Spring 2023

Activity and Selectivity of Class B Sortase Enzymes

Sophie Jackson
Western Washington University, sophiejackson286@gmail.com

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

Part of the Chemistry Commons

Activity and Selectivity of Class B Sortase Enzymes

By

Sophie N. Jackson

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

ADVISORY COMMITTEE

Chair, Dr. Jeanine Amacher

Dr. John Antos

Dr. Serge Smirnov

GRADUATE SCHOOL

David L. Patrick, Dean

Master's Thesis

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.

Signature: *Sophie N. Jackson*

Date: May 12, 2023

Activity and Selectivity of Class B Sortase Enzymes

A Thesis Presented
to the Faculty of
Western Washington University

In Partial Completion
of the Requirements for the Degree
Master of Science

By
Sophie N. Jackson
May 2023

Abstract

Gram-positive bacteria attach many proteins to their cell walls via sortase enzymes. Sortases are cysteine transpeptidases and are grouped into 6 classes, A-F. Sortase enzymes, particularly sortase A from *Staphylococcus aureus,* have been used extensively for *in vitro* protein ligations. Here, we investigate substrate-binding in sortase A from *Streptococcus pyogenes*. In addition, class B sortases are typically overlooked for research and development due to low *in vitro* activity and incomplete knowledge of substrate specificity. Here, we investigate the activity of class B sortases from *Bacillus anthracis* (baSrtB), *Clostridioides difficile* (cdSrtB), *Listeria monocytogenes* (lmSrtB), and *Staphylococcus aureus* (saSrtB). Of these, baSrtB was the most active in our hands and was selected for further study. Mutant enzymes were created to study the impact of a class B N-terminal α-helix and a structurally conserved, but sequentially variable, loop on baSrtB activity. Mutations to the structurally conserved loop were impactful on enzyme activity, with some mutations decreasing activity while others greatly increased it. A substrate-bound enzyme model generated using Alphafold2 (Galaxy) allowed us to explore enzyme-substrate interactions in greater detail. This model was validated through molecular dynamics simulations and mutagenesis. This work shows that baSrtB is a viable tool for protein engineering studies and lends greater insight into the structural features that underpin sortase activity and selectivity.

Acknowledgements

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abz	2-aminobenzoyl
baSrtB	*Bacillus anthracis* sortase B
cdSrtB	*Clostridioides difficile* sortase B
cpSrtB	*Clostridium perfringens* sortase B
CWSS	cell wall sorting signal
GlcNAc	b-1,4,*N*-acetylglucosamine
HPLC	High Performance Liquid Chromatography
K(Dnp)	2,4-dinitrophenyl lysine
LII	Lipid II
LCMS	Liquid Chromatography Mass Spectrometry
lmSrtB	*Lysteria monocytogenes* sortase B
MurNAc	N-acetylmuramic acid
NOE	nuclear Overhauser effects
RMSD	root-mean-square fluctuation
saSrtA	*Staphylococcus aureus* sortase A
saSrtB	*Staphylococcus aureus* sortase B
SML	Sortase-mediated ligation
spSrtA	*Streptococcus pneumoniae* sortase A
spySrtA	*Streptococcus pyogenes* sortase A

Chapter 1: Introduction

Surface proteins enable bacteria to perform functions that are critical for survival.[1] In pathogens, these proteins are frequently virulence factors.[1] In all Gram-positive bacteria and some Gram-negative bacteria and archaea, many surface proteins are anchored via sortase enzymes.[1–3] A representative sortase structure is shown in Figure 1. Sortases have attracted a great deal of attention because of their role in pathogenicity.[1,4] In the decades since they were first discovered, sortases have also become powerful protein engineering tools.[1,4] In recent years, bacterial genome sequencing projects have greatly expanded the catalog of potential sortase genetic sequences.[3] To date, more than 10,000 such sequences have been identified from over 1,000 bacterial species.[1,3] Many of these sortases have not been explored experimentally and have unknown functions.[3] The sortase superfamily is organized into 6 classes, A-F, based on primary sequence.[1,5] Sortase research has largely focused on the class A sortases.[1,3] Of these, a single class A sortase from *Staphylococcus aureus* (saSrtA) remains the most thoroughly studied and the most utilized for protein engineering.[1,6,7] Much remains to be learned about sortases, particularly class B-F sortases. Class B sortases, the primary focus of this work, are associated with pathogen persistence and have unique structural features and substrate recognition sequences, as well as generally poor *in vitro* reaction rates.[1,5,8] Studying sortase B enzymes will provide insight into how these and other sortases function, with implications for human health and protein engineering.

Figure 1. Class B sortase from *Bacillus anthracis*. Example structure of a sortase enzyme (PDB ID: 1RZ2).

1.1: *In Vivo* Sortase Function

Sortases are cysteine transpeptidases that operate at the surface of the cell membrane.[1,5,9] Sortase enzymes extend toward the cell wall matrix and are anchored within the cell membrane by a hydrophobic N-terminus.[4] Substrate proteins are synthesized in
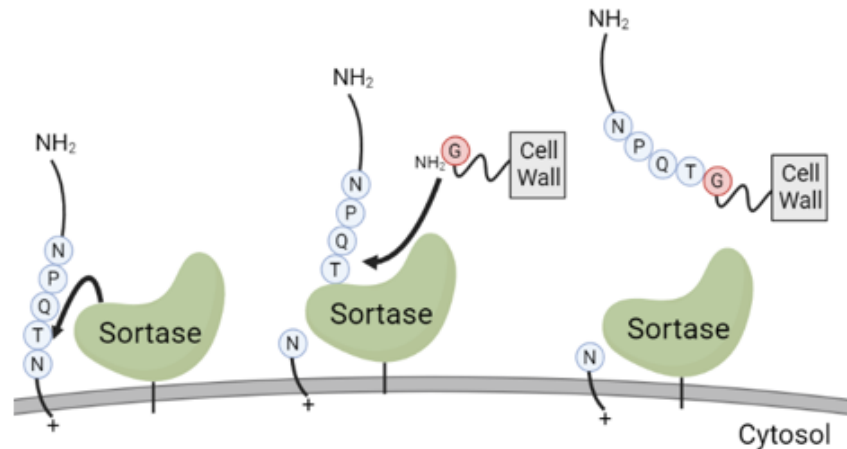
Figure 2. Sortase enzymes perform transpeptidation reactions at the cell surface. Sortase enzymes recognize short amino acid sequences and sever substrates from C-terminal transmembrane anchors before ligating substrates to a surface amine group, such as a component of the cell wall.

the cytoplasm and transported to the cell envelope via the secretory pathway.[4,5] During this process, they become rooted in the cell membrane by a C-terminal region that consists of a sequence of hydrophobic (transmembrane) residues and ends with a positively charged (cytosolic) tail.[1,4,10] Connecting the substrate's C-terminal transmembrane region to the cell wall-facing N-terminus is the cell wall sorting signal, an amino acid sequence that forms the basis of sortase substrate recognition.[1,4,5] The cell wall sorting signal, also referred to as a motif, is recognized and cleaved by the sortase enzyme.[1,4,5] The C-terminus is left rooted in the cell membrane, while the N-terminal protein is ligated to an amine nucleophile.[1,4,5] Most commonly, this nucleophile is a component of the cell wall, although in the case of pilus polymerization, the nucleophile is a lysine ε-amino group from another pilin subunit.[1,4,5] The *in vivo* sortase reaction process is summarized in Figure 2 with a representative sorting signal motif and nucleophile.[11]

There is considerable diversity in sortase substrates.[1] Some sortases anchor many proteins to the cell wall, while others anchor only one.[5,12] In one extraordinary example, sortase A from *Listeria monocytogenes* has at least 43 distinct protein substrates.[3] Substrate proteins contain a sorting signal motif that enables sortase recognition.[1,4] The motif is canonically 5 amino

acid residues in length.[1,4] The various residues of the motif can be designated by proximity to the cleavage site. Using the NPQTN motif as an example, P4 = N, P3 = P, P2 = Q, P1 = T, and P1′ = N. The amide bond joining P1 and P1′ is cleaved during catalysis, and P1 is subsequently ligated to the surface amine group. Sortases are specific to a particular motif and different sortase enzymes often recognize different motifs.[1,4,5,12] In fact, many bacteria express more than one sortase enzyme and in some cases these sortases serve non-redundant roles and have mutually exclusive recognition sequences.[1,5] For example, the class A sortase from *Staphylococcus aureus* (saSrtA) recognizes the LPXTG motif, where X is any amino acid, while sortase B from this same species (saSrtB) recognizes NPQTN.[12] Several proteins contain the LPXTG motif and are recognized by saSrtA, while only one protein is recognized by saSrtB.[2,12] In general, the different sortase classes recognize distinct sorting signal motifs, although there is variability within classes and knowledge of the motifs recognized by different sortases is still growing.[1] General sortase class motif patterns are shown in Table 1.[1,5,13,14] There is also diversity of the amine nucleophile that serves as the ligation partner and is the second sortase substrate. Depending on the sortase, this nucleophile may be an N-terminal pentaglycine strand, dialanine, or *meso*-diaminopimelic acid, among other possibilities.[9,10,15]

Sortase Class	Recognition Motif				
A	L	P	X	T	G
B	N	P	Q/K	T	N/G
C	L/I	P	X	T	A
D	L	P	X	T	A
E	L	A	X	T	G
F	*Unknown*				

Table 1. Common recognition motifs by sortase class. Sortases tend to share motif patterns with other members of their class. X denotes any amino acid residue.

1.2: Major Structural Features

Sortases share certain structural features. A cartoon representation of saSrtA, considered the archetypal sortase structure, is shown in Figure 3.[1] All characterized sortases share an 8-stranded β-barrel core.[3,16] This feature is conserved even in sortases with markedly different

primary sequences.[16] Flexible loops connect one β-strand to the next and are designated accordingly (e.g. the loop connecting the β6 sheet to the β7 sheet is the β6-β7 loop).[1] These loops are relatively structurally diverse.[16] In saSrtA, the major catalytic residues (H120, C184, and R197, respectively) are positioned in close proximity to one another within the β-barrel fold.[1] This region constitutes the active site. A substrate-binding groove is observed at the active site in saSrtA and is common among sortases.[1] The binding groove floor is formed by residues in the β4 and β7 strands and the walls are formed by the β2-β3, β3-β4, β6-β7, and β7-β8 loops.[1,17]

Beyond these broadly conserved structural features, there are several regions of class-associated structural variation.[1] These include the N-terminal region before the β-barrel, the β6-β7 loop, the β7-β8 loop, and the

Figure 3. SaSrtA is the archetypal sortase structure. SaSrtA (PDB ID: 1T2P) displays several of the common features that unite sortase enzymes. These include the characteristic β-barrel fold (shown in blue) and the "catalytic triad": highly conserved histidine, cysteine, and arginine residues (labeled, side chains shown as sticks).

C-terminus extending away from the β-barrel.[1] Several of these features in class B sortases differ notably from the class A sortases (Figure 4).[1,3] The class B sortase β6-β7 loop is markedly longer than that of other sortases and includes an elongated α-helix.[1,2] Additionally, compared to saSrtB, the B class sortases have additional N-terminal helices.[1] The role of these N-terminal helices is currently unknown.[1] Overall, sortase B enzyme structures appear to be largely similar across different homologs, including those that perform pilus polymerization and protein anchoring.[1]

Figure 4. Class B sortases display structural differences compared to class A. Compared to class A sortases (**A**, saSrtA PDB ID 1T2P), class B sortases (**B**, saSrtB PDB ID 4LFD) have additional N-terminal α-helices and long β6-β7 loops with an elongated α-helix (dark blue). Catalytic histidine, cysteine, and arginine residue side chains shown in stick representation (pink).

