P SpySrtA .........................................................81
Figure S14. UV-vis trace of mono(spy)7/8. ..........................................................82
Figure S15. UV-vis trace of spy(mono)7/8.............................................................83
Table S2. Table of Peptide masses .................................................................84
Table S3. Details of the molecular dynamics simulation size ...............................84
Introduction

Sortases: An Overview

Cell walls protect both gram-negative and gram-positive bacteria from the extra cellular environments they persist in as well as give shape to the cell and play critical roles in overall cell function.\(^1\) Gram positive bacteria lack the lipopolysaccharide outer membrane that gram-negative bacteria have and instead have a thick wall made up of layers of peptidoglycan peppered with long anionic polymers, called teichoic acids.\(^1\) In addition to teichoic acids, there is the presence of various proteins that are displayed on the cell surface.\(^1\) Because there is no outer membrane, all of these proteins contain some feature that anchors them in or near the membrane in the form of membrane spanning helices or are attached to lipid anchors embedded in the membrane.\(^1\)–\(^3\) Still other proteins are covalently attached to or associated with peptidoglycan or bind to teichoic acids. Of the proteins that are displayed on the cell surface, those that are covalently attached are done so through the use of an amino-terminal signal sequence that allows the protein to be secreted through the cell membrane via the sec translocon and a carboxy-terminal pentapeptide cell wall sorting signal (CWSS).\(^1\)–\(^4\)

The enzymes that are responsible for recognizing the CWSS and attaching the secreted protein to the cell wall are called sortases. Sortase enzymes are a family of cysteine transpeptidases found in most known gram-positive bacteria and are organized into six classes, A-F, based on sequence homology and substrate preference.\(^1\)–\(^4\) Of these the most well researched sortases are from that of class A which are found in all but two gram-positive bacteria: *Mycobacterium* and *Microplasma*.\(^5\) The archetypal sortase A enzyme was first identified from *Staphylococcus aureus* which selects on a CWSS consisting of an LPXTG motif,
where X is any amino acid. This motif is followed by a hydrophobic segment of amino acids and a tail composed of mostly positively charged residues. Sortases are found on the extracellular side of the membrane, and act on their protein substrate after it is partially secreted, cleaving the LPXTG between the T and G residues then ligating the cleaved LPXT- to a lipid II molecule embedded in the cell wall. The mechanism with which this takes place is discussed further down.

**Figure 1.** Sortase mediated ligation scheme. Acyl donors and acyl acceptors can take the form of solid supports, particles, cells, peptides, and proteins.

There are two main functions that sortases execute to serve the cell: (1) attaching proteins directly to the cell wall as stated above and (2) the assembly of pili. Many of the proteins attached by sortases and displayed on the cell wall are involved in functions that are critical to a bacterial cell’s survival and in pathogenic bacteria, their virulence. These functions include nutrient acquisition, immune system evasion, host cell invasion, and cell adhesion to name a few. For these reasons sortases have become a highly researched enzyme as an attractive therapeutic target. Class A sortases have also attracted great interest in the protein engineering field. This is due to their ability to be solubilized and their highly selective nature allowing them to be used to ligate distinct biomolecules together in vitro through a covalent peptide bond. Some of these applications include ligating fluorophores or drugs to antibodies and proteins, attaching
proteins to cells, the preparation of protein nanoparticles, and the cyclization of proteins and peptides.\textsuperscript{2,12,14} These applications are referred to as sortase mediated ligations (SML), sortagging, or sortylation a scheme of which is shown in Figure 1.

\textbf{Sortase Structure}

Sortases have a highly conserved structure consisting of 3 regions: (i) an amino terminal sequence that allows for secretion across the cell membrane, (ii) a nonpolar section of amino acids that remains embedded in the phospholipid bilayer, (iii) and the conserved, water soluble carboxy terminal catalytic domain depicted here in Figure 2 from \textit{Streptococcus pyogenes}.\textsuperscript{2} Class A sortases from \textit{Staphylococcus aureus} were the first to be characterized by Ilangovan et al using NMR spectroscopy and revealed the conserved antiparallel eight-stranded beta-barrel canonical sortase fold.\textsuperscript{4} Within this fold is the highly conserved catalytic triad consisting of His-Cys-Arg.\textsuperscript{2–4} An area that has significant variation among class A sortases is that of the loops surrounding the binding cleft.\textsuperscript{11}

Depicted in Figure 3 is the binding pocket from \textit{S. aureus} which shows the labeled areas S4 to S1 which correspond to the labeling on the CWSS; P4-P3-P2-P1-P1'.\textsuperscript{2} The base of the binding pocket is formed by residues in strands β4 and β7 and the walls are formed by surface loops between β strands 6 and 7, 3 and 4 and the β2 strand and the H2 helix.\textsuperscript{2,4} The S4 region is a highly hydrophobic binding pocket found across
most class A sortases and as such many of the catalyzed substrates accept a leucine here.\textsuperscript{2,4,6,11}

Hydrophobic contacts with proline are found in the S3 region and proline is a highly selected on residue in the P3 position.\textsuperscript{2,15} Less is known or discussed on the S2 and S1 binding regions.

While this description paints a picture for understanding the binding pocket architecture, there is structural variation in class A homologs from reported structures of \textit{S. pyogenes} (SpySrtA), \textit{S. agalactiae} (SagSrtA), \textit{B. anthracis} (BaSrtA), and \textit{S. mutans} (SmSrtA).

![Figure 3. Staphylococcus aureus and Streptococcus pyogenes binding region. PDB ID: 2KID and 3FN5](image)

Structural variation and loop dynamics can vary across the class A sortase homologs in such a way to affect binding and catalysis of substrate. Thus researchers have been extensively investigating β4/5, β6/7, and β7/8 loop interactions and dynamics.\textsuperscript{11,16,17} One feature found in many of these enzymes is the presence of a short helix in the β6/7 loop which may contact the CWSS upon binding.\textsuperscript{6,18} There is variation in this loop with respect to dynamics as it has been shown that in SaSrtA the loop is disordered and mobile until binding which triggers a disordered to ordered transition state, whereas in all other class A sortases this loop remains in an ordered state containing a short helix.\textsuperscript{2,6,19} Another notable structural variation is the presence of a second
groove leading into the active site.\textsuperscript{2} This was first discovered in the SpySrtA structure where the H1 helix and β7/8 loop from the walls of this groove while the β4/5 loop forms the base.\textsuperscript{20} Furthermore there are some class A enzymes that possess a flexible N-terminal appendage that may influence substrate binding.\textsuperscript{19,21} This was first noted in the NMR structure of \textit{B. anthracis} SrtA (BaSrtA) where the N-terminus residues 57 to 79 wrap around the protein then come into contact with the active site.\textsuperscript{21} It is thought that this appendage modulates substrate access to the binding site increasing efficiency by limiting the possibility of hydrolytic cleavage to occur.\textsuperscript{21}

**Catalytic mechanism**

The catalytic mechanism, presented in Figure 4, by which sortases ligate proteins destined for the cell wall was first elucidated by characterizing the archetypal \textit{Staphylococcus aureus} sortase A.\textsuperscript{4,6} The currently accepted mechanism for most class A sortases is evidenced through kinetic studies to occur in a ping-pong fashion.\textsuperscript{2,22,23} From here on the LPXTG will be interchangeably referred to as P4 to P1’ where P4 corresponds to the L residue, P3 to P residue, X is to P2, T is to P1, and the G is the P1’ residue. First, class A sortase will recognize the LPXTG motif within the CWSS of the protein substrate that was secreted.\textsuperscript{2,3,6} The binding grooves made up by the β4/5, β6/7, and β7/8 loops and strands of the sortase enzyme are responsible for the recognition of the CWSS and subsequent binding.\textsuperscript{11,16,18} As binding occurs the highly conserved proline at the P3 position allows for the P4 residue to be oriented in to the hydrophobic S4 pocket while the C-terminal end of the CWSS is oriented towards the active site.\textsuperscript{6} The next step is the cysteine residue performs a nucleophilic attack on the carbonyl carbon in the peptide bond between the P1 and P1’ residues.\textsuperscript{2,3,7,23} This reaction forms a tetrahedral intermediate that collapses to form a semi-stable thioacyl intermediate where the sortase is covalently attached to the substrate at its catalytic cysteine residue.\textsuperscript{13,20,22} Sortase then recognizes
a lipid II molecule embedded in the cell wall and catalyzes a reaction of the glycine nucleophile where the N-terminal primary amine group performs a nucleophilic attack on the carbonyl carbon atom sharing the thioacyl bond.\textsuperscript{2,3,20} A second tetrahedral intermediate is formed but quickly collapses to the protein-lipid II product via a newly formed peptide bond.\textsuperscript{2,3} This lipid II linked protein is attached to the cell wall through the transpeptidation and glycosylation reactions that help form the cell wall itself.\textsuperscript{2}

**Figure 4.** Catalytic Mechanism of Sortase A. Jacobitz et al. 2017.

**Previous work**

In recent years there has been a surge in research to increase the utility of sortases in these ligation strategies. Strategies such as directed evolution screening and phage, or yeast
display have been employed to create mutants of *Staph aureus* that are more catalytically efficient as well as a mutant that removed the need for a Ca$^{2+}$ cofactor.$^{10,13,24}$ While most of the work with sortases has pertained to that of *S. aureus* since it was the first to be discovered, more work has been done to find other sortase homologs that may provide yet another tool in SML reactions. For example, work has been done by Schmol et al. that discovered the promiscuity of Streptococci sortase A and that it actually has greater selectivity on the LPXLG motif over the canonical LPXTG motif.$^7$ While more and more research provided interesting insights into the possible relaxed selectivity of sortase A homologs there was still a lack of research on characterizing in vitro reactivity of many homologs as well as a limited investigations that probed new substrates. This is where the Antos lab came in.

In a 2018 paper from the Antos lab, their group was able to use model peptides and eight sortase A homologues from different bacterial species to probe what their preferred substrates were and if there were any new unknown substrates.$^{15}$ Their results were fascinating, not only were they able to show that there was promiscuity among some of these homologues, namely *Streptococcus pneumoniae* and *Listeria monocytogenes*, but they also discovered that these sortases were able to catalyze a transacylation reaction at alternative positions within the substrate sequence.$^{15}$

**Our work**

Understanding the structure of a protein is critical to understanding that protein’s function and the mechanism with which an enzyme performs that function. Knowing which regions are playing an active role could lead to an increase in the tunability of sortases, making them an even more powerful tool in protein engineering. And so, being that the Amacher lab is one focused on
structural biochemistry, we set about to answer the questions of how sortases recognize and bind their substrates. Initially this work set about by creating chimeric sortases by swapping out loops thought to be critical to substrate binding such as the β7/8 loops, then checking their activities against the canonical substrate as well as others that are known to be active in both donors to the chimeric protein and subsequently crystallizing those constructs.\textsuperscript{16,18} Then more recent studies were conducted with\textit{S. pyogenes} using a catalytically inactive mutant for X-ray crystallography and molecular dynamic simulations to glean insights into substrate recognition.\textsuperscript{11} Through this work and that of others it has been shown that the structurally conserved β6/7 and β7/8 loops directly affect sortase target recognition and activity.\textsuperscript{6,11,16,18,21}

These studies and that of Nikghalb and colleagues prompted the question of what residues are taking an active role in catalysis and binding when alternative cleavage takes place and is there possibly an alternative catalytic residue that may be responsible? In effort to answer this question we have begun to perform NMR studies, x-ray crystallography, and alternative cleavage assays with loop swapped and point mutants to glean information into which residues are playing active roles in alternative cleavage. Our aims here are to, (1) investigate the possibility of an alternative cleavage mechanism due to the presence of an adjacent histidine residue to the catalytic histidine; (2) probe the biophysical parameters responsible for the binding of class A sortases to their native substrates focusing on the S2-S1’ regions; and (3) investigate the biophysical parameters that determine binding and cleavage of alternative substrates by\textit{Listeria monocytogenes, Streptococcus pyogenes, and Streptococcus agalactiae}. 
Chapter 1

Structures of *Streptococcus pyogenes* Class A sortase in complex with substrate and product mimics provide key details of target recognition

D. Alex Johnson,†,‡ Isabel M. Piper,†,‡ Brandon A. Vogel,† Sophie N. Jackson,† Justin E. Svendsen,†,‡ Hanna M. Kodama,† Darren E. Lee,† Katy M. Lindblom,† James McCarty,† John M. Antos,§ Jeanine F. Amacher†

†Department of Chemistry, Western Washington University, Bellingham, WA, 98225

§These authors contributed equally to this work.


Brandon A. Vogel’s contributions to this work included all aspects of the preparation and execution of the molecular dynamics simulations excluding the modeling of the acyl-enzyme intermediate. He also assisted with the preparation and purification of C208A SpySrtA and the crystallization of SpySrtA-LPATS.
Introduction

Bacterial sortases are cysteine transpeptidase enzymes that play important roles at the cell wall of Gram-positive bacteria. Despite over 20 years since the discovery of the first sortase enzyme in *Staphylococcus aureus*, a complete picture of how these critical enzymes recognize their ligands has remained elusive due to limited structural information involving sortases in complex with their substrates. This type of characterization is essential to understanding how sortases perform their role of attaching protein factors to the bacterial cell wall. A thorough understanding of this process is also relevant to human health and disease in two significant ways; sortases are used in protein engineering, e.g., sortase-mediated ligation (SML), *sortagging*, or *sortylation* applications, and are also therapeutic targets for the development of antibiotics.