1.3: Sortase-Mediated Ligation

Sortase *in vitro* activity was first observed in 1999 using saSrtA.[1] Since then, sortases, primarily saSrtA, have been used extensively to perform protein ligations *in vitro*, termed sortase-mediated ligation (SML).[1,18] A general SML scheme is shown in Figure 5. An astonishing variety of proteins have been created using SML, including drug-antibody conjugates, cyclized proteins that resist degradation, fluorescently-labeled protein tags, and biotinylated proteins, among many other applications.[7,18,19] SML has several advantages over other protein ligation strategies. First, the sortase substrate motif is a relatively short sequence of naturally occurring amino acids. For saSrtA, the nucleophile substrate includes another naturally occurring amino acid residue (glycine).[19] This means that the substrate motif and nucleophile can be introduced into target proteins through recombinant genetic expression if they do not occur naturally.[19] This is an advantage over protein ligation techniques that require unnatural reactive groups that must be chemically installed on the target protein prior to ligation.[19] This also means that the end result of

SML is typically the formation of a new native peptide bond.[18] Additionally, SML reactions can be performed under mild conditions with relatively low concentrations of substrates.[18,19] Sortase reactions are also specific enough that SML can be performed in complex matrices, including living cells and cell lysate.[7,18,19]

However, there are several notable drawbacks to SML. Sortases have a generally slow *in vitro* reaction rate.[1] There are likely many factors contributing to this. Perhaps most notably, only a small fraction of sortases are estimated to be in the active state at physiological pH.[1,2,18] For saSrtA specifically, that number is estimated to be around 1%.[1] This is even more remarkable given that the *in vitro* activity of saSrtA is at least 20 times higher than that of other characterized sortases.[1,3,7] Additionally, SML reactions are typically reversible.[7,18] This is primarily driven by the leaving group, which can often serve as the nucleophile and form a new peptide bond that re-creates the starting substrate.[7] Undesirable hydrolysis products can also form when water serves as the nucleophile.[19] Another drawback can be the saSrtA stringent specificity for the LPXTG motif.[7] This specificity is advantageous in many applications, but it also limits the pool of suitable SML substrates.[7] Although mutations can be introduced to substrates via recombinant genetic expression, in some cases this may be impractical or undesirable.[20]
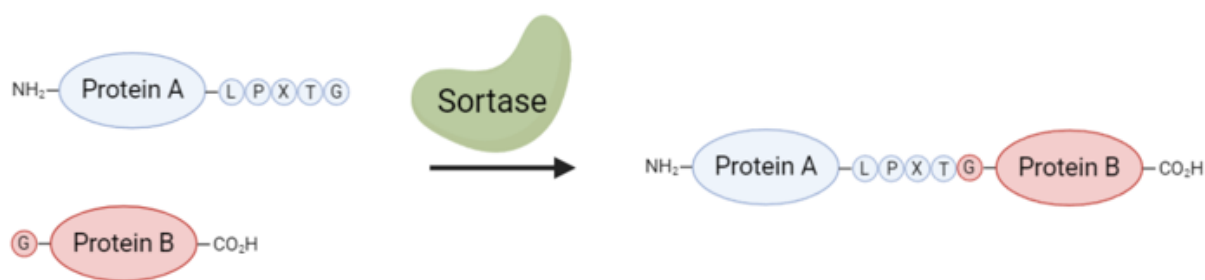
Figure 5. Sortase-Mediated Ligation (SML). Sortase enzymes can be used to perform highly specific ligations *in vitro*. A protein target with a C-terminal recognition motif can be joined to a second protein that displays an appropriate N-terminal nucleophile (typically glycine). The ligations performed typically result in a native amide bond.

1.4: SML Optimization

Much research has been devoted to optimizing SML, including work improving or altering sortases themselves.[7] Several modifications have been made to saSrtA to improve *in vitro* activity.[7] Five point mutations resulting from a directed evolution experiment increased saSrtA catalytic efficiency by 120-fold.[7] This increased activity primarily resulted from improved substrate binding.[7] However, in some instances the resulting pentamutant produces lower yields and more undesirable side products than the wild-type enzyme.[7] Still, these experiments show that sortase activity can be increased through mutations. Much work has also been done to expand the scope of suitable SML substrates. SaSrtA mutants have been generated that display alternative motif selectivity.[7] Generally, the resulting mutants show reduced activity and have not been widely adopted for SML.[6,15] Sortases other than saSrtA also represent promising avenues for development.[6,7,19] SpySrtA is currently the only sortase other than saSrtA to see widespread use.[6] SpySrtA can accommodate an LPXTA recognition motif as well as an N-terminal alanine nucleophile, whereas saSrtA can only effectively accommodate the LPXTG motif and an N-terminal glycine nucleophile.[7] This orthogonal selectivity has been utilized for dual labeling applications, demonstrating the value of developing sortases with alternative selectivity profiles.[7,18] Additionally, a screen of 8 class A sortases from different bacterial species showed that several of these enzymes displayed altered or relaxed selectivity.[6] At least one of these enzymes was successfully used to catalyze a transpeptidation reaction that would not have been possible using saSrtA.[6] Notably, the class B sortases appear to have markedly different recognition motifs than class A.[5] This makes them intriguing candidates for research expanding the catalog of suitable SML substrates.

1.5: *In vitro* Activity of Sortase B

Few class B sortases have been tested for *in vitro* activity.[1] There remains much to be learned about this enzyme class, including the function of the unique sortase B N-terminal α-helix

and the mechanics of substrate recognition.[1] Class B sortases recognize a unique sorting signal motif, typically including an unusual P4 asparagine (e.g. <u>N</u>PQTN), and appear to show more variability in substrate motif than class A sortases.[1,5] These features make sortase B enzymes appealing for SML, however they have not been adopted for this use in large part because they are less active *in vitro* than many class A sortases.[1,8,21] As noted above, poor *in vitro* activity is common in sortase enzymes and may be improved through strategic mutations. Investigating the biochemical and structural features affecting sortase B activity may therefore lead to discoveries that have implications for SML optimization. These enzymes also represent new opportunities to study factors driving sortase activity and selectivity, which may be generalizable to the other sortase classes.

1.6: Research Aims

The principal research aims are to:

1. Identify promising sortase B homologs for study;
2. Characterize the interactions that drive sortase activity by:
   a. Generating crystals structures and models of substrate-bound sortases;
   b. Characterizing the impact of single residues and major structural features, including the class B sortase N-terminal α-helix and the β7-β8 loop, on sortase activity; and
   c. Investigating the relationship between sortase activity and substrate sequence and nucleophile identity; and
3. Identify methods to improve class B sortase *in vitro* activity through substrate and enzyme (mutational) modifications.

Chapter 2: 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

[‡]These authors contributed equally to this work.

Contributions by Thesis Author

I expressed and purified the spySrtA mutants T207A and R216A. I also performed HPLC assays and LC-MS characterization of spySrtA-catalyzed reactions using LPATS, LPATA, and LPAT-LII peptide substrates. Additionally, I performed the fluorescence assay for sortase activity comparing wild-type spySrtA versus the T207A and R216A mutants. I also performed LC-MS/MS characterization of purified T207A and R216A spySrtA proteins. I am a co-first author on this manuscript.

2.1: 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.[23–25] This type of characterization is essential to understanding how sortases perform their role of attaching protein factors to the bacterial cell wall.[5] 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.[7,26]

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).[5] 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.[1,20,27] 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.[5] 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.[20] 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.[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).[20] These structures have revealed that sortases share a conserved core antiparallel 8-stranded b-barrel structure, termed the *sortase fold*.[1,29] This was first identified in the *Staphylococcus aureus* Class A sortase (saSrtA) structure and is consistently found in wild-type and chimeric sortase enzymes.[9,20,29] 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 *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 *in vitro*.[6,17,20,30] 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.[31]

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 *Streptococcus pyogenes* sortase A (spySrtA) (Figure 6). The apo structure of spySrtA was solved using X-ray crystallography in 2009, and its catalytic triad consists of His142, Cys208,and Arg216.[16] 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*.[15,16,21,32–35] 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").[1]

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
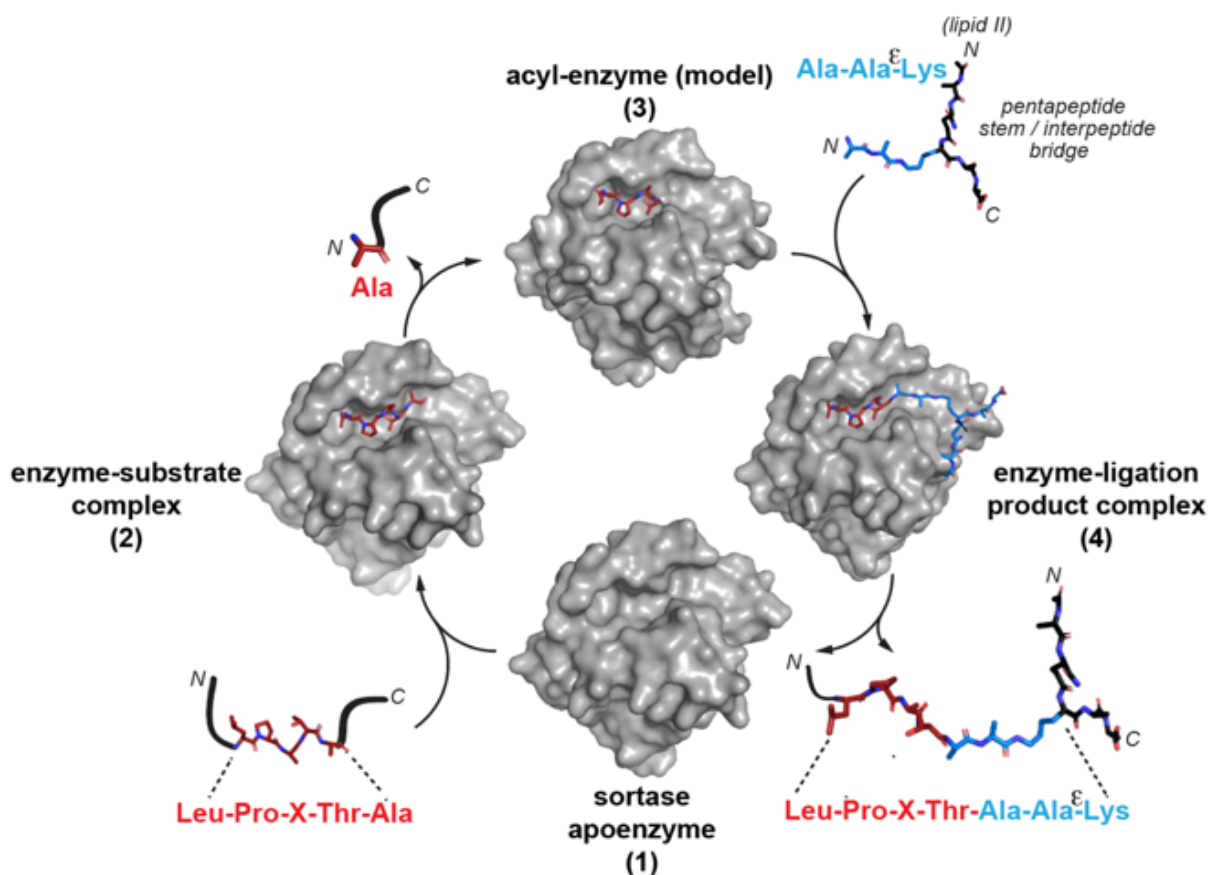
Figure 6. Structural model of the spySrtA catalytic mechanism. A summary of the spySrtA catalytic mechanism, as supported by biochemical and structural data in the field, including studies presented here. A portion of the lipid II ligation partner from *S. pyogenes* is shown, and the structure of this 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.

13

provides new structural insights for important states in the SrtA catalytic mechanism (Figure 6), significantly increasing our understanding of target recognition in this important protein family.

2.2: Results

2.2.1: 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.[36,37] 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*.[33,35,37,38] 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*.[15,16,21,32–34,39] 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.[6,39,40]

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 (Figure A1).[35] 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.[16] From this, we succeeded in crystallizing and solving two structures of C208A spySrtA bound to the model peptides (P1' position in **bold**) *Abz*-LPAT**A**GK(Dnp)-*NH₂* and *Ac*-LPAT**S**G-*NH₂* (Figure 7A). 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*).[6,20,35,41–43] For both substrates, LC-ESI-MS was used to confirm that they were cleaved by wild type spySrtA in a model transacylation reaction (Figure A2). 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.