Sortases are widespread in Gram-positive bacteria, and are currently grouped into multiple classes (A-F), including several that are considered general housekeeping enzymes (e.g., Classes A and E), and those that assemble pili (Class C). The sortase mechanism involves two catalytic steps: i. Recognition and cleavage of a target sequence, and formation of an acyl-enzyme intermediate, followed by ii. Nucleophilic attack by a second reactant, initiating a ligation reaction that creates a new peptide bond, or isopeptide in the case of the bacterial pilus. For Class A sortases, the general consensus sequence, which is found within the cell wall sorting signal (CWSS), includes a pentapeptide motif, Leu-Pro-X-Thr-Gly (or LPXTG), where X = any amino acid. Positions are defined with respect to the location of the cleavage site between the threonine and glycine residues, with P1’ = Gly, P1 = Thr, P2 = X, P3 = Pro, and P4 = Leu. For protein-anchoring to the bacterial cell surface, the nucleophile in the second step of the reaction mechanism
is the cell-wall precursor lipid II, thus allowing for incorporation of the protein into the growing peptidoglycan layer.\textsuperscript{28}

The majority of knowledge to date on sortase structure and mechanism is focused on Class A sortases, however there are available structures of representative sortases from all six classes (A-F), e.g., Class A (PDB ID 2KID), Class B (1NG5), Class C (3O0P), Class D (2LN7), Class E (5CUW), and Class F (5UUS).\textsuperscript{16} These structures have revealed that sortases share a conserved core antiparallel 8-stranded $\beta$-barrel structure, termed the \textit{sortase fold}.\textsuperscript{2,29} This was first identified in the \textit{Staphylococcus aureus} Class A sortase (saSrtA) structure and is consistently found in wild-type and chimeric sortase enzymes.\textsuperscript{4,16,29} As of early 2022, there were over 65 structures of sortases in the Protein Data Bank, including from all 6 classes and SrtA structures from 10 different organisms. Despite this, there is a notable lack of structural information about ligand recognition in sortases. Of the three SrtA structures that contain ligands, two approximate the acyl-enzyme intermediates of saSrtA and \textit{Bacillus anthracis} SrtA (baSrtA) using cleverly designed peptidomimetic ligands (PDB IDs 2KID, 2RUI). However, because it is not present, these structures do not provide information about recognition of the P1’ residue, a position for which SrtA enzymes have shown variable selectivity \textit{in vitro}.\textsuperscript{6,15,16,19} The third structure contains a complex between saSrtA and a non-covalently bound LPETG peptide that is shifted by several Angstroms in the peptide-binding pocket (PDB ID 1T2W), revealing a geometry that is not consistent with known biochemical data.\textsuperscript{30}

In this work, we have sought to fill remaining gaps in the understanding of SrtA target recognition through the structural characterization of multiple states in the catalytic mechanism of \textit{Streptococcus pyogenes} sortase A (spySrtA) (Fig. 5). The apo structure of spySrtA was solved using X-ray crystallography in 2009, and its catalytic triad consists of His142, Cys208, and Arg216.
Using similar crystallization conditions, we were able to crystallize and solve the structures of a catalytically inactive C208A spySrtA mutant bound to the peptides LPATA and LPATS, which are sequences that are known to serve as spySrtA substrates in vitro. In addition, we synthesized a model peptide (LPAT-LII) of the ligation product between the LPAT fragment and the in vivo nucleophile lipid II, and solved the structures of two complexes between C208A spySrtA and LPAT-LII where the peptide is in the “Thr-in” and “Thr-out” conformations, terminology previously used to describe the side chain of the P1 Thr as protein-interacting (“Thr-in”) or solvent-interacting (“Thr-out”).

Because these are the first solved peptide-bound sortase structures that include the P1’ residue and initial cleavage site, we wanted to investigate the relative dynamics of ligand-binding. We ran 900 nanosecond molecular dynamics simulations using four structures (apo spySrtA (PDB ID 3FN5), spySrtA-LPATA, spySrtA-LPATS, and spySrtA-LPAT-LII) to assess positional flexibility and the overall dynamics of the sortase-peptide complex. Finally, we used our peptide-bound structures to model the acyl-enzyme intermediate of spySrtA-LPAT. Taken together, this
work provides new structural insights for important states in the SrtA catalytic mechanism (Fig. 5), significantly increasing our understanding of target recognition in this important protein family.

![Figure 5. Structural model of the spySrtA catalytic mechanism.](image)

**Figure 5. Structural model of the spySrtA catalytic mechanism.** A summary of the spySrtA catalytic mechanism, as supported by biochemical and structural data in the field, including studies presented here. A portion of the lipid II ligation partner from *S. pyogenes* is shown, and the structure of this component varies between bacterial species. In this model, LPXTA is shown as the target sequence in the first step of the reaction to reflect the ability of spySrtA to recognize a P1’ Ala residue in vitro, and to be consistent with the structural data described in this work. The sortase apoenzyme in state 1 is PDB ID 3FN5. The structures of the enzyme-substrate complex (state 2) and enzyme-ligation product complex (state 4) are the experimental structures presented in this study. As discussed in the main text, the acyl-enzyme (state 3) is a model generated from experimentally determined structures of the enzyme-substrate complex.

**Results**

**Peptide-bound spySrtA crystallization and structure determination**

Like other Class A sortases, the majority of predicted and verified *in vivo* targets of spySrtA possess LPXTG substrate sequences. In addition, prior work from ourselves and others has demonstrated that spySrtA readily accepts LPXTA and LPXTS substrates *in vitro*, despite the fact that these particular sequence variants do not appear to be present in naturally occurring spySrtA
substrates in vivo. The spySrtA enzyme also accepts alanine- or serine-based nucleophiles, which is a characteristic that has been exploited for dual-labeling SML strategies and is consistent with the presence of N-terminal alanines in the interpeptide bridge of lipid II in S. pyogenes. Notably, the ability of spySrtA to recognize non-glycine nucleophiles and to accept substrates that vary at the P1’ position is in stark contrast to saSrtA, which is narrowly selective for glycine at these sites.

In order to gain a stereochemical understanding of target recognition by spySrtA and other Class A sortases, we sought to co-crystallize a catalytically inactive mutant (C208A) of spySrtA with a range of model peptides containing known substrate sequences (LPATG/S/A). Briefly, spySrtA protein containing the inactivating C208A mutation was expressed and purified as previously described for the wild-type protein and analyzed via SDS-PAGE and LC-ESI-MS (Fig. S1). Purified protein (at ~1.1 mM) was incubated in a 1:1 ratio with 1 mM peptide for 1 h prior to crystallization by hanging drop vapor diffusion. Crystallization conditions were optimized from those used for apo spySrtA (PDB ID 3FN5), and are described in the Materials and Methods. From this, we succeeded in crystallizing and solving two structures of C208A spySrtA bound to the model peptides (P1’ position in bold) Abz-LPATAGK(Dnp)-NH₂ and Ac-LPATSG-NH₂ (Fig. 6A). The former is an example of a FRET quencher probe that is commonly used for monitoring sortase enzymatic activity, while the latter is a simplified target containing an acetyl (Ac-) cap and C-terminal primary amide (-NH₂). For both substrates, LC-ESI-MS was used to confirm that they were cleaved by wild type spySrtA in a model transacylation reaction (Fig. S2). Notably, we also crystallized C208A spySrtA with peptides containing the canonical LPXTG sequence.
(Abz-LPATGGK(Dnp)-NH₂ and fluorescently labeled 5-FAM-Ahx-LPATGG-NH₂), however, the crystals obtained were not of suitable diffraction quality.

Figure 6. The spySrtA complex structures with LPATS and LPATA peptides. (A) Structures of model peptides co-crystallized with spySrtA. (B) The spySrtA protein is in gray surface representation, and the non-covalently bound LPATS and LPATA peptides are in sticks and colored by heteroatom (N=blue, O=red, C=yellow/cyan as labeled). The N-terminal Abz moiety on the LPATA peptide and the P2' Gly residue for both peptides are not shown in order to focus on the target recognition sequences. The inset box shows a zoomed-in version of peptide binding, with all peptide positions labeled. The spySrtA side chains are shown as sticks and colored by heteroatom. The catalytic triad (H142-C208A-R216) are labeled. (C) The 2Fo-Fc electron density map for the LPATS peptide and catalytic triad is shown in blue mesh and rendered at 1.0 σ. The structure is shown as in the inset of (B). (D) Alignment of the the apo spySrtA (PDB ID 3FN5) and spySrtA-LPATS structures reveals an RMSD = 0.158 Å (508 main-chain atoms). The proteins are shown in cartoon representation and colored as labeled. The LPATS peptide is shown as sticks and colored by heteroatom.
For simplicity, we will hereafter refer to the solved enzyme-substrate complexes as spySrtA-LPATA and spySrtA-LPATS (Fig. 6B). All diffraction and refinement statistics for these complexes are in Table 1. In general, crystals grew stacked and were relatively unstable in traditional cryo solutions (e.g., with 10-20% (v/v) glycerol added). As a likely result of these challenges, the crystal ultimately used for spySrtA-LPATA structure determination contained pseudo-symmetry. We predict this may be due to lattice disruption during crystal harvesting. The space group of this crystal was \( P 2_1 2_1 2_1 \) and contained 2 protomers in the asymmetric unit. We refined it to a \( R_{\text{work}}/R_{\text{free}} = 0.21/0.24 \) at 1.4 Å resolution (Table 1). Relatively high R-factors are a consequence of pseudo-symmetry in crystal packing. Optimization of cryo conditions, namely using PEG 400 as a cryoprotectant, resulted in better quality diffraction data for the crystal used to solve the spySrtA-LPATS structure, as described in the Materials and Methods. This crystal diffracted to 1.4 Å resolution and the resulting structure was solved in space group \( P 2_1 \) to a \( R_{\text{work}}/R_{\text{free}} = 0.17/0.19 \), with two spySrtA molecules in the asymmetric unit (Table 1). The unit cell and space group are very similar between spySrtA-LPATS and apo spySrtA. 43

Alignment of chain (or protomer) A of spySrtA-LPATS with the two molecules of spySrtA-LPATA revealed very similar structures, with pairwise RMSD values for main-chain atoms of all protomers of both structures <0.13 Å (Fig. S3A). We were able to model all residues of the enzyme in protomer B of spySrtA-LPATA and protomer A of spySrtA-LPATS revealing an additional N-terminal helix not previously seen in the apo structure (Fig. S3B). Because of the large degree of similarity between these structures, unless otherwise noted, our analyses will focus on spySrtA-LPATS protomer A.
Stereochemistry of target recognition by spySrtA

We next used our peptide-bound crystal structures to analyze the stereochemistry of target recognition by Class A sortases. In both structures, we see clear peptide density and modeled the entire pentapeptide motif for all spySrtA protomers (Fig. 6B,C). Unbiased electron density maps, created by omitting the peptide atoms and running a round of refinement, confirm strong electron density for peptide residues (Fig. S3C). Alignment of spySrtA-LPATS with the two (A and B)
protomers of apo spySrtA revealed RMSD values for main chain atoms of: 0.158 Å (508 atoms) and 0.189 Å (541 atoms), respectively. The largest difference between these structures is an approximately 1 Å displacement in the backbone of the β7-β8+3, β7-β8+4, and β7-β8+5 loop residues (Fig. 6D). Here, superscript numbering refers to the residue position with respect to the catalytic C208 residue, as previously defined. 16 This suggests that very small structural rearrangements are needed in order to accommodate the target peptide.

We were able to model the 2-aminobenzoyl (Abz) moieties in the spySrtA-LPATA protomers, although the 2,4-dinitrophenyl lysine residue (K(Dnp)) was unresolved. In the A-protomer of spySrtA-LPATA we see a potential hydrogen bond between the 2-amino group of Abz and the carbonyl of P188 (Fig. S4A). While interesting, we do not consider this interaction to be critical for the binding of this substrate, as it is not observed in the B-protomer of the spySrtA-LPATA complex. This is further supported by the successful binding and co-crystallization of the Ac-LPATSG-NH₂ peptide, which lacks the Abz unit.

We next analyzed position-specific interactions in the LPATX motif of the CWSS. The highly conserved Leu residue at P4 interacts with a hydrophobic pocket formed by V186, V191, and V193 of the β6-β7 loop, as well as V206 in β7 and I218 in β8 (Fig. 7A). A similar pocket was previously identified in the NMR structure of saSrtA with a covalent peptidomimetic (LPAT*), PDB ID 2KID. 6 The proline residue in P3 interacts weakly via van der Waals interactions with V206 and A140, residues in the β4 and β7 strands, as well as M125 in the β3-β4 loop (Fig. 7A). The distances between these residues are of equal magnitude or shorter to those seen in the saSrtA-LPAT* structure where strong intermolecular nuclear Overhauser effects (NOEs) were observed that supported P3 Pro interactions with residues in the β4 and β7 strands (Fig. S4B). 6
There are several backbone atoms in the LPATX motif that form non-covalent interactions with residues in spySrtA (Fig. 7B). In both the LPATA and LPATS structures, the carbonyl oxygens of the P4 Leu and P3 Pro residues are hydrogen bonded with nitrogen atoms in R216, the catalytic arginine residue. In the LPATA complex, R216 also interacts with the P2 Ala carbonyl, whereas in the LPATS structure this carbonyl is rotated ~180° and interacting with solvent (Fig. 7B). In all structures, the orientation of the P2 and P1 residue side chains (AT, respectively) observed are rotated ~180° as compared to the saSrtA-LPAT* structure, agreeing more closely with the structure of B. anthracis SrtA (baSrtA-LPAT*) from the same group (Fig. S4C).6,19 As described above, the conformation observed in spySrtA-LPATA and spySrtA-LPATS is referred to as “Thr-in” to describe the P1 Thr side chain oriented toward the enzyme.16 The carbonyl oxygen of P1 Thr further interacts with the amide of C208A and side chain hydroxyl of T207 as well as the amide of H143, the residue immediately C-terminal to the catalytic histidine, H142 (Fig. 7B). The methyl group of the P1 Thr is oriented towards the side chain atoms of A140 and V206, and the side chain hydroxyl interacts with the amide of the catalytic C208A residue, as well as forms intrapeptide hydrogen bonds with its own amide and the carbonyl of the P3 Pro (Fig. 7B).

Finally, the P1’ Ser in spySrtA-LPATS interacts with a weakly negative ridge formed by the β7-β8 loop, specifically due to E212, the β7-β8+4 residue (Fig. 7C). A spatially analogous P1’ binding site, albeit with some differences in morphology and overall charge, was predicted in the previously reported saSrtA-LPAT* structure (PDB ID 2KID) (Fig. 7C). In our spySrtA-LPATS structure, we also observe a hydrogen bond with the hydroxyl group of the P1’ Ser and the carbonyl of I211 (Fig. 7B). This interaction is necessarily absent from the spySrtA-LPATA complex, and therefore we do not consider it a requirement for substrate binding. In general, the binding site for the P1’ position in spySrtA does not appear to be particularly selective, which is consistent with
our previous work on *S. pneumoniae* SrtA. Due to the observed similarities in these *Streptococcus* SrtA proteins, as well as our previous work investigating the β7-β8 loop in these proteins, we hypothesize that spySrtA is also non-selective at this position and can accommodate a wide variety of P1’ amino acids.

Overall, the observed location for the P1’ Ser, as well as the adjacent P1 Thr, renders the LPATS peptide ideally positioned for nucleophilic attack by the catalytic cysteine residue. Specifically, the methyl group of C208A in the spySrtA-LPATS structure is 3.4 Å from the P1 Thr carbonyl carbon (the corresponding distance in spySrtA-LPATA is 3.2 Å) (Fig. 7D). The scissile amide bond of the P1-P1’ linkage is also held in close proximity to the catalytic histidine (His142), which is consistent with the suggested role of this residue in facilitating proton transfers to the excised P1’ fragment and from the incoming lipid II nucleophile. Taken together, these observations support the validity of the spySrtA-LPATS and spySrtA-LPATA complexes as reasonable models for target recognition by Class A sortases that are consistent with the current understanding of the sortase catalytic mechanism.
Figure 7. Stereochemistry of the spySrtA-LPATS interaction. (A) The interactions of the P4 Leu and P3 Pro ligand residues with spySrtA are highlighted. The spySrtA enzyme is in cartoon and surface representation, with residues that interact directly with the Leu-Pro shown as sticks and colored by heteroatom (C=gray, S=golden yellow). The ligand is shown as a cyan cartoon, with the side chain sticks of Leu-Pro shown and colored by heteroatom (C=cyan, N=blue). (B) There are several non-covalent interactions between the LPATS ligand and spySrtA enzyme, shown as black dashed lines with distances labeled. There are also intramolecular interactions between the P1 Thr sidechain and its own amide, as well as the P3 Pro carbonyl oxygen, as labeled. The ligand is in stick representation and colored by heteroatom (C=cyan, N=blue, O=red). The spySrtA enzyme is shown as sticks and colored by heteroatom. The catalytic triad (H142-C208A-R216) is labeled. (C-D) The electrostatic potential surface maps for SrtA in spySrtA-LPATS (C) and saSrtA-LPAT* (D, PDB ID 2KID) were created using APBS in PyMOL and are shown from ±5 eV, with red = negative and blue = positive. The ligands are shown as sticks and colored by heteroatom. The location (C) or predicted location (D) of the P1’ site is circled, and the β7-β8 loop is labeled. (E) The distance between the CB atom of C208A and the C of the P1 Thr is shown as black dashed lines and labeled. This is the site of nucleophilic attack by C208. The structures are rendered as in (B), with the exception that spySrtA is in cartoon representation with only the side chain sticks of C208A shown.
Model of the acyl-enzyme intermediate

Next, we used our spySrtA-LPATS complex structure to model the acyl-enzyme intermediate (Fig. 8A). The model was constructed as described in the Materials and Methods. Briefly, coordinates for the cleaved peptide were determined and fit into the experimental electron density for spySrtA-LPATS. In addition, C208A was mutated \textit{in silico} to the wild-type cysteine and a round of refinement was run to validate the peptide geometry. We then performed a steepest descent energy minimization of the acyl-enzyme model to obtain the final geometry (Fig. S4D). The resulting acyl-enzyme model is therefore very similar to the spySrtA-LPATS structure, including nearly identical positions for the P4-P2 residues of the LPATS substrate (Fig. 8B). Slight differences were observed, however, in the case of the P1 Thr residue. As discussed above, the P1 Thr carbonyl in spySrtA-LPATS appears to be stabilized by the amides of H143 and C208A, as well as the side chain hydroxyl of T207. These interactions were largely maintained in our model, however a slight rotation of the P1 carbonyl towards T207 was observed (Fig. 8B). Specifically, in the geometry of the acyl-enzyme model, the T207 hydroxyl is 3.3 Å from the P1 Thr carbonyl (Fig. 8C). This distance is 3.4 Å in the solved structure of spySrtA-LPATS (Fig. 7B).

With respect to catalytic mechanism, a feature of the acyl-enzyme model that was also shared by both the spySrtA-LPATS and spySrtA-LPATA structures was the absence of a clear interaction between the P1 Thr carbonyl group and the putative catalytic arginine (R216) side chain. This is significant as this arginine has been proposed to stabilize high energy oxyanion intermediates generated during the sortase ligation reaction. \textsuperscript{2,6,46} The P1 Thr carbonyl in our acyl-enzyme model and solved structures was actually observed to point away from the R216 side chain, and the distance between these sites is >6 Å (Fig. 8C). Nonetheless, R216 was found to be
essential for spySrtA function, as mutating it to an Ala residue resulted in complete loss of enzyme activity when tested with model LPATG/S/A peptide substrates (Fig. 8D, Fig. S4E).

In terms of oxyanion stabilization, our structures are more consistent with a key role for the side chain hydroxyl of T207. This residue, along with the amides of H143 and C208, is ideally positioned to bind to the P1 Thr carbonyl and potentially stabilize tetrahedral oxyanion intermediates formed immediately prior to the acyl enzyme state and following nucleophilic attack by lipid II (Fig. 8C). This type of role for the Thr immediately preceding the catalytic Cys has indeed been suggested in previous computational studies. Moreover, sequence analysis of 400 sortase A enzymes in the NCBI database reveals that over 90% (363 total) contain a Thr residue immediately preceding the catalytic Cys, which suggests a fundamentally important role for this Thr such as stabilization of key reaction intermediates. Consistent with this hypothesis, we found that a T207A mutant of spySrtA exhibited a near total loss of enzymatic activity (Fig. 8D, Fig. S4E). Notably, a dramatic drop in enzyme activity has also been reported when mutating the corresponding Thr residue (T183) of saSrtA.
Building from our peptide-bound structures, we next explored the nature of the interaction between spySrtA and its in vivo nucleophile, lipid II. The lipid II molecule has been identified as the anchor for sortase-catalyzed attachment of many proteins to the bacterial cell wall and serves as a key precursor for the production of peptidoglycan. The nature of this peptidoglycan layer and the cell exterior as a whole is what differentiates Gram-positive and Gram-negative bacteria. Whereas Gram-negative bacteria have an inner membrane surrounded by a relatively thin peptidoglycan layer, followed by a second lipoprotein outer membrane, Gram-positive bacteria...
lack the outer membrane and contain a relatively thick peptidoglycan layer. Although there are exceptions and possible modifications, the main glycan moiety of the peptidoglycan layer consists of alternating $\beta$-1,4-$N$-acetylglucosamine (GlcNAc) and $N$-acetylmuramic acid (MurNAc) residues that are further crosslinked by peptide subunits.

The lipid II building block itself consists of the GlcNAc-MurNAc disaccharide attached to a polyisoprenoid membrane anchor and a pentapeptide stem that is linked via an amide bond to the C-3 D-lactyl ether of MurNAc. While the structure of the pentapeptide stem varies, a common sequence in Gram-positive bacteria such as S. pyogenes is: L-alanine, D-isoglutamine, L-lysine, D-alanine, and D-alanine. In many of these organisms, the L-lysine is subsequently modified by peptidyltransferases to create an *interpeptide bridge*, which are the residues that ultimately serve as the nucleophile for sortase-mediated ligation of surface proteins to the peptidoglycan layer. The nature of this interpeptide bridge is variable, but commonly includes L-Gly/Ala/Ser residues, e.g., for *S. aureus* = Gly, *Enterococcus faecalis* = Ala-Ala, *Streptococci* = Ala/Ser-Ala.

To visualize the interaction of spySrtA with lipid II and its related ligation products, we synthesized a model branched peptide representing the ligation of a LPATX substrate to the interpeptide bridge/pentapeptide stem portion of lipid II from *S. pyogenes* (Figs. 9A). Synthesis and characterization are described in the Materials and Methods and Supplementary Information. Specifically, this structure (LPAT-LII) possesses an Abz-LPAT fragment derived from the *Abz-LPATAGK(Dnp)-NH$_2$* substrate described above covalently linked to a lipid II mimetic via a dialanine interpeptide bridge. To our knowledge, there is no evidence of specific interactions between the glycan residues of lipid II and the sortase enzyme, therefore those portions were omitted and replaced with a simple acetyl group on the terminal L-alanine residue. We also note
that some structural heterogeneity in the interpeptide bridge/pentapeptide stem of *S. pyogenes* is likely. Examples of this include variable numbers of alanine residues in the interpeptide bridge and even low levels of hydroxylysine.\(^{49,52}\) However, our LPAT-LII model is consistent with structural features reported in the literature and should therefore be representative of a significant fraction of lipid II structures in *S. pyogenes*.\(^{20,49,53}\)

As a preliminary assessment of whether our LPAT-LII model was recognized by the enzyme, it was used in a model spySrtA-catalyzed reaction and found to be efficiently cleaved at the expected site between the Thr and Ala residues (Figs. 9B, S5). Indeed, we found LPAT-LII to react more rapidly than the related Abz-LPATAGK(Dnp)-NH\(_2\) peptide, suggesting that the added interpeptide bridge/pentapeptide stem portion may be enhancing binding and recognition by spySrtA (Fig. S5).

We next crystallized and solved the structure of C208A spySrtA non-covalently bound to our LPAT-LII mimetic. Two distinct conformations were observed, with the peptide Thr residue in both the “Thr-in” and “Thr-out” conformations previously observed in other SrtA structures (Fig. 9C-D).\(^6,54\) These structures will be referred to as spySrtA-LPAT-LII “Thr-in” and spySrtA-LPAT-LII “Thr-out.” Crystallization was performed similarly to the peptide-bound structures described above, and as in the Materials and Methods. Microseeding was used in this case to obtain crystals of suitable diffraction quality. Both the “Thr-in” and “Thr-out” structures crystallized in the space group \(P 2_1 2_1 2_1\) with one molecule in the asymmetric unit and to a resolution of 1.8 Å and 1.9 Å, respectively. The spySrtA-LPAT-LII “Thr-in” structure was refined to a final \(R_{\text{work}}/R_{\text{free}} = 0.18/0.21\), and the spySrtA-LPAT-LII “Thr-out” structure to a final \(R_{\text{work}}/R_{\text{free}} = 0.18/0.23\) (Table 1). Overall, the structures are very similar and the main chain atoms of spySrtA align with an RMSD = 0.082 Å (559 atoms).
In the spySrtA-LPAT-LII “Thr-in” structure, the stereochemistry of the LPATA portion is consistent with our peptide-bound structures (Fig. S6A). The main chain atoms align to the A- and B-protomer of spySrtA-LPATA with an RMSD = 0.218 Å (518 atoms) and 0.205 Å (495), respectively. Values are almost identical for spySrtA-LPAT-LII “Thr-out,” at 0.218 Å (497) and 0.207 Å (489) for the spySrtA-LPATA A- and B-protomers. The positions of the interpeptide bridge dialanine and ε-amine/ε-carbon of the L-lysine residue are also well conserved between the “Thr-in” and “Thr-out” structures (gray arrow in Fig. 9D). These sites make contacts with residues of the β7-β8 loop and appear to be stabilized by a hydrophobic pocket in spySrtA formed by four amino acids (I119 in α1, I144 and I147 in the β4-α2 loop, and V247 at the C-terminus) (Fig. 9E). Moving beyond the ε-carbon of L-lysine, there is more variability in the conformation of the pentapeptide stem between the two structures; this reflects the weaker electron density for these residues (Fig. 9C-D). Indeed, in both structures, there is only one observed non-covalent interaction with the lipid II pentapeptide and spySrtA enzyme, a hydrogen bond formed between the spySrtA α1 Y120 hydroxyl and the amide of the lipid II D-isoglutamine residue (Fig. S6B). In each, there are also multiple interactions with the lipid II pentapeptide and spySrtA enzyme of molecules related by symmetry (Fig. S6C).

Taken together, our crystallographic findings suggest that while the interpeptide bridge likely plays an important role in SrtA recognition of lipid II, the pentapeptide stem does not substantially interact with the enzyme. As noted above, the electron density for the pentapeptide stem was weaker than that of the LPAT segment and interpeptide bridge dialanine, suggesting flexibility in the stem region of the LPAT-LII ligand (Fig. 9C). Nonetheless, the clear electron density for the dialanine interpeptide bridge revealed a discrete binding site with potential implications for substrate binding outside of the standard LPXTG substrate motif, specifically at
the P2’ position. Interestingly, several predicted in vivo substrates of *S. pyogenes* and other streptococcal species possess LPXTGE motifs, with glutamic acid occupying this P2’ position.  

In our hands preliminary experiments suggest that spySrtA recognizes LPATGG and LPATGE peptides similarly, but additional work is ongoing to investigate P2’ specificity (data not shown).
Figure 9. The structure of spySrtA bound to a peptide model of the LPAT-lipid II ligation product. (A) Chemical structure of the Abz-LPAT-lipid II (LPAT-LII) model with the ligation site between the LPAT fragment and lipid II highlighted. (B) RP-HPLC chromatograms showing efficient cleavage (91% conversion) of the LPAT-LII model in the presence of spySrtA and an excess of alanine amide (A-NH₂) nucleophile. Percent conversion was estimated by comparing RP-HPLC peak areas for the unreacted LPAT-II substrate and the LPATA-NH₂ product. (C) The 2 Fo-Fc electron density maps for the “Thr-in” and “Thr-out” LPAT-LII peptides are shown at 0.6 σ (blue mesh, left figures), highlighting the specific, but relatively weak, density in the pentapeptide stem region. The LPAT-LII ligand is in sticks and colored as in (D). Unbiased Fo-Fc maps at 2.0 σ are also shown (green mesh, right figures), which were created by deleting the peptide density and running a round of refinement. The black arrows indicate the Thr side chain in both structures. (D) The structure of spySrtA and the LPAT-LII molecule. This complex was crystallized in both P1 “Thr-in” and “Thr-out” ligand conformations and both peptides are shown here. SpySrtA is very similar for both structures and alignment reveals RMSD = 0.082 Å (559 main chain atoms). Therefore, only the spySrtA protein for the “Thr-in” structure is shown (gray cartoon and in surface representation). The LPAT-LII residues are colored as labeled and by heteroatom (N=blue, O=red). The black arrow highlights the Cε atom of the Lys residue of LII, which is the atom at which the conformations of the two solved structures begin to vary, indicating flexibility in the LII pentapeptide. (E) Interactions at the interpeptide bridge of the LPAT-LII ligand are highlighted. Side chain atoms in spySrtA that form a hydrophobic interaction surface in the vicinity are shown as spheres and the residues are labeled. The “Thr-In” peptide is rendered as in (D) and the arrows point to the Ala-Ala residues of the interpeptide bridge, as labeled.
Molecular dynamics simulations of spySrtA bound to target peptides

During structure refinement and model building for spySrtA-LPAT-LII, we observed reduced electron density for the pentapeptide stem as compared to the LPAT sequence and interpeptide bridge, which suggested variations in conformational dynamics for different segments of the LPAT-LII ligand (Fig. 9C). To probe this further, as well as investigate the molecular dynamics of our other spySrtA substrate complexes, we ran ~900 nanosecond molecular dynamics simulations of apo spySrtA (PDB ID 3FN5), spySrtA-LPATA, spySrtA-LPATS, and spySrtA-LII “Thr-in” structures (Table S1). Briefly, MD simulations were performed in full atomistic detail with explicit water using the AMBER99SB*-ILDN force fields. 55 The starting structures were solvated with ~10,000 TIP3P water molecules in a cubic box with periodic boundary conditions. The system was neutralized with an ionic concentration of 150 mM. These simulations are described in further detail in the Experimental Procedures and Appendix A.