For simplicity, we will hereafter refer to the solved enzyme-substrate complexes as spySrtA-LPATA and spySrtA-LPATS (Figure 7B). All diffraction and refinement statistics for these complexes are in Table 2. In general, crystals grew stacked and were relatively unstable in traditional cryo solutions (e.g., with 10-20% (*w/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_{work}/R_{free} = 0.21/0.24 at 1.4 Å resolution (Table 2). Relatively high R-factors are a consequence of pseudo-symmetry in crystal packing.[44] 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_{work}/R_{free} = 0.17/0.19, with two spySrtA molecules in the asymmetric unit (Table 2). The unit cell and space group are very similar between spySrtA-LPATS and apo spySrtA.[16]
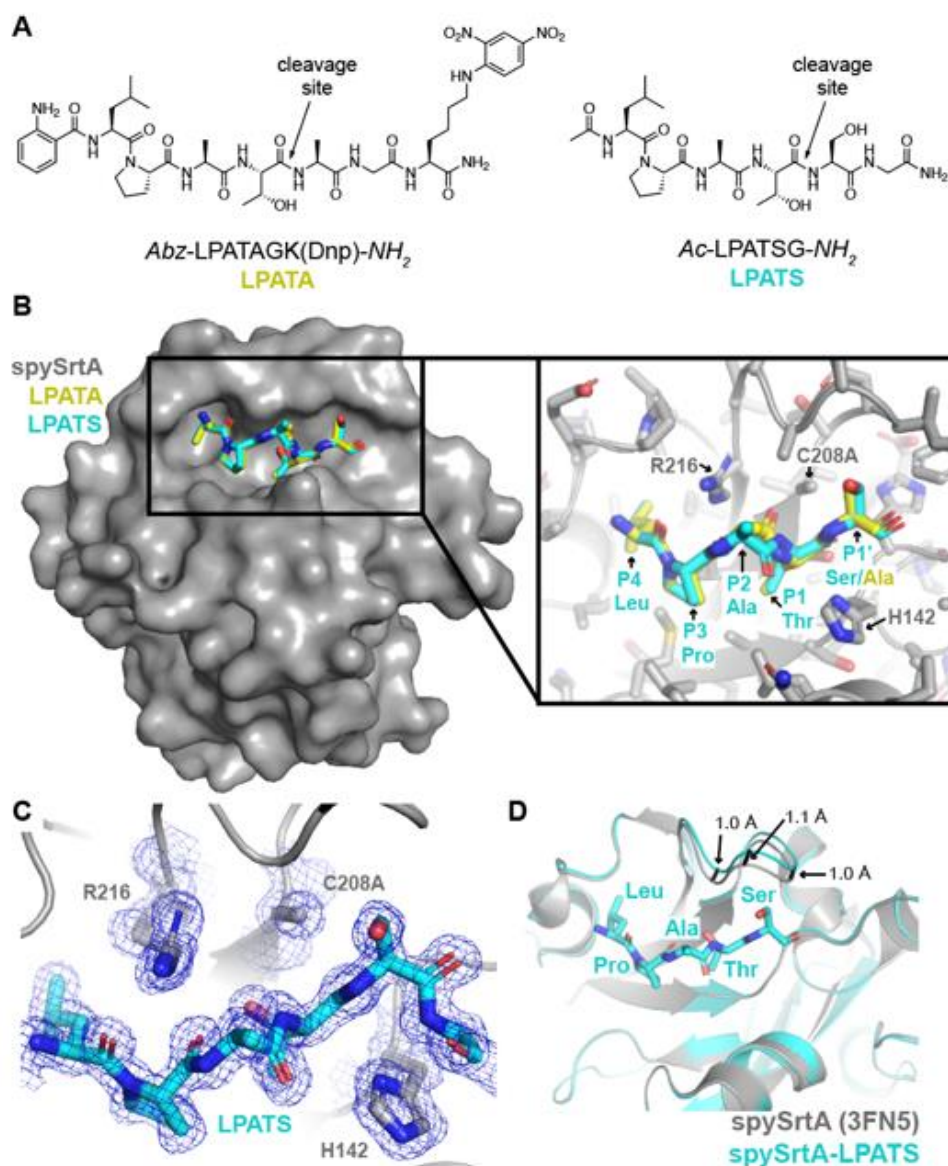
Figure 7. The spySrtA complex structures with LPATS and LPATA peptides. (**A**) Structures of model peptides co-crystallized with spySrtA. **(B)** The spySrtA protein is in gray surface representation, and the non-covalently bound LPATS and LPATA peptides are in sticks and colored by heteroatom (N=blue, 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 $2F_o$-F_c 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 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.

Table 2. Data Collection and Refinement Statistics.

	spySrtA-LPATA	spySrtA-LPATS	spySrtA-Lipid II "Thr-in"	spySrtA-Lipid II "Thr-out"
Data Collection				
Space Group	$P\,2_1\,2_1\,2_1$ (19)	$P\,2_1$ (4)	$P\,2_1\,2_1\,2_1$ (19)	$P\,2_1\,2_1\,2_1$ (19)
Unit cell dimensions				
a,b,c (Å)	58.98,64.57,75.02	38.49,59.1, 64.57	34.3,57.73,72.25	34.32,57.68,71.46
α,β,γ (°)	90, 90, 90	90, 101.7, 90	90, 90, 90	90, 90, 90
Resolution[a] (Å)	48.9-1.4 (1.5-1.4)	43.1-1.4 (1.5-1.4)	45.1-1.8 (1.91-1.8)	44.9-1.9 (2.02-1.9)
R_{sym}[b] (%)	8.1 (40.8)	6.1 (31.5)	7.4 (53.0)	8.3 (54.7)
$I/\sigma I$[c]	12.60 (4.34)	16.86 (4.56)	15.55 (2.47)	16.29 (3.49)
Completeness (%)	99.9 (99.6)	99.6 (98.5)	98.9 (94.0)	98.7 (97.4)
Refinement				
Total # of reflections	56,989	55,864	13,739	11,555
Reflections in the test set	2,838	2,709	684	534
R_{work}[d]/R_{free}[e]	21.3/23.8	16.9/19.4	17.5/20.6	18.4/22.9
Number of atoms:				
Protein	2684	2683	1303	1311
Water	257	449	129	71
Ramachandran plot[f] (%)	99.12/0.88/0	99.42/0.58/0	100/0/0	98.73/1.27/0
B_{av} (Å²)				
Protein	14.4	14.6	24.3	27.4
Bond length RMSD (Å)	0.007	0.006	0.006	0.007
Bond angle RMSD (°)	0.988	0.872	1.135	1.338
PDB Code	**7S51**	**7S40**	**7T8Y**	**7T8Z**

[a]Values in parentheses are for data in the highest-resolution shell.

[b]$R_{sym} = \Sigma_h\Sigma_i\,|I(h) - I_i(h)| / \Sigma_h\Sigma_i\,I_i(h)$, where $I_i(h)$ and $I(h)$ values are the i-th and mean measurements of the intensity of reflection h.

[c]SigAno = $|F(+) - F(-)| / \sigma$

[d]$R_{work} = \Sigma\,||F_{obs}|_h - |F_{calc}||_h / \Sigma\,|F_{obs}|_h$, $h \in \{$working set$\}$

[e]R_{free} is calculated as R_{work} for the reflections $h \in \{$test set$\}$

[f]Favored/allowed/outliers

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 (Figure A3A). 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 (Figure A3B). Because of the large degree of similarity between these structures, unless otherwise noted, our analyses will focus on spySrtA-LPATS protomer A.

2.2.2: 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 (Figure 7B,C). Unbiased electron density maps, created by omitting the peptide atoms and running a round of refinement, confirm strong electron density for peptide residues (Figure A3C). 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 (Figure 7D). Here, superscript numbering refers to the residue position with respect to the catalytic C208 residue, as previously defined.[20] 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 (Figure A4A). 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 (Figure 8A). A similar pocket was previously identified in the NMR structure of saSrtA with a covalent peptidomimetic (LPAT*), PDB ID 2KID (14). The proline residue in P3 interacts weakly via van der Waals interactions with
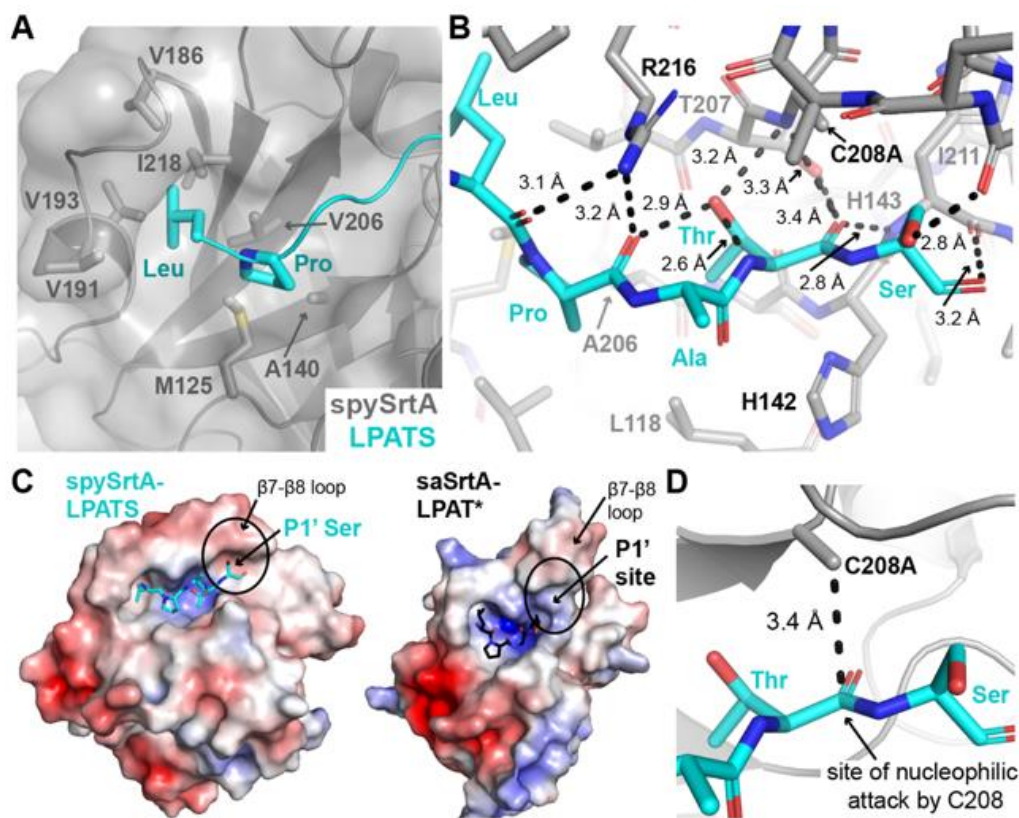
Figure 8. Stereochemistry of the spySrtA-LPATS interaction. (A) The interactions of the P4 Leu and P3 Pro ligand residues with spySrtA are highlighted. The spySrtA enzyme is in cartoon and surface representation, with residues that interact directly with the Leu-Pro shown as sticks and colored by heteroatom (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 (as in **A-B**). 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.

V206 and A140, residues in the β4 and β7 strands, as well as M125 in the β3-β4 loop (Figure

8A). The distances between these residues are of equal magnitude or shorter to those seen in

19

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 (Figure A4B).[17]

There are several backbone atoms in the LPATX motif that form non-covalent interactions with residues in spySrtA (Figure 8B). 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 (Figure 8B). 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 (fiugre A4C).[17,30] 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.[1] 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 (Figure 8B). 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 (Figure 8B).

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 (Figure 8C). 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) (Figure 8C). In our spySrtA-LPATS structure, we also observe a hydrogen bond with the hydroxyl group of the P1' Ser and the carbonyl of I211 (Figure 8B). 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.[6,20] 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.[20,35]

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) (Figure 8D). 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.[1] 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.[1,4,5,45]