Overall, the root-mean-square deviation (RMSD) of backbone atom positions for spySrtA indicated that the enzyme remained stable over the course of all four simulations. (Fig. S7A). The root-mean-square-fluctuation (RMSF) of backbone atoms for spySrtA were also consistent with a well-defined 8-stranded antiparallel β-barrel sortase core structure, in that these regions are relatively inflexible over the course of the simulation, as compared to some α-helical and all loop regions (Fig. 10A). In all the peptide ligands, the LPATX sequences were also relatively inflexible. This was clearly evident in the visualization of representative frames taken over the course of each simulation (Fig. 10B), as well as the average RMSF of backbone atoms in each peptide (Fig. S7B). The RMSF of the P1’ Ser in the LPATS peptide was also similar to that of the P1’ Ala in either the LPATA or LPAT-LII simulations.
We also analyzed the distance distribution between the C208A methyl group (or Cβ atom) and that of the P1 Thr carbonyl C in the spySrtA-LPATA simulation, revealing that the most often sampled distance equals 3.8 Å (Fig. 10C). Surprisingly, our experimentally observed distance of 3.4 Å (Fig. 7D) was observed less than 5% of the time in the simulation; however, considering the C208A mutation and the standard C-S bond length of ~1.8 Å, this distribution of distances still positions the P1 Thr C in an ideal position for nucleophilic attack by the thiol group of the catalytic cysteine.

Finally, analysis of the average RMSF fluctuation of every non-hydrogen atom in the LPAT-LII ligand was consistent with increased conformational dynamics for the pentapeptide stem portion (Fig. S7C). We see a dramatic increase in flexibility in atoms in the LII pentapeptide, as compared to the LPAT and interpeptide bridge sequences (Fig. S7C). Specifically, this increase begins at the Cε atom of the lysine side chain, and gets progressively larger moving down the lysine side chain toward the pentapeptide stem. This was also clearly evident in the alignment of representative frames (taken every 45 ns) from the MD trajectory of the spySrtA-LPAT-LII system (Fig. 10B). Taken together, these molecular dynamics simulations strongly support our described structure-based conclusions.
Figure 10. Molecular dynamics simulations of spySrtA structures. (A) The average root-mean-square-fluctuations (RMSF) of backbone atoms in each residue are shown and colored as labeled. The secondary structure elements are indicated under the curve by arrows for β-strands and curved lines for α-helices. These are based on the apo spySrtA structure (PDB ID 3FN5) and residue boundaries fluctuate slightly amongst the structures. (B) Representative frames from every approximately 45 nanoseconds of simulation time are aligned to the original model for spySrtA-LPATA (top left, yellow peptide), spySrtA-LPATS (top right, cyan peptide), and spySrtA-LPAT-LII (bottom, green peptide) simulations. A clipped version of the spySrtA-LPAT-LII simulation is shown in the bottom right to highlight the relative inflexibility of the interpeptide bridge dialanine residues as compared to the rest of the pentapeptide stem. For all images in panel (B), the spySrtA protein is in gray cartoon, and the peptides are shown in stick representation and colored by heteroatom (N=blue, O=red). (C) The distribution of observed distances between the C208A CB (Cβ) and P1 Thr C (main chain carbonyl carbon) is shown as a function of its probability. The highest probability distance is approximately 3.8 Å.
Discussion

As we highlight in Fig. 5, there are multiple key states in the SrtA catalytic cycle when attaching a protein to the cell surface of Gram-positive bacteria. Facilitated by a conserved Cys-His-Arg triad, the apo enzyme (state 1) recognizes a motif within the CWSS on the C-terminus of a target protein (state 2), and cleaves the peptide between the P1 Thr and P1’ Gly residues (or other P1’ residues in vitro), presumably forming a tetrahedral oxyanion that resolves to generate a thioacyl-enzyme intermediate (state 3). Nucleophilic attack by the N-terminal amine of the interpeptide bridge of lipid II on the carbonyl carbon of the P1 Thr residue leads to a second tetrahedral oxyanion intermediate that collapses into the final ligation product and completes the transpeptidation reaction, whereby the initial target sequence (minus the P1’ residue and all residues C-terminal to this position) is covalently attached to lipid II (state 4). Using spySrtA as a model, we solved structures that experimentally show how the full LPXTX substrate is recognized by the enzyme (state 2) as well as how the final ligation product is accommodated within the enzyme active site (state 4). In addition, we used our peptide-bound structures and energy minimization to model the acyl-enzyme intermediate (state 3); thus, providing a nearly comprehensive structural view of the spySrtA catalytic mechanism.

Considered alongside other SrtA structures that contain bound substrate mimetics, the studies reported here both reaffirm certain common structural features and reveal new insights. As described above, we observe several position-specific interactions similar to those first reported for the peptidomimetic-bound structures of saSrtA and baSrtA. Our observed interactions at the P4 Leu, P3 Pro, P2 Ala, and P1’ Ala/Ser positions also support additional data on substrate selectivity in Class A sortases. Finally, in our LPAT-LII structure, we see both “Thr-in”
and “Thr-out” conformations, molecular orientations that have also been previously described. 2,6,54

However, apart from the orientation of the P1 Thr side chain, other aspects of the positioning of this residue reveal unique attributes of our spySrtA complexes that differ from prior work with baSrtA and saSrtA. 6,54 It was previously suggested that the carbonyl group of the P1 Thr may be stabilized by contacts with the highly conserved Arg residue that forms part of the ubiquitous Cys-His-Arg triad found in sortases. This proposed interaction would further allow the Arg side chain to stabilize tetrahedral oxyanion intermediates generated during the sortase-catalyzed transpeptidation reaction. While we do find that the conserved Arg (R216) of spySrtA is critical for enzyme function (Fig. 8D, Fig. S4E), and appears to play a role in positioning the LPXTX motif through direct contacts with the P4 and P3 carbonyl groups (Fig. 7B), we see no evidence for interactions with the P1 carbonyl. Indeed, the P1 carbonyl in our complexes is projected away from the R216 side chain, and instead forms interactions with a series of other sites (Fig. 7B, Fig. 8C), including the hydroxyl group of a conserved Thr residue (T207) adjacent to the active site Cys (C208). These same contacts would also appear to provide a suitable oxyanion hole for stabilizing high energy reaction intermediates, which is supported by our finding that a T207A mutant of spySrtA was essentially inactive (Fig. 8D, Fig. S4E).

Our observations with the P1 Thr indicate that further work on the exact role of the conserved Arg residue in sortase catalysis is warranted. Along these lines, intriguing results from a recent directed evolution study suggest that the conserved Arg in sortase A enzymes may primarily be responsible for substrate positioning and binding, as opposed to stabilization of catalytic intermediates. Specifically, an engineered variant of saSrtA was reported that is selective for a LMVGG substrate motif. 58 Remarkably, in this enzyme the conserved Arg of wild-type
saSrtA was mutated to Ser and yet it remained an efficient transpeptidase. While it is possible that this highly mutated saSrtA variant acquired a series of compensatory mutations that negated the need for Arg to stabilize high energy oxyanion intermediates, we would argue an alternate interpretation that the wild-type Arg is not critical for creating an oxyanion hole and rather its primary function is substrate binding and controlling substrate selectivity. The Arg to Ser mutation in the LMVGG-specific saSrtA variant is thus understood as contributing to a change in substrate selectivity as opposed to representing a fundamental change in the catalytic mechanism.

We anticipate that our work will also prove useful in the continued development of sortase-mediated ligation (SML) protein modification strategies. Structure-guided engineering efforts have already seen success in generating sortases with altered substrate selectivity or increased activity, as well as a saSrtA mutant that no longer requires a Ca\(^{2+}\) co-factor.\(^{10,16,59,60}\) Moving forward, further optimization of spySrtA and other Class A sortases for use in SML can be envisioned based on the molecular characteristics elucidated in the spySrtA complexes and related structures presented here. Additionally, the extended target binding cleft revealed in our spySrtA-LPAT-LII structures suggest that portions of the substrate outside of the LPXTX motif could make specific contacts with the spySrtA enzyme, for example residues in the P2’ site. A systematic exploration of how these positions impact enzymatic activity \textit{in vitro} may therefore be helpful in optimizing SML using spySrtA and other Class A sortases. A similar approach has already proven beneficial for saSrtA, where it is known that a P2’ Gly residue generally provides superior reactivity \textit{in vitro}.\(^{61}\)

In summary, this work reports the first crystal structures of spySrtA bound to an LPXTX substrate, as well as a model of the \textit{in vivo} ligation product involving lipid II. These structures reveal new details on substrate recognition by bacterial sortases, which may prove valuable for the
use of sortases as tools for protein engineering. More broadly, this work improves our understanding of the fundamental enzymology of this large and clinically-relevant class of bacterial enzymes.

**Experimental Procedures**

*Expression and purification of sypSrtA protein.* Wild-type sypSrtA, C208A sypSrtA, T207A sypSrtA, and R216A sypSrtA genes were recombinantly expressed using *Escherichia coli* BL21 (DE3) cells in the pET28a(+) vector (Genscript), as previously described. The wild-type sequence used matches that of the published sypSrtA structure, PDB ID 3FN5. Briefly, transformed cells were grown at 37 °C in LB media to an OD$_{600}$ 0.6-0.8, followed by induction using 0.15 mM IPTG for 18-20 h at 18 °C. The cells were harvested in lysis buffer [0.05 M Tris pH 7.5, 0.15 M NaCl, 0.5 mM ethylenediaminetetraacetic acid (EDTA)] and whole cell lysate was clarified using centrifugation, followed by filtration of the supernatant. Initial purification was conducted using a 5 mL HisTrap HP column (Cytiva), and wash [0.05 M Tris pH 7.5, 0.15 M NaCl, 0.02 M imidazole, 0.001 M TCEP] and elution [wash buffer with 0.3 M imidazole] buffers.

Following immobilized metal affinity chromatography, the His-tag was proteolyzed off the N-terminus of the C208A sypSrtA protein using Tobacco Etch Virus (TEV) protease overnight at 4 °C and a ratio of ~1:100 (TEV:protein). The proteins used for activity assays (wild-type, T207A, R216A) were not cleaved, consistent with our previous work. After collecting the flow-through of a second 5 mL HisTrap HP column [wash buffer identical to that described above], size exclusion chromatography (SEC) was conducted using a HiLoad 16/600 Superdex 75 column
(Cytiva) in SEC running buffer [0.05 M Tris pH 7.5, 0.15 M NaCl, 0.001 M TCEP]. Purified protein fractions corresponding to the monomeric peak were pooled and concentrated using an Amicon Ultra-15 Centrifugal Filter Unit (10,000 NWML). Protein concentrations were determined using theoretical extinction coefficients calculated using ExPASy ProtParam. Protein not immediately used was flash-frozen in SEC running buffer and stored at -80 °C.

The purity, monomeric state, and identity of purified enzymes were confirmed by SDS-PAGE, analytical SEC, and LC-ESI-MS, respectively. For LC-ESI-MS, analyses were performed on an Agilent 6545XT AdvanceBio Q-TOF system interfaced with an Agilent 1290 HPLC system. Separations upstream of the Q-TOF were achieved with a Phenomenex Aeris™ 3.6 μM WIDEPORE C4 200 Å column (100 x 2.1 mm) [H₂O (0.1% formic acid) / MeCN (0.1% formic acid) mobile phase at 0.3 mL/min, method: hold 10% MeCN 0.0-1.0 min, linear gradient of 10-90% MeCN 1.0-9.0 min, hold 90% MeCN 9.0-11.0 min, linear gradient of 90-10% MeCN 11.0-11.1 min, re-equilibrate at 10% MeCN 11.1-15.0 min)]. Deconvolution of protein charge ladders was achieved using Agilent MassHunter BioConfirm software (version 10.0). The expected and observed molecular weights for all proteins in this study were as follows: wild-type spySrtA (calculated average MW = 20657.5 Da, observed = 20657.6 Da), C208A spySrtA (calculated average MW = 18573.3 Da, observed = 18573.4 Da), T207A spySrtA (calculated average MW = 20627.3 Da, observed = 20627.5 Da), R216A spySrtA (calculated average MW = 20572.3 Da, observed = 20572.5 Da). Representative mass spectrometry data for wild-type and C208A spySrtA is also provided in Fig. S1.

**Peptide synthesis.** Model peptide substrates used in crystallization and/or enzyme assays with the general structure Abz-LPATXGK(Dnp)-NH₂ (Abz = 2-aminobenzoyl, Dnp = 2,4-dinitrophenyl,
NH₂ = C-terminal primary amide) were synthesized and purified as previously described. ¹⁶ The 
Ac-LPATSG-NH₂ peptide (Ac = acetyl, NH₂ = C-terminal primary amide) used for spySrtA-LPATS co-crystallization was purchased from Biomatik. 5-FAM-Ahx-LPATGG-NH₂ (5-FAM-Ahx = 5-carboxyfluorescein linked via an aminohexanoic acid linker, NH₂ = C-terminal primary amide) used in attempted co-crystallization studies was also purchased from Biomatik. Finally, the synthesis of Abz-LPAT-lipid II (LPAT-LII) was achieved via manual Fmoc solid phase peptide synthesis. Full experimental details for the preparation of LPAT-LII are provided in the Supplementary Information and Fig. S8.

HPLC and LC-MS characterization of spySrtA-catalyzed reactions. LPATS/LPATA/LPAT-LII peptide substrates (50 μM), alanine amide nucleophile (5 mM), and wild-type spySrtA enzyme (5 μM in the reaction with LPATS, otherwise 1 μM), were combined at room temperature and incubated for the times indicated. All reactions contained 10% (v/v) sortase reaction buffer (500 mM Tris pH 7.5, 1500 mM NaCl), as well as ≤1.1% (v/v) residual DMSO from the peptide substrate stock solutions. Reactions were analyzed using a Dionex Ultimate 3000 HPLC system interfaced with an Advion CMS expression mass spectrometer. Separations were achieved with a Phenomenex Kinetix® 2.6 μM C18 100 Å column (100 x 2.1 mm) [aqueous (95% H₂O, 5% MeCN, 0.1% formic acid) / MeCN (0.1% formic acid) mobile phase at 0.3 mL/min, gradients adjusted for each substrate to achieve separation between relevant reaction components].

Fluorescence Assay for Sortase Activity. Enzyme assays for assessing the reactivity of wild-type spySrtA versus the T207A and R216A mutants were conducted using a Biotek Synergy H1 plate reader as previously described (1). Briefly, Abz-LPATXGK(Dnp)-NH₂ peptide substrates (50 μM
final concentration) were incubated with hydroxylamine nucleophile (5 mM) and sortase enzyme (5 μM) at room temperature. All reactions contained 10% (v/v) 10x sortase reaction buffer (500 mM Tris pH 7.5, 1500 mM NaCl) and small amounts of residual DMSO (≤ 0.9% v/v) from the peptide stock solutions. The fluorescence intensity of each reaction well was measured at 2-min time intervals over a 2-hr period (λ_ex = 320 nm, λ_em = 420 nm, and detector gain = 75). All reactions were performed in triplicate, and fluorescence intensity (in relative fluorescence units, RFU) over time was plotted using Kaleidagraph 5.01.

**Crystallization of spySrtA complex structures.** The C208A spySrtA protein was crystallized at approximately 20 mg/mL, or 1.1 mM. Peptide (LPATA, LPATS, or LPAT-LII), at 1 mM final concentration, was incubated with protein in a 1:1 ratio at room temperature for approximately 1 h prior to crystallization by hanging drop vapor diffusion using a 500 μL well solution to protein solution ratio of 1:1, for a final drop volume of 4 μL (2 μL + 2 μL). Crystallization conditions were optimized using those for the wild-type apo protein. The crystallization conditions for the crystals used for data collection were (for all, containing C208A spySrtA): LPATA [0.1 M sodium acetate, 34% (w/v) PEG 8000, 0.1 M Tris pH 6], LPATS [0.1 M sodium acetate, 30% (w/v) PEG 8000, 0.1 M Tris pH 6], LPAT-LII “Thr-in” [0.15 M sodium acetate, 26% (w/v) PEG 8000, 0.1 M Tris pH 6], and LPAT-LII “Thr-out” [same conditions as LPAT-LII “Thr-in”]. Microseeding was used to obtain crystals of suitable diffraction quality for structure determination with the LPAT-LII-bound complexes, using initial crystals that grew in conditions of higher PEG 8000 concentration (>30% (w/v), consistent with the other conditions described). As described in the main text, glycerol was used as a cryoprotectant for the C208A spySrtA-LPATA crystal [cryo: crystallization conditions plus 12% (w/v) glycerol], but for the other crystals PEG 400 was used.
[cryo: 0.15 M sodium acetate, 10% (w/v) PEG 8000, 40% (w/v) PEG 400, 0.1 M Tris pH 6]. The crystals were flash-cooled by plunging into liquid nitrogen.

Data collection, structure determination, and protein analyses. Data were collected at the Advanced Light Source (ALS) at Lawrence Berkeley National Laboratory (LBNL) on beamline 5.0.1 and 5.0.2, at λ = 1.00004 nm or 0.97741 nm over 360°, with Δϕ=0.25° frames and an exposure time of 0.5 s per frame. Data were processed using the XDS package (Table 1). Molecular Replacement was performed using Phenix with spySrtA (PDB ID 3FN5) used as the search model. Refinement was performed using Phenix, manual refinement was done using Coot, and model geometry was assessed using MolProbity and the PDB validation server. Coordinates for the Abz moiety in C208A spySrtA-LPATA and LPAT-LII were initially determined using phenix.eLBOW from the SMILES (Simplified Molecular Input Line Entry System) strings rendered using ChemDraw. All crystal data and refinement statistics are in Table 1. Structural analyses and figure rendering were done using PyMOL. PDB accession codes for the structures presented here are in Table 1 and are (for all, containing C208A spySrtA): LPATA (7S51), LPATS (7S40), LPAT-LII “Thr-in” (7T8Y), and LPAT-LII “Thr-out” (7T8Z).

Molecular dynamics simulations of spySrtA. All molecular dynamics simulations were performed using GROMACS 2020.4 with the AMBER99SB*-ILDN force fields. Additional details and relevant references are in Appendix A.
For energy minimization of the spySrtA-LPAT model, a steepest descent energy minimization was performed on the solvated system with a maximum force tolerance of 500 kJ/mol/nm. The steepest descent converged in 2998 steps.
Chapter 2

Investigations into Alternative Substrate Binding and Cleavage

Brandon A. Vogel, Jadon Blount, James McCarty,¹ John M. Antos,¹ Jeanine F. Amacher¹

¹Department of Chemistry, Western Washington University, Bellingham, WA, 98225

Brandon A. Vogel’s contributions to this work include the preparation and purification of the proteins used, the preparation of the peptides used in enzyme assays, and the execution of the mass spectrometry assays. Jadon Blount assisted in the completion of the molecular dynamics simulation and assisted in the completion of protein preparation and purification.
Introduction

Since the discovery of the first Sortase enzymes from *S. aureus* in 1999, there has been an explosion of research leveraging the unique utility of these enzymes. Researchers have been able to solubilize the enzyme, use it in protein engineering applications, and use it as a target for therapeutics. While there has been a great deal of work done over the last 20 plus years finding downstream applications of the enzyme, there are still many fundamental aspects of the enzyme we have yet to fully comprehend.

In 2018, our collaborators in the Antos lab at Western Washington University published a paper characterizing the cleavage profile of many different sortase enzymes hoping to leverage the information gathered to expand the protein engineering capabilities of class A sortases. They were able to discover unique substrate preferences between sortase homologues and an unexpected discovery of a cleavage motif that occurred one residue away from the expected P1/P1’ site, at a P1’/P2’ site, this site is depicted in Figure 11. This alternative cleavage did not occur with all homologues and was specific to certain peptides only.

Figure 11. Abz-LPATGGK-Dnp peptide. The full peptide shown here is Abz-LPATGGK(Dnp)-NH₂ (Abz = 2-aminobenzoyl, Dnp = 2,4-dinitrophenyl, NH₂ = C-terminal primary amide). The LPATG portion is the canonical cell wall sorting signal for class A sortases.
Our lab became interested in finding out more as to why this alternative cleavage was taking place. We decided to investigate this phenomenon further hoping to leverage our expertise in protein structure to determine the causative factors involved in this alternative cleavage event. Our original hypothesis was that there might be a residue or structural motif that is the main causative agent in alternative cleavage. With this in mind we ran a multi-sequence alignment and found that there was a highly conserved His residue located adjacent to the catalytic His residue (Fig. 13). We were also curious about why the *L. monocytogenes* had a differing cleavage motif as to that of the *Streptococcus* enzymes. To do this we began by recapitulating the results from the Antos lab with wild type proteins in 18-hour end-point mass spectrometry assays. We then began to create point mutations at specific residues of interest and chimeric enzymes to investigate both residue specific and structure specific contributions to alternative cleavage. Then

![Synthesized peptides](image)
we tested these mutations in the aforementioned mass spectrometry assays in the presence of three different peptides; one containing the canonical LPATG, one with LPATY, and one with LPATL depicted in figures 11 and 12.

What we found here is still ambiguous and requires further investigation. We found that there is no one residue or specific structure in particular that contributes solely to alternative cleavage but, as an aside, perhaps these mutations could be leveraged to modulate cleavage of peptide sequences of interest. Here we report the findings of these end point mass spectrometry assays, specifically that the conserved histidine residue adjacent to the catalytic residue is not utilized in an alternative mechanism leading to alternative cleavage but that it can be mutated to decrease the amount of P1’/P2’ cleavage. We also report the finding that incorporating the β7/8 loop from L. monocytogenes into the wild type

Streptococcus pyogenes sortase A construct reduced alternative cleavage products and increased cleavage at the P1/P1’ site of the non-canonical LPATLG and LPATYG peptides.

Results

Alternative cleavage assays
Our collaborator, Dr. Antos, finding that some class A sortases cleave their substrate in the P1’/P2’ position instead of the canonical P1/P1’ raised questions as to how this occurs and if this can be modulated. The first step in the process of elucidating this newly discovered cleavage was to investigate the mode of action that Sortases are using to recognize their substrate and subsequently cleave at a site distal to the P1/P1’ position. To do this we first wanted to recapitulate results from the Antos lab and then build on that knowledge using point mutations and chimeric enzymes at sites that could contribute to a possible alternative mechanism or reveal structural contributions to substrate binding. We were successful in recapitulating both the

<table>
<thead>
<tr>
<th>A</th>
<th>P1/P1’ cleavage</th>
<th>Abz-LPATGGK-Dnp</th>
<th>Abz-LPATLGK-Dnp</th>
<th>Abz-LPATYGK-Dnp</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT Spy</td>
<td>80.7 ± 0.5</td>
<td>5.59 ± 0.2</td>
<td>42.7 ± 1.0</td>
<td></td>
</tr>
<tr>
<td>WT lm</td>
<td>75.4 ± 1.6</td>
<td>5.37 ± 0.7</td>
<td>40.8 ± 1.1</td>
<td></td>
</tr>
<tr>
<td>WT Saga</td>
<td>77.0 ± 1.0</td>
<td>7.12 ± 0.9</td>
<td>10.2 ± 0.8</td>
<td></td>
</tr>
<tr>
<td>H143A_Spy</td>
<td>77.8 ± 0.3</td>
<td>4.30 ± 0.6</td>
<td>10.8 ± 0.4</td>
<td></td>
</tr>
<tr>
<td>spy(mono)7/8</td>
<td>78.4 ± 0.4</td>
<td>26.7 ± 2.0</td>
<td>67.3 ± 0.5</td>
<td></td>
</tr>
<tr>
<td>lm(spy)7/8</td>
<td>21.9 ± 0.1</td>
<td>Unobserved</td>
<td>Unobserved</td>
<td></td>
</tr>
<tr>
<td>I211P_Spy</td>
<td>77.9 ± 1.7</td>
<td>7.71 ± 0.3</td>
<td>34.7 ± 0.8</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B</th>
<th>P1'/P2’ cleavage</th>
<th>Abz-LPATGGK-Dnp</th>
<th>Abz-LPATLGK-Dnp</th>
<th>Abz-LPATYGK-Dnp</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT Spy</td>
<td>Unobserved</td>
<td>73.6 ± 1.9</td>
<td>18.2 ± 0.8</td>
<td></td>
</tr>
<tr>
<td>WT lm</td>
<td>Unobserved</td>
<td>6.81 ± 1.3</td>
<td>1.30 ± 0.2</td>
<td></td>
</tr>
<tr>
<td>WT Saga</td>
<td>Unobserved</td>
<td>30.1 ± 3.8</td>
<td>10.3 ± 0.9</td>
<td></td>
</tr>
<tr>
<td>H143A_Spy</td>
<td>Unobserved</td>
<td>41.2 ± 5.8</td>
<td>7.25 ± 0.3</td>
<td></td>
</tr>
<tr>
<td>spy(mono)7/8</td>
<td>Unobserved</td>
<td>23.6 ± 1.4</td>
<td>3.56 ± 0.4</td>
<td></td>
</tr>
<tr>
<td>lm(spy)7/8</td>
<td>Unobserved</td>
<td>Unobserved</td>
<td>Unobserved</td>
<td></td>
</tr>
<tr>
<td>I211P_Spy</td>
<td>Unobserved</td>
<td>62.3 ± 4.6</td>
<td>18.2 ± 0.5</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Percent conversion table. Presented here is the percent of standard cleavage (A) and percent alternative cleavage (B) data from 18-hour endpoint MS-MS assays. Table A shows the triplicate average of the percent of standard cleavage product formation. Table B presents the triplicate average of the alternative cleavage product formation.
published and unpublished results of the Antos lab and were able to then begin experiments to expand on this information evidenced in Table 2.

Table 1A presents the results of 18-hour endpoint assays performed by Q-TOF MS showing the cleavage that takes place between the P1/P1’ position of the substrates LPATG, LPATL, and LPATY. All of the wild type enzymes behaved as expected in the presence of native LPATG substrate, cleaving between 75-80% of available peptide at the P1/P1’ position. There is a marked decrease in activity at the P1/P1’ position for the wild type sortases here in the presence of LPATL, with all wild types cleaving less than 10% at the same position. Similarly, there is a large decrease in cleavage product formation in the presence of LPATY with WTSpySrtA and WTlmSrtA cleaving 42.7% and 40.8% respectively while WTSagSrtA cleaved approximately 10% of available substrate.