2.2.3: Model of the Acyl-Enzyme Intermediate

Next, we used our spySrtA-LPATS complex structure to model the acyl-enzyme intermediate (Figure 9A). 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 *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 (Figure A4D). 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 (Figure 9B). 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
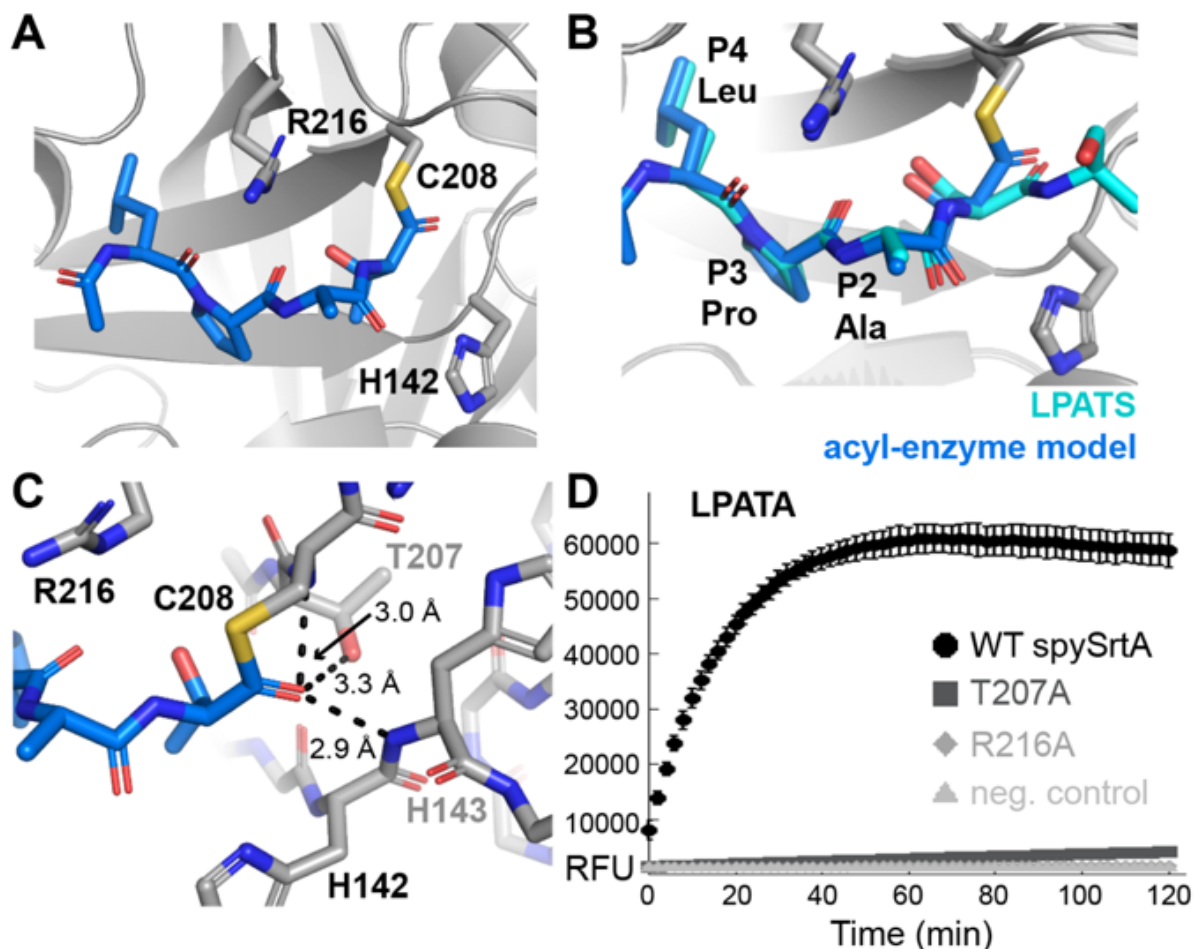
Figure 9. A model of the acyl-enzyme intermediate of spySrtA-LPAT. **(A-C)** The spySrtA protein for the acyl-enzyme intermediate model is in gray cartoon **(A-B)** or stick representation **(C)**, with catalytic residue side chains (H142, C208, and R216) and other relevant positions shown as sticks, colored by heteroatom (C=blue/cyan, O=red, N=blue, S=golden yellow), and labeled. The spySrtA protein from the spySrtA-LPATS structure **(B)** is rendered similarly, but is darker gray. The peptides are in stick representation, with carbons colored as labeled. **(C)** Predicted interactions between the P1 Thr carbonyl and spySrtA T207 side chain, C208A amide, and H143 amide are shown as dashed black lines, with distances labeled. **(D)** Triplicate fluorescence data (in relative fluorescence units, RFU) for the reaction of *Abz*-LPATAGK(Dnp)-*NH$_2$* and H$_2$NOH in the presence of WT (black circles), T207A (dark gray squares), and R216A (gray diamonds) spySrtA protein, and a no protein control (light gray triangles). Data for the *Abz*-LPATGGK(Dnp)-*NH$_2$* and *Abz*-LPATSGK(Dnp)-*NH$_2$* peptides is in figure A4E.

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 (Figure 9B).

Specifically, in the geometry of the acyl-enzyme model, the T207 hydroxyl is 3.3 Å from the P1

Thr carbonyl (Figure 9C). This distance is 3.4 Å in the solved structure of spySrtA-LPATS (Figure 8B).

      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.[1,2,17] 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 (Figure 9C). 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 (Figure 9D, Figure A4E).

      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 (Figure 9C). This type of role for the Thr immediately preceding the catalytic Cys has indeed been suggested in previous computational studies.[46] 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 (Figure 9D, Figure A4E). Notably, a dramatic drop in enzyme activity has also been reported when mutating the corresponding Thr residue (T183) of saSrtA.[47]

2.2.4: Structure and Biochemical Analyses of SpySrtA Bound to a Lipid II Mimetic

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 (38). Although there are exceptions and possible modifications, the main glycan moiety of the peptidoglycan layer consists of alternating b-1,4-*N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) residues that are further crosslinked by peptide subunits.[48,49]

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.[49] 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.[48–50] 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$_5$, *Enterococcus faecalis* = Ala-Ala, *Streptococci* = Ala/Ser-Ala.[48,49]

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* (Figure 10A). 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)-*NH₂* 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.[48,51] 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.*[16,48,52]

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 (Figures 10B, A5). Indeed, we found LPAT-LII to react more rapidly than the related *Abz*-LPATAGK(Dnp)-*NH₂* peptide, suggesting that the added interpeptide bridge/pentapeptide stem portion may be enhancing binding and recognition by spySrtA (Figure A5)*.*

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 (Figure 10C-D).[17,30] 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 $R_{work}/R_{free} = 0.18/0.21$, and the spySrtA-LPAT-LII "Thr-out" structure to a final $R_{work}/R_{free} = 0.18/0.23$ (Table 2). 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 (Figure A6A). 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 e-amine/e-carbon of the L-lysine residue are also well conserved between the "Thr-in" and "Thr-out" structures (gray arrow in Figure 10D). 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) (Figure 10E). Moving beyond the e-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 (Figure 10C-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 (Figure A6B). In each, there are also multiple interactions with the lipid II pentapeptide and spySrtA enzyme of molecules related by symmetry (Figure A6C).

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 (Figure 10C). 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
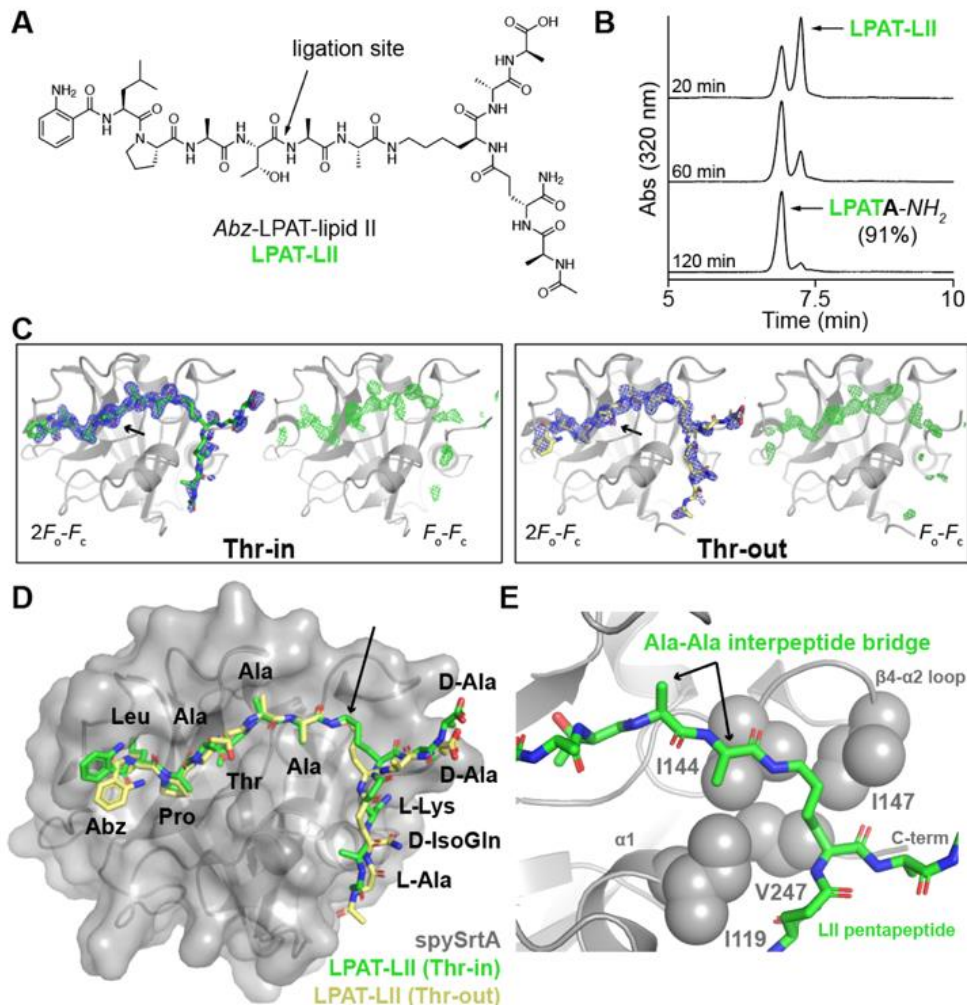
Figure 10. The structure of spySrtA bound to a peptide model of the LPAT-lipid II ligation product. (A) Chemical structure of the *Abz*-LPAT-lipid II (LPAT-LII) model with the ligation site between the LPAT fragment and lipid II highlighted. **(B)** RP-HPLC chromatograms showing efficient cleavage (91% 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 F_o-F_c 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 F_o-F_c 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.

27

streptococcal species possess LPXTGE motifs, with glutamic acid occupying this P2' position.[37] 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).

2.2.5: 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 (Figure 10C). 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 A1). Briefly, MD simulations were performed in full atomistic detail with explicit water using the AMBER99SB*-ILDN force fields.[53] 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 Materials and Methods and Supplementary Information.

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. (Figure A7A). The root-mean-square-fluctuation (RMSF) of backbone atoms for spySrtA were also consistent with a well-defined 8-stranded antiparallel b-barrel sortase core structure, in that these regions are relatively inflexible over the course of the simulation, as compared to some a-helical and all loop regions (Figure 11A). 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 (Figure 11B), as well as the average RMSF of backbone atoms in each peptide (Figure A7B). 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.

28
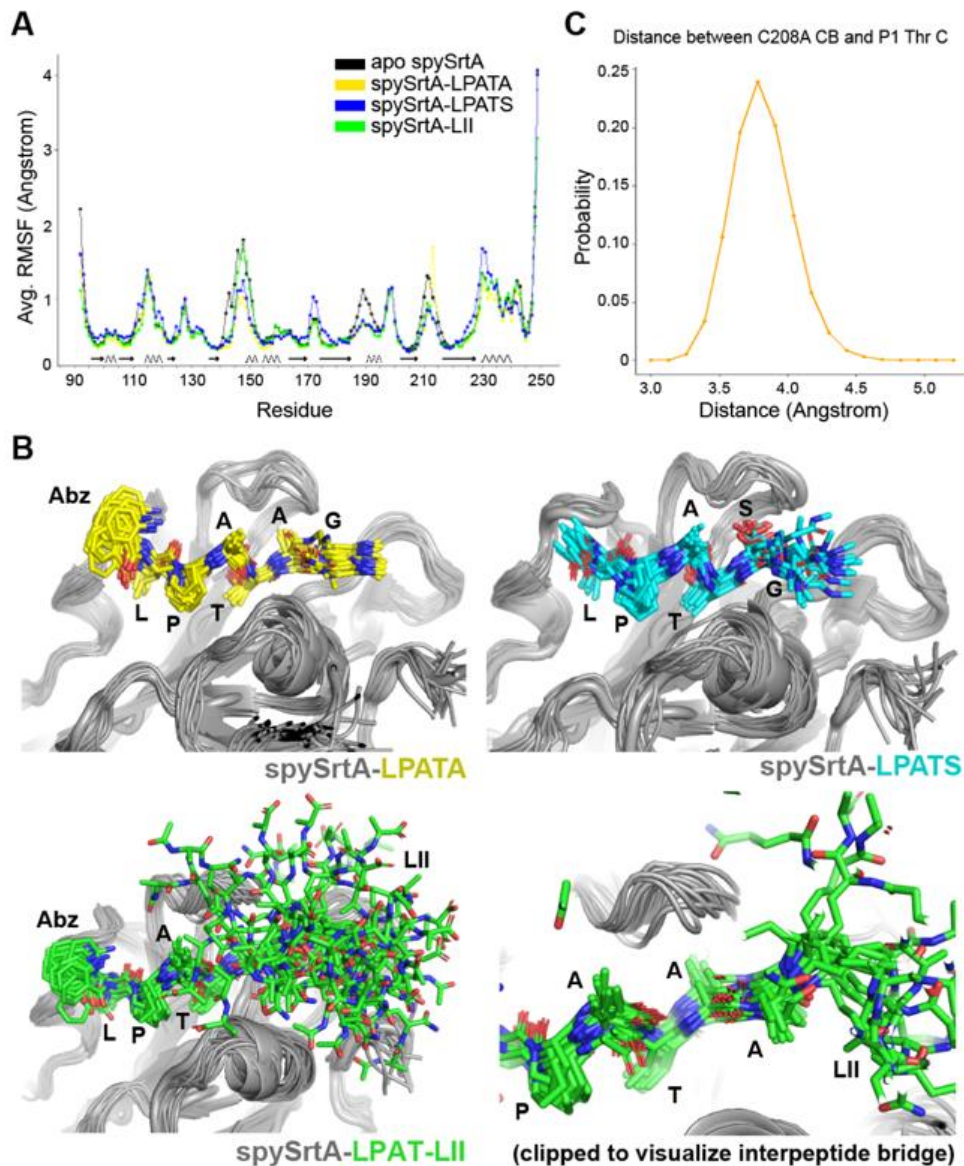
Figure 11. Molecular dynamics simulations of spySrtA structures. **(A)** The average root-mean-square-fluctuations (RMSF) of backbone atoms in each residue are shown and colored as labeled. The secondary structure elements are indicated under the curve by arrows for β-strands and curved lines for α-helices. 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 (C_β) 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.

We also analyzed the distance distribution between the C208A methyl group (or C_b 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 (Figure 11C). Surprisingly, our experimentally observed distance of 3.4 Å (Figure 8D) 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 (Figure A7C). We see a dramatic increase in flexibility in atoms in the LII pentapeptide, as compared to the LPAT and interpeptide bridge sequences (Figure A7C). 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 (Figure 11B). Taken together, these molecular dynamics simulations strongly support our described structure-based conclusions.

2.3: Discussion

As we highlight in Figure 6, 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).[1,28,54] 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.[17,30] 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.[6,20,39,55] Finally, in our LPAT-LII structure, we see both "Thr-in" and "Thr-out" conformations, molecular orientations that have also been previously described.[1,17,30]

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.[17,30] 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 (Figure 9D, Figure A4E), and appears to play a role in positioning the LPXTX motif through direct contacts with the P4 and P3 carbonyl groups (Figure 8B), 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 (Figure 8B, Figure 9C), 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 (Figure 9D, Figure A4E).

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.[56] 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.[20,57–59] 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 *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 *in vitro.*[60]

In summary, this work reports the first crystal structures of spySrtA bound to an LPXTX substrate, as well as a model of the *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.

2.4: Materials & Methods

2.4.1: Expression and Purification of SpySrtA Protein

Wild-type spySrtA, C208A spySrtA, T207A spySrtA, and R216A spySrtA genes were recombinantly expressed using *Escherichia coli* BL21 (DE3) cells in the pET28a(+) vector (Genscript), as previously described.[35] The wild-type sequence used matches that of the published spySrtA structure, PDB ID 3FN5.[16] 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 spySrtA 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.[20,35] 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.[61] 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 mM WIDEPORE C4 200 Å column (100 x 2.1 mm) [H_2O (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 Figure A1.

2.4.2: 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.[20] 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 Figure A8.

2.4.3: 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^L mass spectrometer. Separations were achieved with a Phenomenex Kinetix® 2.6 mM C18 100 Å column (100 x 2.1 mm) [aqueous (95% H_2O, 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].

2.4.4: 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.[20] 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 (I_{ex} = 320 nm, I_{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.

2.4.5: 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 mL well solution to protein solution ratio of 1:1, for a final drop volume of 4 mL (2 mL + 2 mL). Crystallization conditions were optimized using those for the wild-type apo protein.[16] 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.

2.4.6: 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 l= 1.00004 nm or 0.97741 nm over 360°, with Df=0.25° frames and an exposure time of 0.5 s per frame. Data were processed using the XDS package (Table 2).[62,63] 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.[64–66] 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.[67] All crystal data and refinement statistics are in Table 2. Structural analyses and Figure rendering were done using PyMOL. PDB accession codes for the structures presented here are in Table 2 and are (for all, containing C208A spySrtA): LPATA (7S51), LPATS (7S40), LPAT-LII "Thr-in" (7T8Y), and LPAT-LII "Thr-out" (7T8Z).

2.4.7: Molecular Dynamics Simulations of SpySrtA.

All molecular dynamics simulations were performed using GROMACS 2020.4 with the AMBER99SB*-ILDN force fields.[53,68–70] Additional details and relevant references are in the Supplementary Information.[71–80]

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.

Data Availability

All data are contained in the article and the Supporting Information.

Acknowledgments

Chapter 3: Literature Review and Preliminary Experiments with Sortase B Enzymes

Sophie N. Jackson, Darren E. Lee, Jadon M. Blount, Hanna M. Kodama, Kyle M. Whitham,

James McCarty, John M. Antos, Jeanine F. Amacher

Contributions by Thesis Author

I performed the primary literature review, in conjunction with Dr. Jeanine Amacher. I also expressed and purified wild-type and mutant sortase B enzymes. I selected the sequences of peptide mimetic substrates and contributed to the synthesis, purification, and characterization of peptide substrates. I chose reaction conditions for assays and performed assays (both FRET-based and HPLC) and LC-MS characterization sortase B-catalyzed reactions. I performed data analysis for the assays presented here. I contributed to crystallization experiments and generated the substrate-bound cdSrtB AlphaFold2 (Galaxy) model.

3.1: Introduction

To begin our study of sortase B *in vitro* activity, we first conducted a literature review in order to select promising sortase B homologs for study. Specifically, we were searching for sortase B enzymes for which:

1. Crystal or NMR structures had been solved;

2. Substrates had been identified; and

3. *In vitro* activity had either been observed or had not been tested

Using these criteria, we selected sortase B enzymes from *Bacillus anthracis* (baSrtB)*, Clostridioides difficile* (cdSrtB)*, Listeria monocytogenes* (lmSrtB), and *Staphylococcus aureus* (saSrtB)*. After selecting these sortase B enzymes, we performed preliminary *in vitro* activity tests using peptide substrates. Substrate sequences were selected based on previously published studies and in all cases included at least one residue upstream and two residues downstream of the pentapeptide motif. We also sought to gain insights into key enzyme-substrate interactions, first by attempting to co-crystalize enzymes with peptide substrates and then by generating substrate-bound enzyme models using AlphaFold2 (Galaxy). *In vitro* activity was tested using a fluorescence-based assay and LCMS analysis. As expected, *in vitro* activity was observed for baSrtB, cdSrtB, and saSrtB. BaSrtB was by far the most active sortase B enzyme in our hands and was selected for further selectivity and mutagenic study as described in chapters 4 and 5 of this work. LmSrtB was inactive under the conditions tested.

Though significant, prior research on class B sortases has been limited. In this work, we identify several promising sortase B homologs for further study. We also developed substrate-bound models that may be used to better understand the enzyme-substrate interactions that govern sortase B activity. These represent the first steps towards developing our research on class B sortases, and form the foundation of the following chapters, which focus on class B sortase from *Bacillus anthracis*, the most active sortase homolog identified in this work.

3.2: Results

3.2.1: Literature Review and Selection of Sortase B Homologs for Study

A PDB search revealed solved structures for sortase B enzymes from *Streptococcus pyogenes* (PDB code 3PSQ), *Clostridium perfringens* (5B23), *Bacillus anthracis* (multiple; 1RZ2), *Clostridioides difficile* (multiple; 4UX7), *Staphylococcus aureus* (multiple; 1NG5), and *Listeria monocytogenes* (5JCV) (Figure 12). These show broad structural agreement, including the characteristic sortase β-barrel core and flexible loops. They also have the expected elongated β6-β7 loop and additional N-terminal α-helices expected for sortase B enzymes.
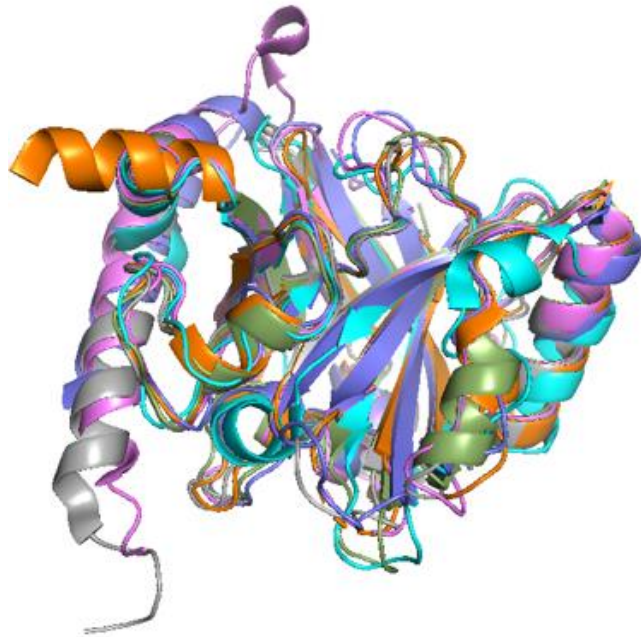
Figure 12. Experimentally determined structures for sortase B enzymes. Structures have been solved for sortase B enzymes from *S. pyogenes* (3PSQ, cyan), *C. perfringens* (5B23, green), *B. anthracis* (1RZ2, pink), *C. difficile* (4UX7, gray), *S. aureus* (1NG5, purple), and *L. monocytogenes* (5JCV, orange). These show broad structural agreement.

Streptococcus pyogenes sortase B (spySrtB) is a pilin polymerizing sortase that recognizes two protein substrates with motif sequences EVPTG, QVPTG, VPPTG, and VVPTG, depending on the strain.[81] However, spySrtB does not cleave the peptide KDF<u>EVPTG</u>VAM *in vitro*.[81] The cofactor protein SipA is necessary for spySrtB *in vivo* activity and has been proposed to act as a chaperone protein, although the exact role of SipA in spySrtB activity is unclear.[5,82] However, incubating spySrtB and spySrtB substrate protein with SipA does not result in observable *in vitro* activity.[82] Sortase B from *Clostridium perfringens* (cpSrtB) does not have an established substrate or recognition motif.[83] We did not pursue spySrtB or cpSrtB for our initial activity assays due to inactivity and unknown substrate identity, respectively.

BaSrtB anchors the heme-scavenging protein IsdC to the cell wall and recognizes the motif NPKTG.[8,84] BaSrtB has been shown to cleave substrate mimetics containing this motif and select variations of it *in vitro* and to catalyze ligation to an N-terminal glycine-containing nucleophile.[8,84] Unusually for sortase B, the enzyme from *Clostridioides difficile* (cdSrtB) appears to recognize 7 protein substrates.[13,85] The proposed recognition motifs for cdSrtB are SPKTG, PPKTG, SPSTG, and SPQTG.[85] CdSrtB has been shown to cleave both SPKTG- and PPKTG-containing peptides *in vitro* and to ligate peptides to *m*-diaminopimelic acid as well as glycine nucleophiles.[13,85–88] *Staphylococcus aureus* sortase B (saSrtB) performs the same endogenous function as baSrtB: anchoring IsdC to the cell wall.[89] SaSrtB recognizes the motif NPQTN and catalyzes cleavage and transpeptidation reactions with NPQTN-containing substrate peptides and glycine nucleophiles *in vitro.*[2,89] *Listeria monocytogenes* sortase B (lmSrtB) anchors two proteins thought to be involved in iron acquisition to the cell wall.[14,90] Mariscotti et al (2009) showed that lmSrtB recognizes the motifs NAKTN and NPKSS. To the best of our knowledge, lmSrtB *in vitro* activity has not previously been tested. BaSrtB, cdSrtB, saSrtB, and lmSrtB met our criteria for study as outlined in 3.1: Introduction. Information about the endogenous substrates

Sortase	Substrate	Substrate Role	Substrate Motif	Nucleophile
BaSrtB	IsdC	Iron acquisition	NPKTG	*meso*-diaminopimelic acid*
CdSrtB	CD0386	Adhesin*	SPKTG	
	CD0183*	Cell wall hydrolase	SPSTG	
	CD2537*	Nucleotidase	SPKTG	
	CD2768*	Cell wall hydrolase	SPQTG	*meso*-diaminopimelic acid*
	CD2831*	Adhesin*	PPKTG	
	CD3246*	Surface protein	SPKTG	
	CD3392*	Surface protein	SPKTG	
LmSrtB	Lmo2185	*Unknown*	NAKTN	*meso*-diaminopimelic acid*
	Lmo2186		NPKSS	
SaSrtB	IsdC	Iron acquisition	NPQTN	Pentaglycine

Table 3. *In vivo* substrates. BaSrtB and saSrtb anchor the heme transporter IsdC to the cell wall. LmSrtB anchors Lmo2185 and Lmo2186, however the function of these proteins is currently unknown. Seven possible substrates have been identified for cdSrtB, but only one, CD0386, has been demonstrated experimentally. The identified substrate motif is indicated as is the endogenous nucleophile (a component of the cell wall in all cases shown). Asterisks (*) mark information that is suspected but not experimentally confirmed.

for these enzymes is shown in Table 3. Detailed information about published *in vitro* activity assays is included Table A2, parts A-B.

3.2.2: Attempts to Crystalize Substrate-Bound BaSrtB, CdSrtB, and SaSrtB

In *Chapter 2: Structures of* Streptococcus pyogenes *Class A Sortase in Complex with Substrate and Product Mimics Provide Key Details of Target Recognition*, we crystallized a catalytically inactive spySrtA in the presence of known *in vitro* substrates LPATA and LPATS.[22] We attempted a similar co-crystallization here using the inactive mutants C233A baSrtB, C239A cdSrtB, and C223A saSrtB, where the catalytic cysteine has been mutated to an alanine. We attempted to co-crystallize C233A baSrtB with *Ac*-DNPKTGDE-*NH₂,* C239A cdSrtB with *Ac*-PVPPKTGDS-*NH₂,* and C223A saSrtB with *Ac*-KVENPQTNAG-*NH₂* using a range of conditions based on previously described conditions that were used to crystalize apo-enzymes.[2,87,91] We also performed a crystallization condition screen (PEG/Ion Screen Macromolecular Crystallization Kit, Hampton Research HR2-126) for C233A baSrtB with *Ac*-DNPKTGDE-*NH₂,* and attempted co-crystallization based on several conditions that formed apo-enzyme crystals. Though crystals were obtained for many srtB conditions, crystals were either not diffraction quality and/or did not have bound substrate (data not shown).