Presented in Table 2 are the point mutations and loop swap chimeric sortase enzymes. These are labeled as H143A_Spy, I211P_Spy, lm(spy)7/8, and spy(mono)7/8. The point mutations are H143A_Spy and I211P_Spy mutating a histidine to an alanine and a isoleucine to a proline respectively. The H143A mutation was performed in order to test if this histidine residue acts as an alternative catalytic residue during alternative cleavage or if it plays a role at all. The I211P mutation was performed in order to impart a more constrained character to the 7/8 as this residue is a proline in the L. monocytogenes homolog which is not observed to perform alternative cleavage. The loop swaps lm(spy)7/8 and spy(lm)7/8 are a L. monocytogenes sortase A construct with the β7/8 loop from S. pyogenes and vice versa for the spy(lm)7/8. These chimeric enzymes were constructed in order to evaluate the effects that the β7/8 loop may have on alternative cleavage overall, as it has been shown in previous work that the β7/8 loops are likely important to substrate selectivity in class A sortases.
Of the point mutations and chimeric enzymes all but lm(spy)7/8 were largely able to retain wild type activity at the P1/P1’ position with H143Aspy and spy(mono)7/8, and I211Pspy cleaving approximately 78% of available substrate. The chimeric lm(spy)7/8 was only capable of cleaving around 22% of substrate from the native LPATG peptide and lost the ability to cleave non-canonical peptides entirely. Notably, the spy(mono)7/8 construct presented an increase in cleavage of the LPATL and LPATY substrate at the P1/P1’ position cleaving approximately 27% and 67% respectively.

Table 1B presents the percentage of substrate cleaved at the P1’/P2’ position. At the P1’/P2' position none of the wild type, chimeric, or point mutation constructs had observable alternative cleavage peaks in the presence of LPATG. WTlmSrtA appeared to be the least promiscuous, cleaving around 6% of the LPATL and only 1.3% of the LPATY substrate at the P1’/P2’ position. WTsagSrtA was able cleave approximately 30% of the LPATL and 10% of the LPATY substrates at this position. WTspySrtA however, has near wild type activity cleaving close to 74% of available substrate at the P1’/P2’ position with LPATL but had markedly lower alternative cleavage of LPATY only cleaving around 18% of available substrate.

Of the point mutations I211P_Spy cleaves around 62% of LPATL and 18% of LPATY at the alternative position. This is about an 11% reduction in alternative cleavage from the wildtype enzyme with LPATL whereas the enzyme had nearly no catalytic change with LPATY. The H143A_Spy construct saw a large reduction in total cleavage of both LPATL and LPATY cleaving approximately 41% of LPATL and 7% of LPATY, a 32% and 11% reduction respectively.

*Molecular Dynamics Simulations*
In order to probe the effects of ligand dynamics and conformational changes that are contributing to alternative cleavage we have begun to perform molecular dynamic simulations. Our first completed simulation to date is that of SpySrtA complexed with an Ace-LPATLA-Nme of which representative measurements are shown in Figure 14. There is more work to be done with this simulation, but these measurements are compelling considering the simulation is stable and the peptide occupies this conformation in the binding cleft for nearly the entirety of the 1 µs simulation.

We observe that the P4 Leu begins the simulation just outside of the binding cleft before it moves down and finds the canonical hydrophobic binding pocket described elsewhere.²,¹¹,¹⁸ In both Figures 14A and 14B the P1 Thr is occupying a conformation that is solvent oriented, previously described as the “Thr out” conformation facilitating the overall puckered shape of the peptide that is observed.²,⁶,¹¹,¹⁸,⁵⁴ Within this puckered, Thr out conformation there is enough space for the P1’ Leu side chain to occupy a position analogous to where the Thr residue would reside if it were in the Thr in conformation. What we then observe is numerous hydrophobic contacts as seen in Figure 14C where Leu113, Leu118, Ala140, and Val206 form what appears to be a hydrophobic pocket or groove. We also observe potential hydrophobic contacts made between the P3 Pro and the side chain of Leu113, as well as a possible hydrophobic interaction between the methyl side chain of the P2 Ala and the side chain of the P1’ Leu. Furthermore, we also observe a hydrogen bond between the carbonyl oxygen of the catalytic Cys208 and the His143. The Cys208 is then oriented close to the amide bond of the P1’ Leu as depicted in Figure 14D.
Figure 14. Measurements from Molecular dynamics simulations. Presented here are measurements taken from a frame of our molecular dynamics run of SpySrtA complexed with a modeled Ace-LPATLA-Nme peptide. A) shows the LPATLA peptide bound within the binding cleft of the enzyme, occupying a “puckered” conformation where the Thr residue is oriented away from the enzyme toward solvent and the P1’ Leu side chain is occupying a position similar to that of the Thr side chain when it is oriented toward the enzyme. B) Presents the peptide in the binding cleft, here we observe that the β2/3 loop which occupies more of a helical structure is in close proximity to the peptide bringing the Leu113 side chain residue close enough to make hydrophobic contacts with the P1’ Leu of the peptide. C) Here we observe that the P1’ Leu is oriented into a hydrophobic pocket formed by Leu113, Leu118, Ala140, and Val206 along the base of the binding cleft. This likely contributes to increased stabilization of this conformation of the peptide. D) Observed here is a hydrogen bonding interaction between the His143 residue and the catalytic Cys208 which, may orient and provide stability to the β7/8 loop, keeping the Cys close to the scissile bond between the P1’ Leu and P2’Ala.
Discussion

Investigating the alternative cleavage of substrates by *S. pyogenes*, *S. agalactiae*, and *L. monocytogenes* revealed limited information into what causes this specific cleavage motif but did lead to further questions and experiments regarding substrate orientation being the likely causative agent of alternative cleavage. Upon reviewing the alternative cleavage data, we see that we were able to successfully purify soluble Sortase A proteins from all wild type constructs of interest that largely retained expected wild type activity. That activity ranges from approximately 75-80% cleavage of available substrate at the P1/P1’ position as shown in Table 2a.

We were primarily intrigued by a particular residue that appears to be highly conserved among class A sortases at first, that being a His residue in the 143 (H143) position of WT*SpySrtA*. This residue is located next to the catalytic residue of H142 and led us to question whether it may play a role in alternative cleavage. It was thought that perhaps H143 was contributing to an alternative mechanism where instead of H142 deprotonating the catalytic cysteine, the H143 was actually responsible for this deprotonation or perhaps there was a different mechanism altogether utilizing this residue. What we found was that when we created an H143A mutation in the SpySrtA construct we were able to retain nearly wild type catalytic function with successful cleavage of approximately 77% of native LPATGG substrate at the P1/P1’. We also found that this mutation contributed to a nearly 32% reduction in cleavage of the LPATYG substrate at the same canonical P1/P1’ position.

The H143A SpySrtA mutant also provided an interesting result in that it did in fact reduce the alternative cleavage at the P1’/P2’ position with LPATLG from 73.6% to 41.2%
little change in the presence of LPATYG. These results indicate that there is likely no alternative mechanism as we would have expected total abrogation of alternative cleavage at the P1′/P2′ position. However, there is some role that is played at this residue as there is a significant reduction in off target catalytic function that may be worth exploring since its mutation does not lead to reductions in overall catalytic function.

Upon further investigation of the H143 residue, the tau nitrogen of the imidazole side chain appears to form a critical hydrogen bond with the oxygen of the carbonyl group of the catalytic cysteine residue. This interaction appears to contribute to the stabilization of the β7/8 loop and orient the cysteine toward the binding cleft, into the P1/P1′ position. In molecular dynamics simulations with an LPATLA substrate we observe that the P1′ Leu is stabilized by a hydrophobic groove formed by Leu113, Leu118, Ala140, and Val206 allowing for a strained conformation of the substrate that is shifted in the binding pocket depicted in Figure 14. This shifting brings the catalytic cysteine closer to the P1′/P2′ position of the substrate allowing for the alternative cleavage at this position. When His143 is mutated to Ala, the enzyme may lose this hydrogen bond and allow for more flexibility in the 7/8 loop therefore allowing the catalytic Cys to come closer to the P1/P1′ bond and cleave as it normally would. L. monocytogenes sortase A does not allow for alternative cleavage in the same way as its Streptococcus homologs because it contains a highly charged β7/8 loop that directly interacts with the P1′ position, disallowing the presence of hydrophobic residues at this position. Thus, there is a massive decrease in catalysis of LPATL as seen in Table 1. Tyrosine is only partially accommodated due to its relatively polar nature.

Another idea that we thought might contribute to alternative cleavage had to do with the β7/8 loop since it is known to play a role in substrate recognition.16,18 What we noticed from the
previous work of our collaborator was that WTlmSrtA was capable of alternative cleavage, albeit at a low total percentage, of the bulky hydrophobic side chains of Phe and Tyr but not of Leu. To further investigate this phenomenon we created three mutations, a point mutation of I211P SpySrtA, and chimeric β7/8 loop swaps of SpySrtA and lmSrtA aptly named spy(mono)7/8, WTSpySrtA with the WTlmSrtA B7/8 loop, and the lm(spy)7/8 where the WTlmSrtA construct has the WTSpySrtA β7/8 loop. Our synthesis of the LPATFGK peptide failed due to the peptide crashing out in solution and as such was not tested here.

Of the chimeric enzymes the lm(spy)7/8 construct turned out to be nearly completely inactive, likely due to overall protein destabilization with such a large mutation. However, the spy(mono)7/8 construct was found to be active and provided unexpected results. This construct was able to retain near WTSpy activity in the presence of LPATGG cleaving approximately 78% of available substrate at the P1/P1’ position but also showed increased catalytic activity at the same position with both LPATLG and LPATYG with increases of 21.1% and 24.6% respectively over the WTSpySrtA construct.

The spy(mono)7/8 construct presented overall reductions in alternative cleavage. In the presence of the LPATLG substrate there was a 50% reduction as compared to WTSpySrtA and in the presence of LPATYG there was a 14.6% reduction. This reduction is expected since incorporating the lmSrtA β7/8 loop was expected to impart characteristics of the WTlmSrtA construct since this loop is critical to binding and catalysis. We may be able to explain the increase in canonical P1/P1’ cleavage of the spy(mono)7/8 construct due to increased loop constraints as the lmSrtA includes a Pro residue which we thought would lead to decreased entropy of the loop and therefore increased substrate interactions. To test this idea, we decided to look at the I211P SpySrtA mutant.
The I211P SpySrtA mutant was able to retain much of its WT Spy catalytic function as seen in Table 1a. And seen in Table 1b we also note that this construct was largely able to retain the same alternative cleavage functionality as that of WT Spy. This indicates that the Pro residue alone has no real effect on alternative cleavage and the polar and charged residues that contribute to the loop characteristics of ImSrtA as a whole are likely contributing factors to ImSrtA reduced alternative cleavage potential.

To summarize this alternative cleavage work showed that certain residues are not in fact the sole contributing factors of an alternative mechanism leading to alternative cleavage but may have some yet identified role in protein or protein-substrate stabilization and there are likely large structural motifs that playing roles in alternative cleavage. When looking at WT Spy SrtA, H143 likely is playing a role in stabilizing the catalytic C208 residue. This, in conjunction with the hydrophobic pocket formed by L113, L118, A140, and V206 interacting with the hydrophobic side chain in the P1’ position like the Leu presented here, increases the stability of a puckered conformation that facilitates the cleavage at the P1’/P2’ site. The work presented here also led us to begin to think more about the role of the substrates themselves in alternative cleavage as we are now investigating, in silico, the different possible conformations of the substrate in the binding pocket.

Materials and Methods

Expression and purification of Sortase proteins. Wild-type (WT) spySrtA, WT SagSrtA, WT ImSrtA, H143A spySrtA, I211P spySrtA, and the chimeric spy(mono)7/8 and Im(spy)7/8 genes were recombinantly expressed using Escherichia coli BL21 (DE3) cells in the pET28a(+) vector (Genscript), as previously described in chapter 1 and other work. The wild-type sequence used
matches that of the published spySrtA, SagSrtA, and ImSrtA structure, PDB ID 3FN5, 7S56, 5HU4.\textsuperscript{20} Briefly, transformed cells were grown at 37 °C in LB media to an OD\textsubscript{600} 0.6-0.8, followed by induction using 0.15 mM IPTG for 18-20 h at 18 °C. The cells were harvested in lysis buffer [0.05 M Tris pH 7.5, 0.15 M NaCl, 0.5 mM ethylenediaminetetraacetic acid (EDTA)] and whole cell lysate was clarified using centrifugation, followed by filtration of the supernatant. Initial purification was conducted using a 5 mL HisTrap HP column (Cytiva), and wash [0.05 M Tris pH 7.5, 0.15 M NaCl, 0.02 M imidazole, 0.001 M TCEP] and elution [wash buffer with 0.3 M imidazole] buffers.

The proteins were used in activity assays and were not cleaved, consistent with our previous work.\textsuperscript{16,18} After collecting the flow-through of a second 5 mL HisTrap HP column [wash buffer identical to that described above], size exclusion chromatography (SEC) was conducted using a HiLoad 16/600 Superdex 75 column (Cytiva) in SEC running buffer [0.05 M Tris pH 7.5, 0.15 M NaCl, 0.001 M TCEP]. Purified protein fractions corresponding to the monomeric peak were pooled and concentrated using an Amicon Ultra-15 Centrifugal Filter Unit (10,000 NWML). Protein concentrations were determined using theoretical extinction coefficients calculated using ExPASy ProtParam.\textsuperscript{62} Protein not immediately used was flash-frozen in SEC running buffer and stored at -80 °C.

The purity, monomeric state, and identity of purified enzymes were confirmed by SDS-PAGE, analytical SEC, and mass spectrometry, see Appendix B (Table S2) for mass spectrometry data.

\textit{Peptide synthesis.} Model peptide substrates used in crystallization and/or enzyme assays with the general structure \textit{Abz-LPATXGK(Dnp)-NH\textsubscript{2}} \textit{(Abz = 2-aminobenzoyl, Dnp = 2,4-dinitrophenyl, \textit{NH\textsubscript{2} = C-terminal primary amide}) were synthesized and purified as previously described.}\textsuperscript{16}
HPLC/MS-MS endpoint assays. The buffer used in the mass spectrometry 18-hour endpoint assays were completed as previously described, in 10% v/v sortase reaction buffer (500 mM Tris pH 7.5, 1500 mM NaCl), 10 mM hydroxylamine, 10 mM CaCl$_2$. The working concentration of all peptides during the reactions was 200 µM and the working concentration of the enzymes was 25 µM. DMSO concentration was brought to 5%. Reactions were run at room temperature and monitored using an Agilent AdvanceBio 6545XT Q-TOF mass spectrometer interfaced with an Agilent 1290 UHPLC. Separations upstream of the mass spectrometer were achieved with a Phenomenex Kinetex C18 column (2.6µm, 100 Å, 100 x2.1 mm). Samples were analyzed in a 60:40 H2O/ACN with 0.1% formic acid mobile phase with a 10-90% linear gradient to ACN with 0.1% formic acid. Data was analyzed using the Chemstation Masshunter software suite.