A structure for C233A baSrtB indicated that proteins related by crystallographic symmetry may be interacting in such a way that the N-terminus of one protein nears the active sight of another (data not shown). Based on this result, we attempted to crystallize a C233A baSrtB construct that was N-terminally fused to part of the substrate sequence (C233A baSrtB_NPKTG, sequence in appendix of this work) as well as a baSrtB construct that was N-terminally truncated by 7 amino acid residues (C233A baSrtB$_{42-254}$). No diffraction-quality crystals were obtained for these constructs under any conditions tested. However, we also ordered an active version of both mutants. The active baSrtB_NPKTG enzyme reacted with itself to produce a cleaved product (sequence in appendix), which yielded diffraction-quality crystals. We are currently refining these

structures and are performing experiments to determine if this cleaved construct is catalytically active. A preliminary structure is shown in Figure 13 (r_{work}/r_{free} = 0.19 / 0.24 at 1.8 Å resolution). The β7-β8 loop appears to be resolvable in this structure (Figure 13B), and it has not been fully resolved in previous structures (Figure 13C). Interestingly, the β7-β8 loop is in a different conformation than the AlphaFold2 (Galaxy) model described below (Figure 13D). This confirmation is not favorable for substrate binding as indicated in that model. The β7-β8 loop conformation in our structure may be a crystallization artifact, as proteins related by
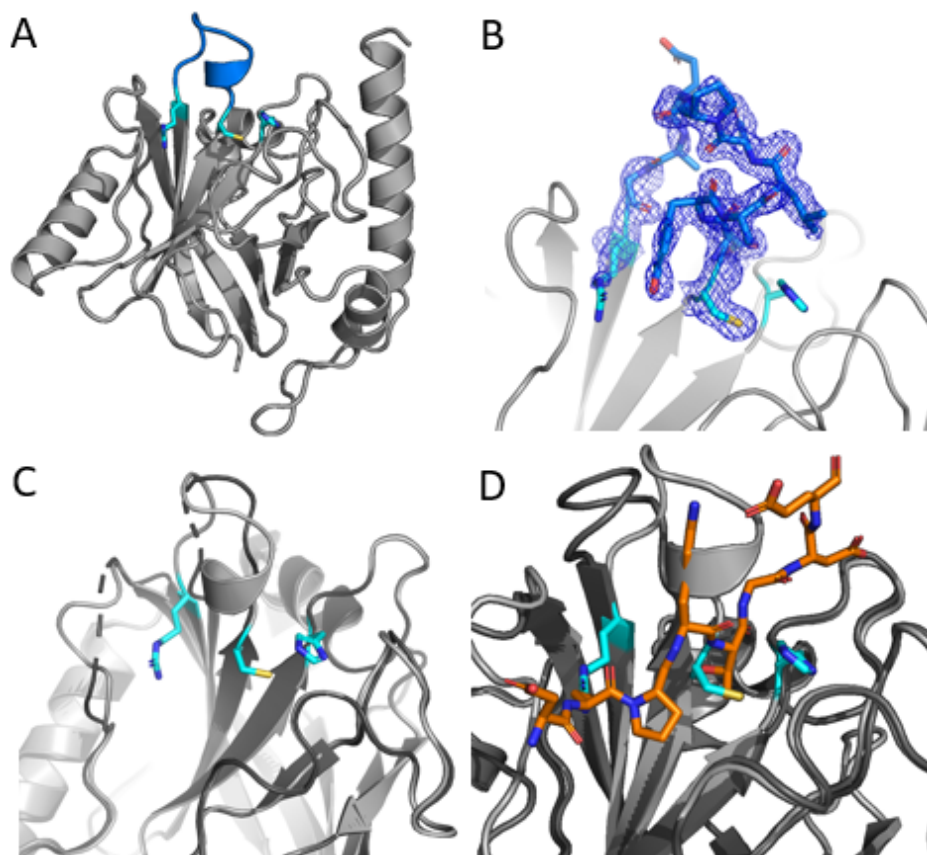
Figure 13. A preliminary baSrtB structure. A preliminary structure determined from a self-cleaved substrate-fused baSrtB protein, shown in cartoon representation **(A)** (β7-β8 loop in dark blue, catalytic triad in cyan with side chains in stick representation). The β7-β8 loop appears to be resolvable **(B)** (electron density in blue mesh, loop colored dark blue with side chains in stick representation). **(C)** The β7-β8 loop has not been resolved in previous baSrtB crystal structures (2OQW, dark gray). **(D)** The β7-β8 loop is in a different conformation than the substrate-fused AlphaFold2 (Galaxy) model (dark gray) and this conformation is not favorable for substrate binding as depicted in that model (orange, side and main chains in stick representation).

crystallographic symmetry appear to be interacting along this region and the N-terminus (data not shown). Alternatively, this β7-β8 loop conformation may reflect a conformation that the loop adopts *in vitro*. While interesting, this structure is also ultimately an apo-enzyme structure and does not provide information about substrate-enzyme interactions.

3.2.3: Creation of Substrate-bound Models using Full Length Enzyme and Substrate Sequences

Full-length substrate protein sequences were linked at the C-terminus to full-length sortase B protein sequences using a flexible glycine-serine linker sequence and submitted to AlphaFold2 (Galaxy), a platform that uses artificial intelligence to predict protein secondary and tertiary structure. Full sequences are included in the appendix of this work. For baSrtB (Figure 14), lmSrtB (Figures 15-16), and saSrtB (Figure 17), substrate proteins appear bound to the respective active sites along the expected motif sequence. Specifically, substrate motifs are positioned within the catalytic pocket and the catalytic cysteine is positioned 3.6 - 3.9 Å away from the carbonyl carbon of the P1 threonine residue. The catalytic cysteine is positioned 3.7 Å away from the P1 serine residue for lmSrtB modeled with NPKSS substrate. The catalytic arginine also forms stabilizing hydrogen bonds with the P4 and P3 carbonyl groups. Additionally, there appears to be an interaction between the P4 asparagine and the β6-β7 loop, consistent with experimental data showing that this loop is important for differentiating between the LPXTG and NPQTN sequence in sortases from *S. aureus*.[92] For cdSrtB, substrate protein appeared shifted within the active site (Figure 18). The catalytic cysteine is positioned 8.0 Å away from the carbonyl carbon of the P1 threonine and it does not appear that the catalytic arginine is stabilizing the substrate (Figure 18). Additionally, AlphaFold2 (Galaxy) rated the pentapeptide region's positioning with low (<50 pLDDT) confidence. Interestingly, the glycine-serine linker in this model was also low confidence and was not resolved in the resulting pdb file. Because of these issues, the substrate-bound cdSrtB model cannot be used to hypothesize about enzyme-substrate interactions.

The baSrtB, lmSrtB, and saSrtB models serve as a reasonable basis to begin to hypothesize about potentially important enzyme-substrate interactions.

Figure 14. A model of substrate-bound baSrtB. AlphaFold2 (Galaxy) was used to generate a model of the endogenous substrate IsdC bound to baSrtB. **(A)** The full-length IsdC protein sequence (grey, NPKTG sequence shown in orange with main and side chains as sticks) is linked to the baSrtB (cyan) N-terminus by a glycine-serine linker (yellow). **(B)** The peptide binding motif NPKTG is positioned within the active site in a way that appears to support experimental data. Catalytic residues are labeled with sides chains shown as sticks.
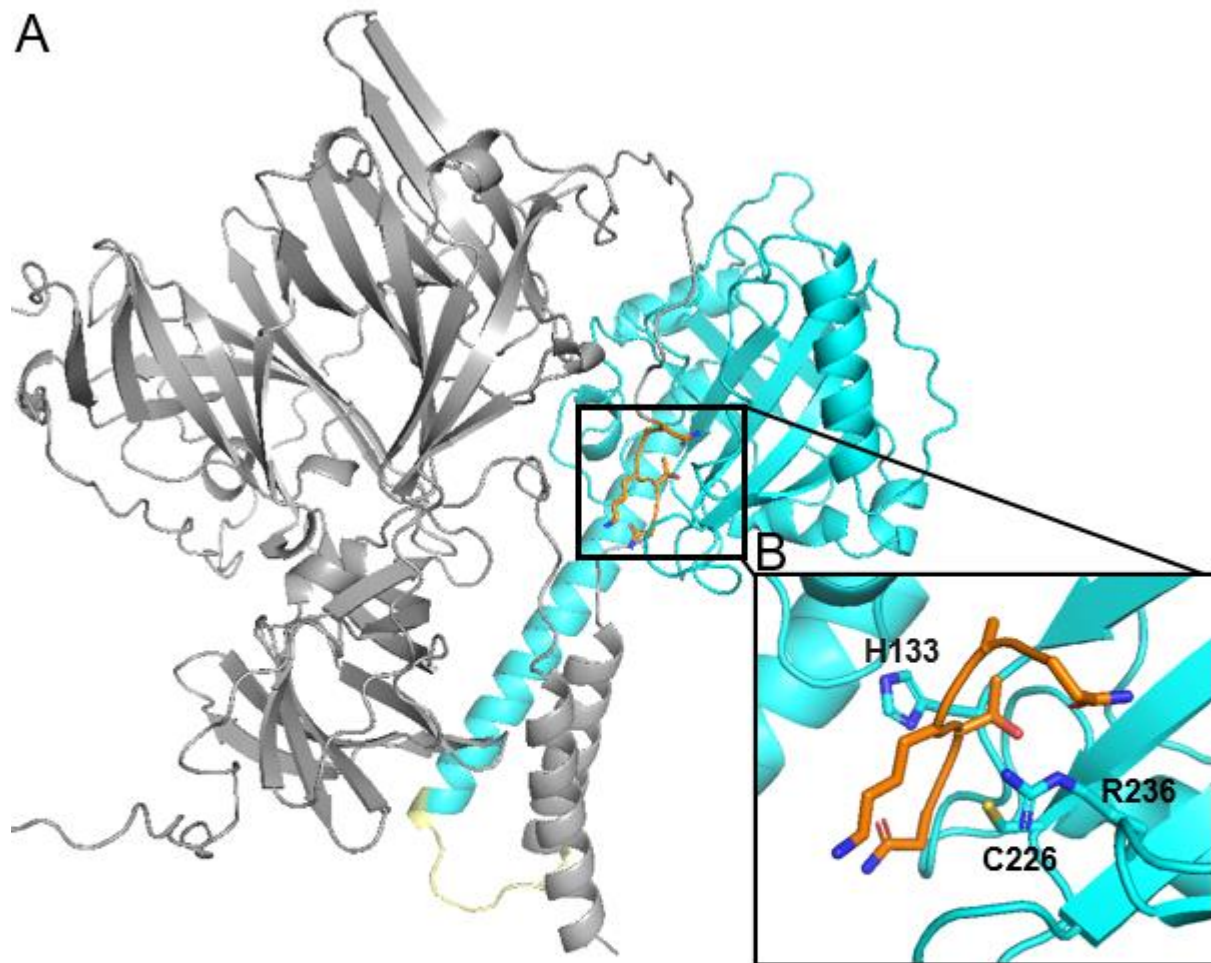
Figure 15. A model of Lmo2185-bound lmSrtB. A model of the endogenous substrate Lmo2185 bound to saSrtB was generated using AlphaFold2 (Galaxy). **(A)** The full-length Lmo2185 protein sequence (grey, portions of a long IDR region not depicted, NAKTN sequence shown in orange with side chains as sticks) was linked to the N-terminus of lmSrtB (cyan) using a glycine-serine linker (yellow). The peptide binding motif NAKTN is positioned within the active site in a way that appears to support catalysis **(B)**. Catalytic residues are labeled with sides chains shown as sticks.
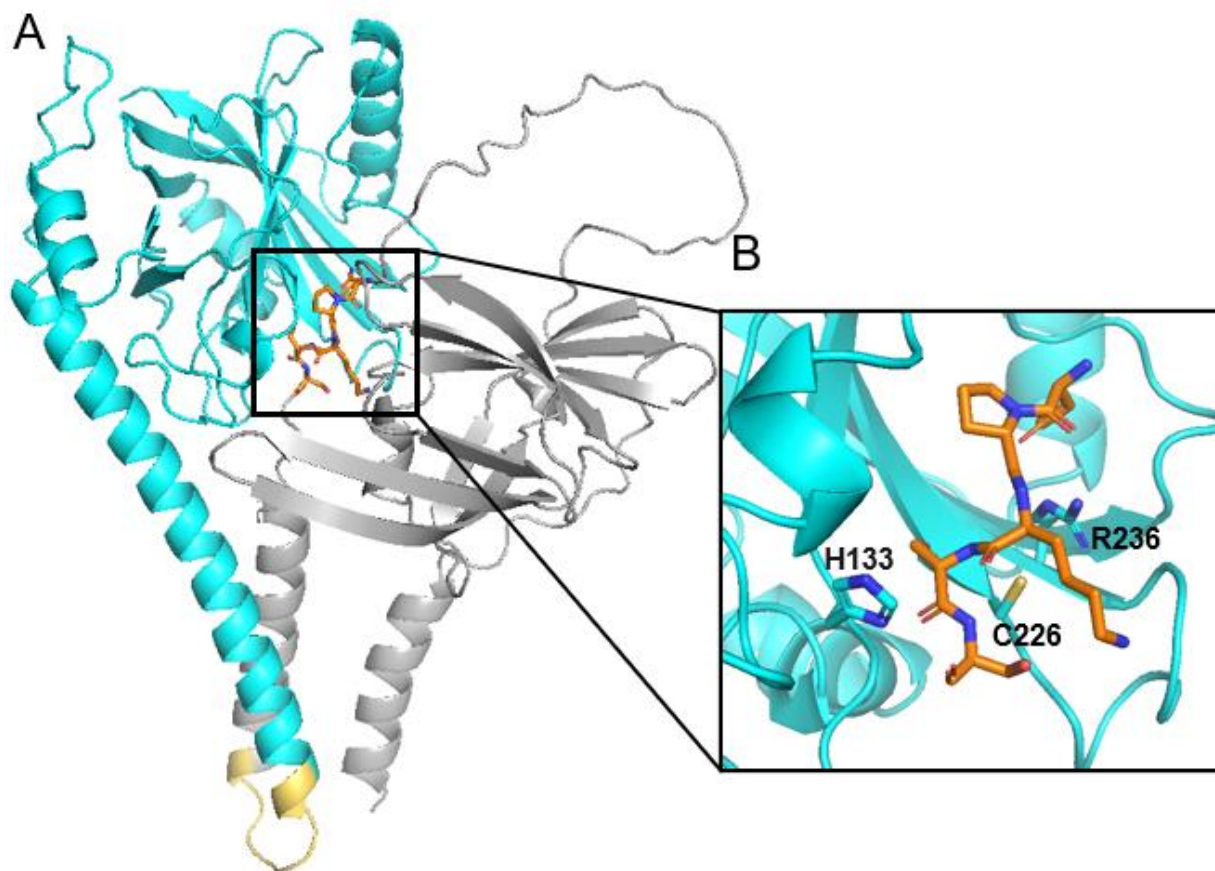
Figure 16. A model of Lmo2186-bound lmSrtB. A model of the endogenous substrate Lmo2186 bound to saSrtB was generated using AlphaFold2 (Galaxy). **(A)** The full-length Lmo2186 protein sequence (grey, portions of a long IDR region not depicted, NPKSS sequence shown in orange with side chains as sticks) was linked to the N-terminus of lmSrtB (cyan) using a glycine-serine linker (yellow). The peptide binding motif NPKSS is positioned within the active site in a way that appears to support catalysis **(B)**. Catalytic residues are labeled with sides chains shown as sticks.
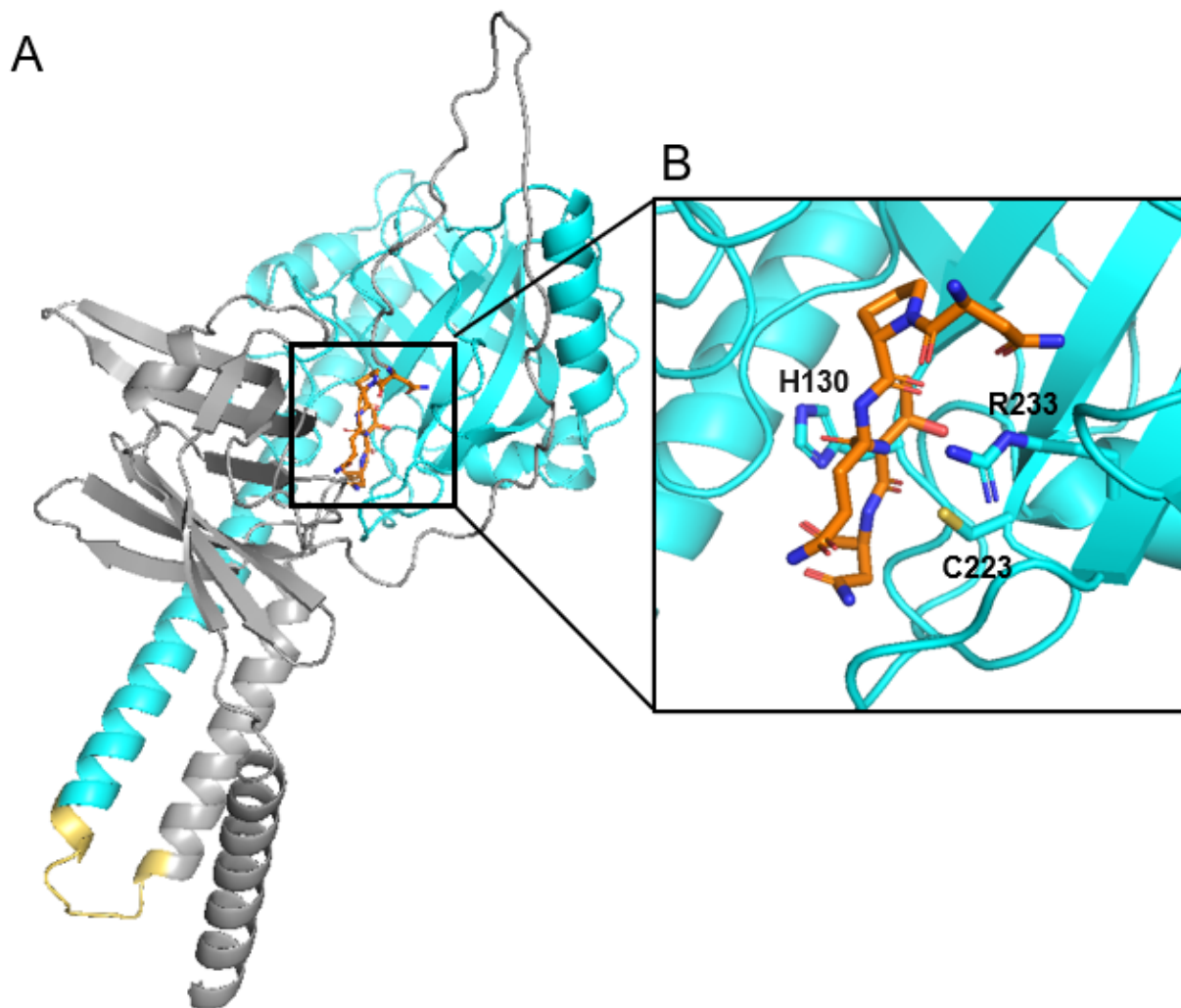
Figure 17. A model of substrate-bound saSrtB. A model of the endogenous substrate IsdC bound to saSrtB was generated using AlphaFold2 (Galaxy). **(A)** The full-length IsdC protein sequence (grey, NPQTN sequence shown in orange with main and side chains as sticks) was linked to the N-terminus of saSrtB (cyan) using a glycine-serine linker (yellow). The peptide binding motif NPQTN is positioned within the active site in a way that appears to support catalysis **(B)**. Catalytic residues are labeled with sides chains shown as sticks.
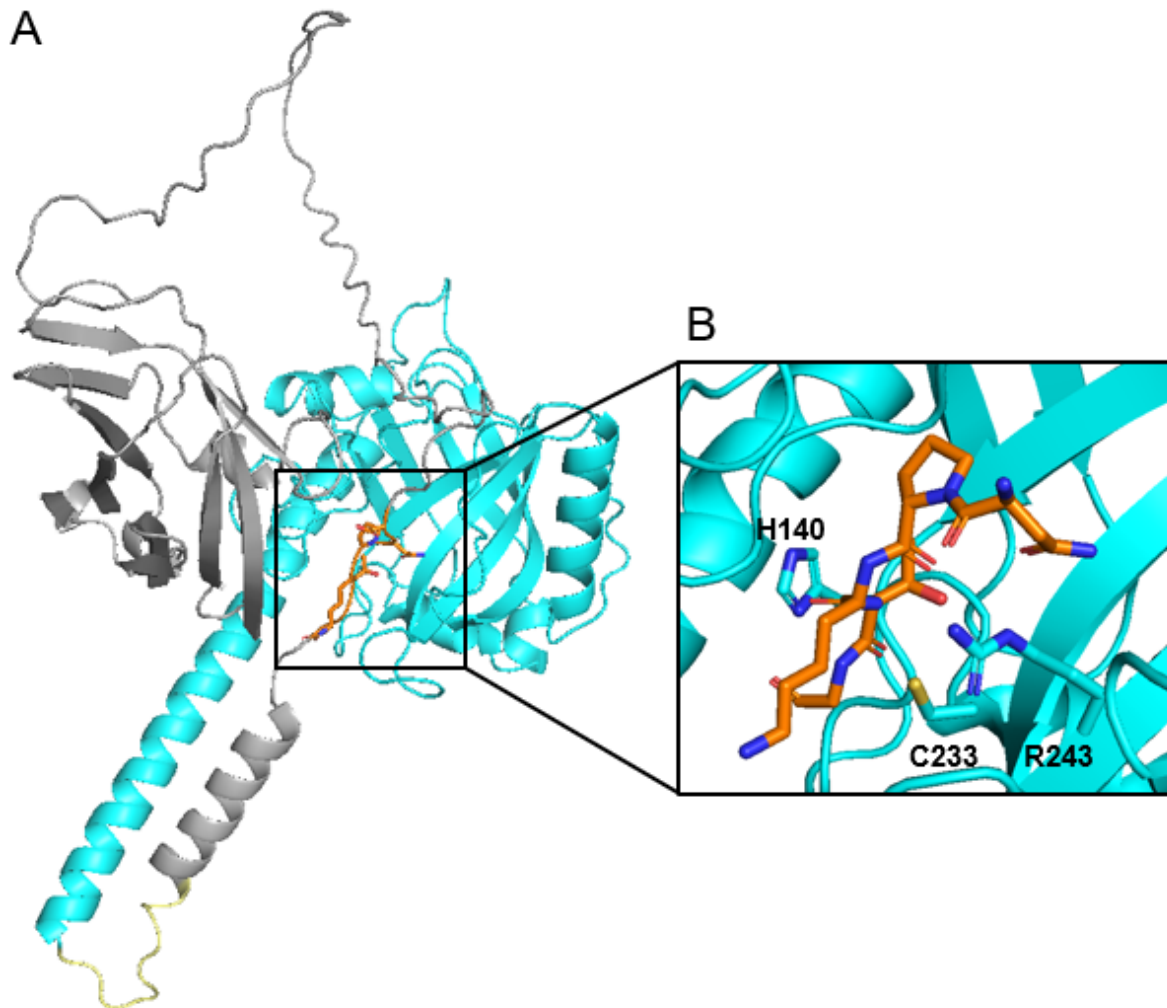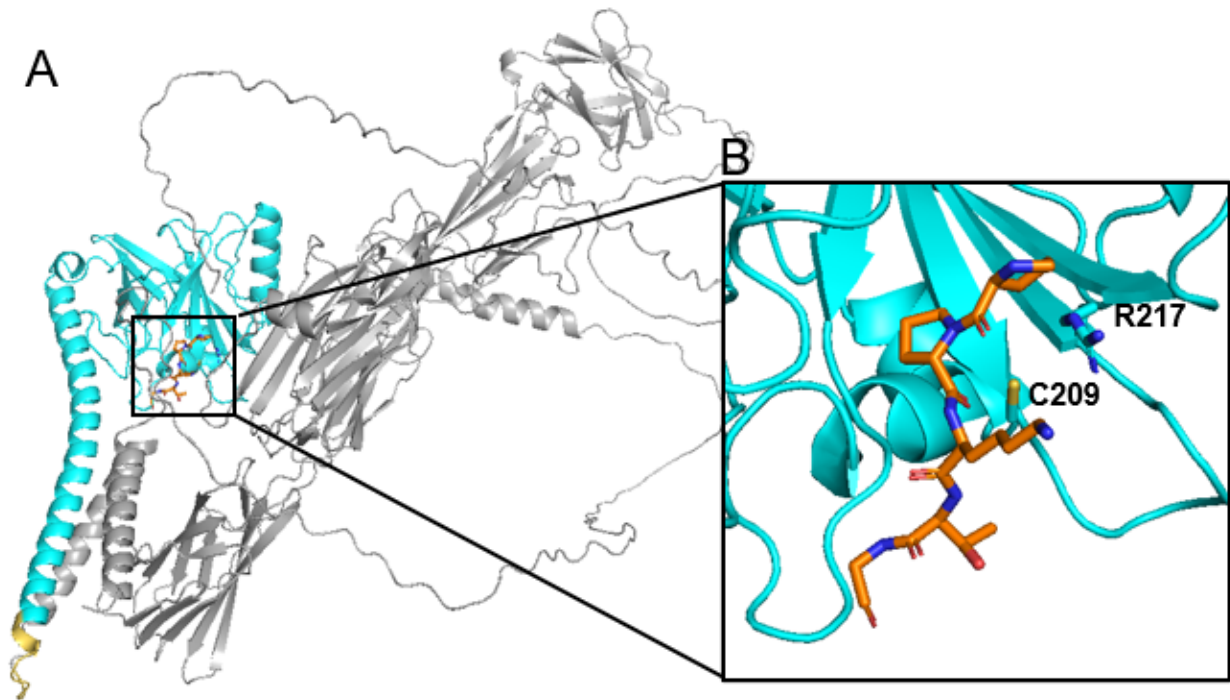
Figure 18. A model of substrate-bound cdSrtB. AlphaFold2 (Galaxy) was used to generate a model of the endogenous substrate CD2831 bound to baSrtB. **(A)** The majority of the full-length CD2831 protein sequence (grey, PPKTG sequence shown in orange with side chains as sticks) was linked to the N-terminus of cdSrtB (cyan) using a glycine-serine linker (yellow). **(B)** The peptide binding motif PPKTG is positioned near the active site but appears shifted away from the catalytic cysteine. Catalytic cysteine and arginine labeled with sides chains depicted as sticks. Histidine is not labeled as Kang et al. (2020) report that histidine is not catalytically necessary in cdSrtB.