Molecular dynamics simulations of spySrtA. GROMACS 2020.4 (GROMACS development teams at the KTH Royal Institute of Technology and Uppsala University) with the AMBER99SB*-ILDN force fields was used and simulations were run for 1000 ns. The starting protein structures were solvated with TIP3P water molecules in a cubic box with periodic boundary conditions, and using a neutralizing ionic concentration of 150 mM (Table S3). All peptides in the starting models were “capped” by acetylation on the N-terminus and amidation on the C-terminus to neutralize charges. The N-terminus of spySrtA was also acetylated; however, because the C-terminus of the protein was included in all tested proteins, we kept the charge on the terminal residue in our simulations. The system was first equilibrated in an NVT ensemble for 100 ps, then in an NPT ensemble for 5000 ps. The starting models used for each simulation were derived from spySrtA-LPATA (7S51), and peptide residues were mutated (and/or shifted) using Coot. Additional details and relevant references are in Appendix B. 69–81
Chapter 3

Summary and Future Directions

In the first chapter of this work, we presented a paper that we published in 2022 investigating details of target recognition by class A sortases of *Streptococcus pyogenes*. In doing so we elucidated the first successful crystal structures of *S. pyogenes* class A sortases non-covalently bound to substrate and product mimics. We also utilized our subsequent structures in molecular dynamics simulations to investigate the dynamics of ligand binding as well as model the acyl-enzyme intermediate of spySrtA-LPAT. This work provided a comprehensive structural view of the spySrtA catalytic mechanism and supported previous data that identified interactions with the P4-P2 and P1′ sites that determine selectivity. We also presented evidence that the commonly accepted catalytic R216 residue that contributes to the stabilization of the tetrahedral oxyanion intermediate may not be critical to this purpose but serves as a critical residue for substrate selectivity. This was evidenced by our crystallographic and the modeled acyl-enzyme intermediate data which do not show any interactions between the R216 side chain and the P1 Thr carbonyl and that the P1 carbonyl in our complexes is in fact projecting away from the side chain of R216. We did, however, observe that there is an interaction between the hydroxyl group of a conserved T207 residue adjacent to the catalytic C208 and the P1 carbonyl. This residue, along with the amides of H143 and C208 are ideally positioned to stabilize the tetrahedral oxyanion intermediate.

The second chapter of this work presented an investigation into the causative agents of the observed class A sortase alternative cleavages. We successfully recapitulated the previous
results of our collaborators and built on that work. We showed that ImSrtA does not participate in alternative cleavage with the same substrates that *Streptococci* homologs do. In completing this work, we found that there does not seem to be a single residue that contributes solely to alternative cleavage but that this is likely a substrate sequence dependent effect. It appears that when there is a sufficiently hydrophobic sidechain occupying the P1’ position, as in Leu, the substrate will occupy a puckered conformation allowing for the amide bond of the P1’/P2’ position to be favored during catalysis. We found limited evidence that H143 may play a role in stabilizing this reaction, but further investigation is required to know if it participates in alternative cleavage.

Our work in these projects has led to more questions and cause for more experiments to further investigate how alternative cleavage is taking place and if it can be leveraged in engineering applications. To date, we have completed preliminary $^{15}$N NMR HSQC experiments where we have successfully prepared and received spectra for a $^{15}$N labeled a catalytically inactive C208A *Streptococcus pyogenes* mutant (C208A SpySrtA). Here we first gathered a spectrum of C208A SpySrtA without any peptide then titrated in an LPATA peptide to see if there was any change in peak positions within the spectra. These peak shifts would indicate a change in the electronic environment of the labeled residues and thus are likely interacting with the peptide. We did in fact observe change in the spectra when peptide was titrated in, this showed that $^{15}$N NMR HSQC experiments are a viable technique to elucidate which residues are exactly interacting with the substrate. However, an issue arose in that we found no spectra of the protein of interest that currently exists, and so peak assignments were not possible. This leaves an opportunity for future researchers to assign peaks and continue experiments with both native
and alternative substrates and even extend the technique to other sortase enzymes further
elucidating how sortases interact with substrate.

Another future direction for this work is the continuation of in silico experiments. Those
being molecular dynamics simulations of WTlmSrtA with peptides of interest, those being
peptides like LPATLG, which are alternatively cleaved by WTSpySrtA. This may allow us to
gain insight into any differences in the structures or residues that may be preventing WTlmSrtA
from this alternative cleavage function.

A final experiment that is planned is the testing of a H142A SpySrtA mutant. This is the
mutation of the catalytic histidine to an alanine which will then be tested in the presence of both
native substrate and alternatively cleaved substrates, like LPATL, to see if there is any formation
of alternative cleavage products. If it is found that this mutation results in alternative cleavage
products, then we would know that there is another residue contributing to an alternative
mechanism.

Overall, we believe that the revelations presented here will go on to better inform the
engineering of sortase variants to expand sortase mediated ligation strategies. We also believe
that what we have shown here has improved the understanding of the function of these clinically
relevant enzymes.
Works Cited


(35) Raz, A. Streptococcus Pyogenes Sortase Mutants Are Highly Susceptible to Killing by Host Factors Due to Aberrant Envelope Physiology. PLoS One 2015, 10 (10).


(44) Bradshaw, W. J. Molecular Features of the Sortase Enzyme Family. FEBS J 2015, 282 (11), 2097–2114.


Supplementary Information Text

Supplemental Methods

**Molecular dynamics simulations.** Nonstandard residues of the lipid II pentapeptide or LPAT-LII ligand that are not represented by the standard AMBER force field were described by the General Amber Force Field (GAFF2) (2). Partial charges for the ligand atoms were obtained using the restrained electric potential fitting method (RESP) with the molecular electric potentials computed at the HF/6-31G* level of theory (3). Quantum chemistry calculations were performed using GAMESS with RESP fitting performed using Multiwfn (4,5). The starting protein structures were solvated with TIP3P water molecules in a cubic box with periodic boundary conditions. The system was neutralized with an ionic concentration of 150 mM. The total number of atoms, box dimensions, and simulation time is reported in Table S1. Long-range electrostatic interactions were treated with the particle mesh Ewald (PME) algorithm (6). A cutoff of 1.0 nm was used for both the real-space Coulombic and Lennard-Jones interactions. Following a steepest descent energy minimization, a short 100 ps simulation was performed with position restraints on all protein heavy atoms in the NVT ensemble at 300 K using the velocity rescaling thermostat (7). This was followed by a 1 ns equilibration in the NPT ensemble at 1 bar without position restraints using a Parrinello-Rahman barostat (8). We used an integration time step of 2 fs. All bonds to hydrogen atoms were constrained using the LINCS algorithm (9).
Production runs of ~900 ns were performed in the NVT ensemble at 300 K. Production simulations were performed on *Expanse*, an NSF-funded system operated by the San Diego Supercomputer Center at UC San Diego, available through the XSEDE program (10). To monitor distances between atoms of interest, we used the PLUMED2 plugin (11).

**Synthesis of LPAT-LII. Reagents and General Procedures.** The synthesis of LPAT-LII was achieved via manual Fmoc solid phase peptide synthesis (SPPS) using Fmoc D-Ala Wang resin (AAPPTec) (Fig. S8A). Unless noted otherwise, all steps (washing, coupling, deprotection) were performed at room temperature and included gentle agitation on a bench-top rocking platform.

Incorporation of D-isoglutamine was achieved using a commercially available building block (Fmoc-D-isoGln-OH) purchased from AAPPTec. The 4-methyltrityl (Mtt) protected lysine residue (Fmoc-Lys(Mtt)-OH) used to create the isopeptide linkage in LPAT-LII was purchased from AAPPTec. Boc-2-aminobenzoic acid for installation of the 2-aminobenzoyl (Abz) fluorophore was obtained from Chem-Impex International. All other materials and reagents were obtained from commercial sources and used without further purification.

**Synthesis.** A 15 mL polypropylene synthesis vessel fitted with appropriate frits and inlet/outlet caps was loaded with 0.286 g (0.2 mmol scale) of Fmoc-D-Ala Wang Resin (S1, 0.7 mmol/g). The resin was then swollen prior to synthesis with ~20 mL of N-methyl-2-pyrrolidinone (NMP) (3x, 10 min per wash). Next, the base-labile Fmoc group was removed with 20 mL of 20% piperidine in NMP (2x, 10 min per treatment), followed by washing with ~20 mL of NMP (3x, 5 min per wash). The resin was then elaborated through sequential coupling of Fmoc-D-Ala-OH,
Fmoc-Lys(Mtt)-OH, Fmoc-D-isoGln-OH, and Fmoc-Ala-OH. For each residue, a coupling solution consisting of Fmoc amino acid (0.6 mmol), O-(benzotriazol-1-yl)-N,N,N’ ,N’-tetramethyluronium (HBTU) (0.6 mmol), and N,N-diisopropylethylamine (DIPEA) (1.0 mmol) in ~6 mL of NMP was used. Following thorough mixing, the coupling solutions were added to the synthesis vessel containing the deprotected resin. If necessary, additional NMP was added to fully suspend the resin. Couplings were incubated for a minimum of 40 minutes at room temperature. Following each coupling, the resin was washed with ~20 mL NMP (3x, 10 min per wash). The resin was then deprotected with ~20 mL of 20% piperidine in NMP (2x, 10-20 min per treatment), and washed with ~10 mL NMP (3x, 5 min per wash). Repeated cycles of coupling and deprotection were then used to assemble the target sequence. Following coupling of the Fmoc-Ala-OH residue and removal of the Fmoc group, the peptide chain was acetylated at its N-terminus via overnight treatment with a capping solution consisting of acetic anhydride (0.94 mL, 10 mmol) and DIPEA (1.74 mL, 10 mmol) in ~6 mL of NMP. This acetylation step yielded resin bound intermediate **S2** (**Fig. S8A**). The 4-methyltrityl (Mtt) protecting group on the lysine side chain was then removed by treatment with 15 mL of a solution of 94:5:1 CH₂Cl₂/TIPS/TFA (3x, 5 min per treatment). The resin was not agitated during this step. The resin was next washed with ~20 mL of NMP (3x, 5 min per wash) to yield resin-bound intermediate **S3** (**Fig. S8A**). At this stage the resin was portioned into two equal portions, and half of this material was extended from the lysine ε-amine to generate the complete sequence of LPAT-LII. The coupling of these additional residues was achieved using methods analogous to those described above. Finally, Boc-2-aminobenzoic acid was coupled to the terminal leucine residue using similar procedures in order to install the 2-aminobenzoyl (Abz) fluorophore. Following completion of the synthesis, the resin was washed with NMP (3x) and CH₂Cl₂ (3x). A 5 mL solution of 95:2.5:2.5
TFA/TIPS/H$_2$O was then used to cleave LPAT-LII from the resin (2x, 30 min per treatment). The resin was not agitated during the cleavage step. The cleaved peptide solution was concentrated on a rotary evaporator, and the remaining residue was added dropwise to 35 mL of diethyl ether chilled over dry ice. The suspension was centrifuged at 4000 rpm for 5 minutes at 4 °C to collect precipitated LPAT-LII peptide. The diethyl ether was decanted and the crude peptide was dried overnight under vacuum.

Crude LPAT-LII was resuspended in a minimum volume of 1:1 MeCN/H$_2$O and purified by RP-HPLC using a Dionex UltiMate 3000 HPLC system equipped with a Phenomenex Luna 5 μm, 100 Å C18 column (10 x 250 mm) [aqueous (95:5 H$_2$O/MeCN, 0.1% formic acid) / MeCN (0.1% formic acid) mobile phase at 4.0 mL/min, method: hold 20% MeCN (0.0-2.0 min), linear gradient of 20-45% MeCN 2.0-7.0 min, linear gradient of 45-55% MeCN 7.0-12.0 min, linear gradient of 55-90% MeCN 12.0-13.0 min, hold 90% MeCN 13.0-14.5 min)]. Pure peptide fractions were concentrated on a rotary evaporator and lyophilized. Purified LPAT-LII peptide was then resuspended in DMSO at a final concentration of 20 mM. The identify and purity of LPAT-LII in this stock solution were confirmed by LC-ESI-MS (Fig S8A) and RP-HPLC (Figure S8B). These analyses were conducted on a Dionex UltiMate 3000 HPLC system interfaced with an Advion CMS expression$^*$ mass spectrometer. Separations were achieved with a Phenomenex Kinetix® 2.6 μM C18 100 Å column (100 x 2.1 mm) [aqueous (95% H$_2$O, 5% MeCN, 0.1% formic acid) / MeCN (0.1% formic acid) mobile phase at 0.3 mL/min, method: hold 10% MeCN (0.0-0.5 min), linear gradient of 10-90% MeCN (0.5-7.0 min), hold 90% MeCN (7.0-8.0 min)].
**Figure S1. Sequences and LC-ESI-MS characterization of wild-type and C208A spySrtA.** Mass spectra were acquired using an Agilent 6545XT AdvanceBio Q-TOF system as described in Materials and Methods. Unprocessed spectra showing full charge ladders for each protein are shown, along with the corresponding deconvoluted spectra.
Figure S2. Model transacylation reactions of spySrtA with LPATA and LPATS substrates. (A) Schemes for the reaction of model peptides (Ac-LPATSG-NH$_2$ or Abz-LPATAGK(Dnp)-NH$_2$) with excess alanine amide (A-NH$_2$) in the presence of spySrtA. Conditions: 50 μM LPATS/LPATA substrate, 5 mM A-NH$_2$, 1 or 5 μM spySrtA, 10% v/v sortase reaction buffer (500 mM Tris, 1500 mM NaCl, pH 7.5), room temperature. (B) ESI-MS spectrum of crude reaction mixture with Ac-LPATSG-NH$_2$ (5 μM spySrtA, 180 min time point). (C) ESI-MS spectrum of crude reaction mixture with Abz-LPATAGK(Dnp)-NH$_2$ (1 μM spySrtA, 260 min time point).
Figure S3. Structural alignments of spySrtA complex structures. (A) Pairwise alignments for all protomers of both structures reveal that the main-chain atoms of spySrtA-LPATS align to protomer A or protomer B of spySrtA-LPATA with RMSD = 0.124 Å (544 atoms) and 0.114 Å (556), respectively. Alignment of protomer B of spySrtA-LPATS with the two protomers of spySrtA-LPATA resulted in RMSD values of: 0.126 Å (569) and 0.124 Å (574), respectively. Within each structure, the protomers align within the experimental error: RMSD = 0.075 Å (558) for spySrtA-LPATA and 0.089 Å (499) for spySrtA-LPATS. All structures are shown in ribbon representation and colored as labeled. The peptides are shown as sticks and colored by heteroatom (O = red, N = blue). (B) Although the apo spySrtA protein previously crystallized (PDB ID 3FN5, gray cartoon) started at residue Ser81, residues N-terminal to Ala89 are unresolved, perhaps due to the presence of an additional 18 residues including a His-tag and Thrombin cleavage site. The spySrtA protein crystallized in our spySrtA-LPATS structure (blue cartoon with the peptide shown as in (A)), starts at Ser80 and all residues are resolved. The N-terminal residues form an additional α-helix, as shown in cartoon, colored cyan, and labeled. (C) Unbiased electron density maps are shown for each of the spySrtA-LPATA and spySrtA-LPATS structures. The peptide atoms were removed and a round of refinement was run using the experimental reflections data. Protomer A of each of the complex structures is shown in gray cartoon and aligned with the refined structure. The modeled peptides are shown as sticks and colored by heteroatom, as labeled. Catalytic triad residues (H142, C208A, and R216) are in gray sticks. In each, the $F_o-F_c$ maps are rendered at 2.5σ and shown in the vicinity of the peptide.
Figure S4. The spySrtA LPATA- and LPATS-complex structures. (A) The Abz moiety at the N-terminus of the Abz-LPATAGK(Dnp)-NH$_2$ peptide is in two conformations in the peptides bound to protomer A (chain C) and B (chain D) in the spySrtA-LPATA structure. The peptides are in stick representation and colored by heteroatom (N=blue, O=red, C=yellow). The spySrtA protein is in gray cartoon, with the interacting residue, P188, and neighboring residues, A187 and E189, as sticks and colored by heteroatom. The distance between the carbonyl of P188 and amide of Abz is labeled and shown as a black dashed line. (B) A comparison of the potential interactions of the P3 Pro with SrtA enzymes, including from the saSrtA-LPAT* (PDB ID 2KID) structure (left) and spySrtA-LPATS structure (right). The peptides are shown as stick residues and colored by heteroatom and as labeled. The spySrtA enzymes are in gray cartoon, with interacting residue side chains as sticks and colored by heteroatom (C=black/cyan, S=golden yellow). Distances are labeled and measurements shown as black dashed lines. (C) The peptides from the saSrtA-LPAT* (2KID), baSrtA-LPAT* (2RUI), and spySrtA-LPATS structures are shown in stick representation and colored by heteroatom (as in (A-B)) to highlight the differences in orientation of the P2 Ala and P1 Thr residues. Orientation of “Ala-in/out” and “Thr-in/out” are indicated by the colored arrows. (D) Energy minimization of the spySrtA-LPAT acyl-enzyme intermediate model solvated in water. (E) Triplicate fluorescence data (in relative fluorescence units, RFU) for the reaction of Abz-LPATGG-K(Dnp), left, or Abz-LPATSG-K(Dnp), right, and H$_2$NOH in the presence of WT (black circles), T207A (dark gray squares), and R216A (gray diamonds) spySrtA protein. The key is the same for both graphs. Data for the LPATA peptide is in Fig. 4D.
Figure S5. Comparison of transacylation reactions with LPAT-LII and LPATA. (A) Model transacylation reactions of Abz-LPAT-lipid II (LPAT-LII) or Abz-LPATAGK(Dnp)-NH$_2$ (LPATA) with excess alanine amide (A-NH$_2$) in the presence of spySrtA. Conditions: 50 μM LPAT-LII/LPATA peptide, 5 mM A-NH$_2$, 1 μM spySrtA, 10% v/v sortase reaction buffer (500 mM Tris, 1500 mM NaCl, pH 7.5), room temperature. Reaction progress was monitored by RP-HPLC, and the identity of all reaction components was confirmed by LC-ESI-MS. Representative RP-HPLC chromatograms for the reactions of (B) LPAT-LII and (C) LPATA are shown. (D) Time course of reactions in panel (A) demonstrating a higher level of substrate conversion for LPAT-LII as compared to LPATA. Percent conversion values were estimated by comparing RP-HPLC peak areas for the unreacted substrates and the specific reaction products labeled in panels (B) and (C). Data points represent three independent experiments (mean ± standard deviation).
Figure S6. Structure of spySrtA bound to LPAT-LII peptide. The structures in these figures are rendered similarly: the spySrtA protein is shown as gray cartoon, with any side chains or highlighted areas as sticks and colored by heteroatom (N=blue, O=red). The LPAT-LII molecules are in green sticks and colored by heteroatom. Molecules related by symmetry are labeled, indicated by single or double apostrophes. Ligand LPAT-LII molecules related by symmetry are in dark green sticks and colored by heteroatom and labeled. Distances are shown as black dashed lines and labeled. (A) The similarities in the LPATA sequence of the spySrtA-LPATA and LPAT-LII bound structures are highlighted. The black arrow indicates the only difference, which is a rotation in the P2 Ala carbonyl. This reflects conformational flexibility seen in the P2 Ala and P1 Thr residues amongst the structures presented in this work. (B-C) These figures illustrate interactions of LPAT-LII with the spySrtA enzyme (B) and molecules, both ligand and protein, related by symmetry (C). In (B), both LII pentapeptides are shown and the residues are labeled. The “Thr-in” peptide (green carbons) is shifted by approximately one residue as compared to the “Thr-out” LII peptide (pale yellow carbons); however, it is the amide of the isoglutamine residue that interacts with the hydroxyl of Y120 in both structures. The “Thr-in” and “Thr-out” structures are labeled in (C).
Figure S7. Molecular dynamics simulations of spySrtA. (A) The average root-mean-square-deviation (RMSD) for backbone atoms over the course of the simulation is graphed for all four simulations, including the apo spySrtA (PDB ID 3FN5, black curve) protein, as well as the structures presented here, spySrtA-LPATA (yellow), spySrtA-LPATS (blue), and spySrtA-LPAT-LII “Thr-in” (green). (B) The average root-mean-square-fluctuation (RMSF) from the average position over the course of each simulation is graphed for the L-P-A-T-A/S residues, colored as in (A) and labeled. These values are an average of the RMSF for the backbone atoms of each residue. (C) The RMSF for each atom of the LPAT-LII molecule is graphed for Abz-LPATAA (left) and pentapeptide of LII (right); Ac = acetyl group. The order of atoms is backbone then side chain for all residues, and the exact order is listed for the LPATAA segment and the Lys residue of the pentapeptide, highlighting the increased flexibility for the Lys atoms starting at the CE atom of the sidechain.
Figure S8. Synthesis and characterization of LPAT-LII. (A) Synthetic scheme for the solid phase synthesis of Abz-LPAT-lipid II (LPAT-LII). The identity of purified LPAT-LII was confirmed by LC-ESI-MS. (B) RP-HPLC characterization of purified LPAT-LII. The presence of a peak in the 320 nm chromatogram is consistent with the expected absorbance of the 2-aminobenzoyl (Abz) fluorophore.
Table S1. Details of the molecular dynamics simulation size.

<table>
<thead>
<tr>
<th>System</th>
<th>Total number of atoms</th>
<th>Cubic box dimensions [nm]</th>
<th>Simulation time [ns]</th>
</tr>
</thead>
<tbody>
<tr>
<td>SpySrtA (Apo)</td>
<td>34123</td>
<td>7.0224</td>
<td>896.6</td>
</tr>
<tr>
<td>SpySrtA-LPATA</td>
<td>34118</td>
<td>7.0186</td>
<td>905.0</td>
</tr>
<tr>
<td>SpySrtA-LPATS</td>
<td>32523</td>
<td>6.9350</td>
<td>905.0</td>
</tr>
<tr>
<td>SpySrtA-LPAT-LII</td>
<td>35836</td>
<td>7.1559</td>
<td>1005.0</td>
</tr>
</tbody>
</table>

PDB ID 3FN5
PDB ID 7S51
PDB ID 7S40
PDB ID 7T8Y

Supplemental References

Figure S9. UV-vis trace of WT SpySrtA. Representative UV-vis 360nm wavelength data from one trial of WT SpySrtA with LPATG (A), LPATL (B), and LPATY (C). Time points were taken every two hours but only the 0th, 10th, and 18th hours are shown for space considerations.
Figure S10. UV-vis trace of WTlmSrtA. Representative UV-vis 360nm wavelength data from one trial of WTlmSrtA with LPATG (A), LPATL (B), and LPATY (C). Time points were taken every two hours but only the 0th, 10th, and 18th hours are shown for space considerations.
Figure S11. UV-vis trace of WTSagSrtA. Representative UV-vis 360nm wavelength data from one trial of WTSagSrtA with LPATG (A), LPATL (B), and LPATY (C). Time points were taken every two hours but only the 0th, 10th, and 18th hours are shown for space considerations.
**Figure S12. UV-vis trace of H143A SpySrtA.** Representative UV-vis 360nm wavelength data from one trial of H143A SpySrtA with LPATG (A), LPATL (B), and LPATY (C). Time points were taken every two hours but only the 0th, 10th, and 18th hours are shown for space considerations.
**Figure S13. UV-vis trace of I211P SpySrtA.** Representative UV-vis 360nm wavelength data from one trial of I211P SpySrtA with LPATG (A), LPATL (B), and LPATY (C). Time points were taken every two hours but only the 0th, 10th, and 18th hours are shown for space considerations.
Figure S14. UV-vis trace of mono(spy)7/8. Representative UV-vis 360nm wavelength data from one trial of mono(spy)7/8 with LPATG (A), LPATL (B), and LPATY (C). Time points were taken every two hours but only the 0th, 10th, and 18th hours are shown for space considerations.
Figure S15. UV-vis trace of spy(mono)7/8. Representative UV-vis 360nm wavelength data from one trial of Spy(mono)7/8 with LPATG (A), LPATL (B), and LPATY (C). Time points were taken every two hours but only the 0th, 10th, and 18th hours are shown for space considerations.
Table S3. Details of the molecular dynamics simulation size.

<table>
<thead>
<tr>
<th>System</th>
<th>Total number of atoms</th>
<th>Cubic box dimensions [nm]</th>
<th>Simulation time [ns]</th>
</tr>
</thead>
<tbody>
<tr>
<td>SpySrtA-LPATLA(^a)</td>
<td>34121</td>
<td>7.018</td>
<td>1000</td>
</tr>
</tbody>
</table>

\(^a\)PDB ID 3FN5

References


Sortase Sequences

>`spySrtA
MESSHHHHHHENLYFQSSVLQAQMAAQQLPVGIGIAIPELGINLPFGBKGLNTLIELYGAGTMKEEQVMMGGNNYSLAHIFGITGSSQMLFSPLERAQNGMSIYLTDKEKIYEYIIKDVFTVAPERVDVIDDTAGLKEVTLVTCTDIEATERIIVKGELKTEYDFDKAPADVLFKAFNHSYNQVST

>`SrtA_List_Mono
MESSHHHHHHGAANYDKDAVVGSIAVPSVDVNLLVFKGNTANLLAATTMRSDQVMGKGNYPLAGHHMRDESMLFGPIMKVKKGDYLTDLLENLYETVTETKTDKAPADVLFKAFNHSYNQVST

>`SrtA_S_agalactiae_79-247
MESSHHHHHHENLYFQSSILSAQTKSHNLPVIGGIAIPDVEINLPFGBKGLNTLIELYGAGTMKENQIMGGPNNYALASHHFGLTGSSKMLFSPLHAHKGMKVYLTDKSVYYTITTEISKVTPEHVEVIDDTPGKSLTLVTCTDPEATERIIVAHELEKTGEFSTADELSKAFSKKYNQINL

>`lmSrtA_pyogenes 7/8
MESSHHHHHHENLYFQSGAANYDKDAVVGSIAVPSVDVNLLVFKGNTANLLAATTMRSDQVMGKGNYPLAGHHMRDESMLFGPIMKVKKGDYLTDLLENLYETVTETKTDKAPADVLFKAFNHSYNQVST

>`spySrtA_b7-b8_mono
MESSHHHHHHENLYFQSSVLQAQMAAQQLPVGIGIAIPELGINLPFGBKGLNTLIELYGAGTMKEEQVMMGGNNYSLAHIFGITGSSQMLFSPLERAQNGMSIYLTDKEKIYEYIIKDVFTVAPERVDVIDDTAGLKEVTLVTCTDIEATERIIVKGELKTEYDFDKAPADVLFKAFNHSYNQVST

>`H143A spySrtA
MESSHHHHHHENLYFQSSVLQAQMAAQQLPVGIGIAIPELGINLPFGBKGLNTLIELYGAGTMKEEQVMMGGNNYSLAHIFGITGSSQMLFSPLERAQNGMSIYLTDKEKIYEYIIKDVFTVAPERVDVIDDTAGLKEVTLVTCTDIEATERIIVKGELKTEYDFDKAPADVLFKAFNHSYNQVST

85
>I211P_spySrtA
MESSHHHHHHENLYFQSSVLQAQMAAQQLPVIGGIAIPELGINLPFKGLGNTELIYAG
TMKEEQVMGGENNYSLASHHIFGITGSSQMLFPLERAQNGMSIYLDKEKIYEYIIKDV
FTVAPERVDVIDDTAGLKEVTTLVCTDPÆATERIIVKGELKTEYDFD KAPADV LGKAFNHS
YNQVST