3.2.4: Preliminary *In Vitro* Activity Assays

3.2.4:1: BaSrtB Preliminary *In Vitro* Activity Assays

We tested baSrtB *in vitro* activity using the peptide substrate mimetic *Abz*-DNPKTGDEK(Dnp)-*NH₂* (pentapeptide motif underlined). A FRET-based activity assay that has been described previously (see *Chapter 2: Structures of Streptococcus pyogenes Class A Sortase in Complex with Substrate and Product Mimics Provide Key Details of Target Recognition*) was used to test for *in vitro* activity. BaSrtB activity was tested using "low" (10 μM enzyme, 50 μM substrate) and "high" concentration conditions (50 μM enzyme, 175 μM

substrate). Though activity was observed in both conditions, the high concentration conditions resulted in a much stronger fluorescence increase in a 1.8-hour activity assay (Figure 19A).

BaSrtB was not active in a FRET-based assay with other sortase B homolog substrate sequences used in this work or with the sortase A substrate sequence Abz- LPATGGK(Dnp)-NH_2 (data not shown).
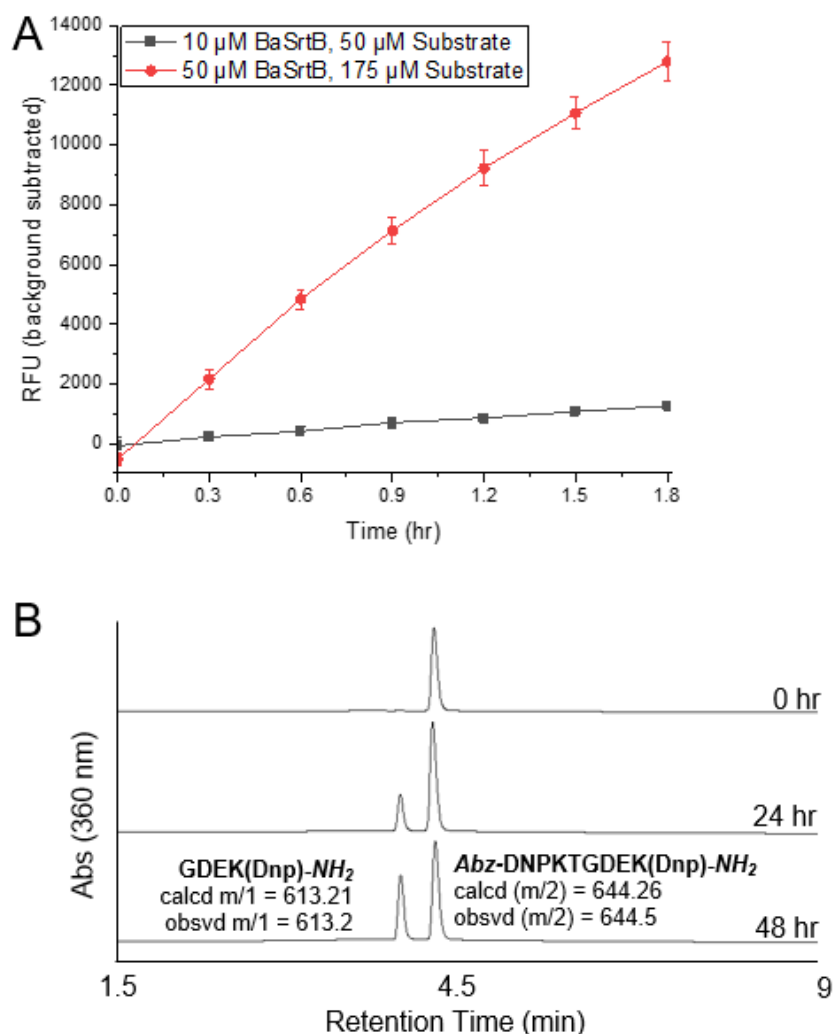
Figure 19. Preliminary baSrtB activity assays. (A) BaSrtB activity can be observed in a 1.8-hour FRET-based assay when enzyme and substrate concentrations are 10 μM and 50 μM, respectively (black squares). However, the signal can be much more clearly observed when concentrations of enzyme and substrate are increased to 50 μM and 175 μM, respectively (red circles). (B) BaSrtB shows substrate cleavage over a 48-hour time period in an HPLC-based assay (10 μM baSrtB, 50 μM substrate).

An HPLC activity assay also showed substrate cleavage at 24 and 48 hours (Figure 19B). The identity of the C-terminal cleavage product was confirmed via LCMS. The N-terminal ligation product was not observed in the preliminary HPLC-based assay, but was subsequently confirmed via LCMS in the nucleophile assay described in *Chapter 4: The Structural and Biochemical Basis of BaSrtB Selectivity and Activity.*

3.2.4.2: Preliminary *In Vitro* Activity Assays: CdSrtB

We tested cdSrtB *in vitro* activity using a peptide substrate mimetic with the sequence Abz-PV<u>PPKTG</u>DSK(Dnp)-*NH₂*. CdSrtB activity was not apparent in a fluorescence assay using this substrate under any of the conditions tested (Table A3) (data not shown). However, *in vitro* activity by cdSrtB was observed in an HPLC-based assay after 24 and 48 hours of incubation (n = 1) (Figure 20). Activity levels were very low, however. Additionally, there was a small peak visible on the 320 nm
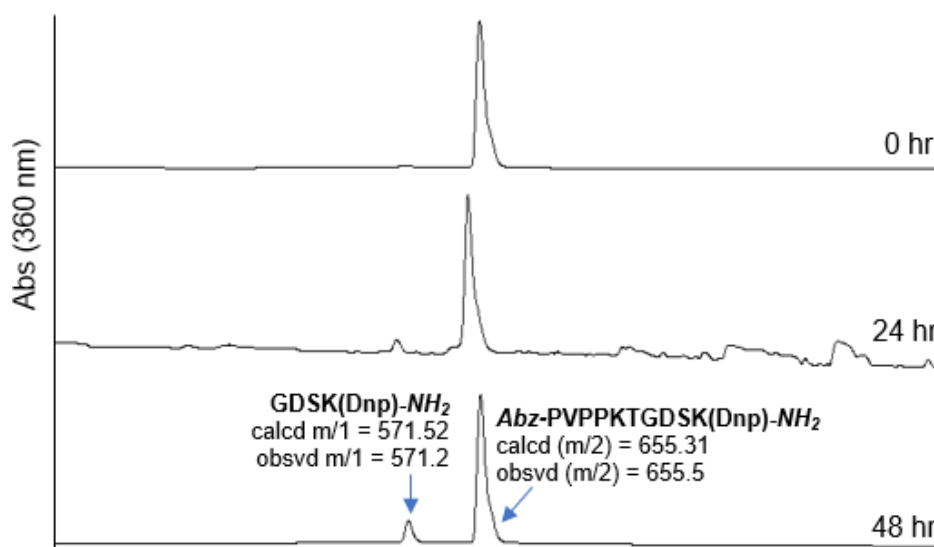
Figure 20. A 48-hour HPLC assay of cdSrtB. CdSrtB cleaves the substrate peptide *Abz*-PVPPKTGDSK(Dnp)-*NH₂* over 48 hours. The identity of the substrate and C-terminal product was verified using LCMS.

chromatogram with a mass that agreed with the expected value for the N-terminal ligation product (Figure A9). The signal of this analyte was small and did not rise above background levels. Verifying the N-terminal ligation product should be revisited in future experiments.

3.2.4.3: Preliminary *In Vitro* Activity Assays: LmSrtB

LmSrtB activity was not observed in a fluorescence assay using the substrate *Abz*-T<u>NPKSS</u>DSK(Dnp)-*NH₂* (data not shown). An HPLC-based assay was performed and showed no product formation after 48 hours of incubation (n = 1) (data not shown).

3.2.4.4: Preliminary *In Vitro* Activity Assays: SaSrtB

A small increase in fluorescence was observed in a saSrtB FRET-based assay using Abz-KVE<u>NPQTN</u>AGK(Dnp)-*NH₂* (n = 2) (Figure 21A). The negative control fluorescence decreased during the assay on a similar scale to the increase seen in reaction fluorescence. For this reason, the background RFU was not subtracted to avoid artificially increasing the reaction fluorescence. A separate reaction mixture was analyzed via HPLC after approximately 20 hours of incubation at room temperature (Figure 21B). The identity of the C-terminal cleavage product was verified using LCMS. Further study of saSrtB was significantly
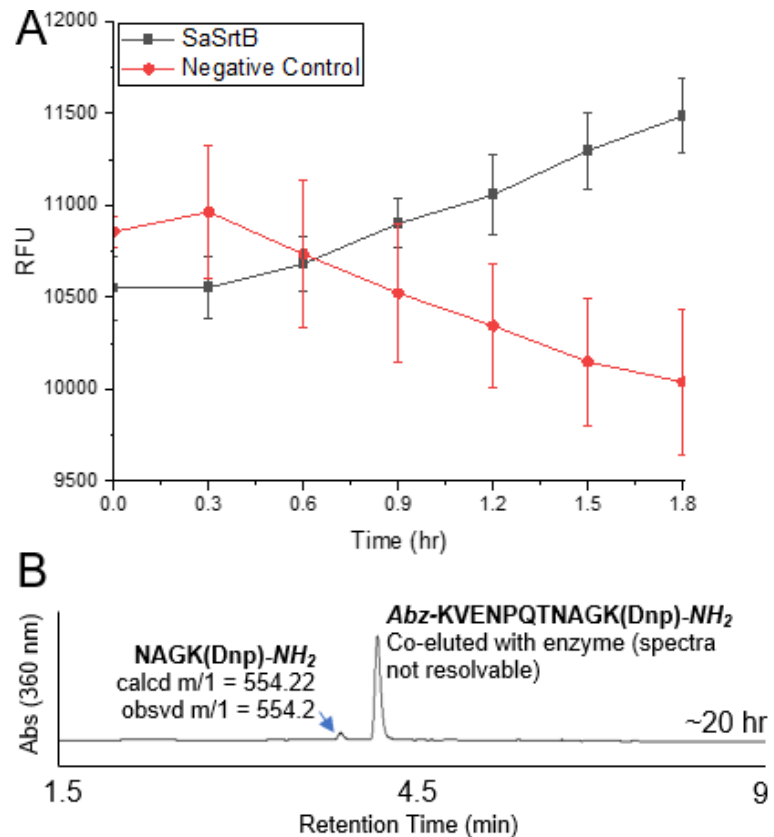
Figure 21. SaSrtB cleaves *Abz*-KVENPQTNAGK(Dnp)-*NH₂* in vitro. (A) SaSrtB shows a low level of activity in a FRET-based activity assay over 1.8 hours. **(B)** SaSrtB cleaves *Abz*-KVENPQTNAGK(Dnp)-*NH₂* after approximately 20 hours of incubation at room temperature. The C-terminal product (NAGK(Dnp)-*NH₂*) is visible in an HPLC chromatogram, and the mass was verified using LCMS. The intact substrate peak could not be verified because it co-eluted with the protein and could not be distinguished from the protein ladder, however the retention time is as expected based on analyses during peptide purification.

hindered by the relative insolubility of peptide mimetic substrates, which made purification and storage of stock solutions challenging. SaSrtB showed no activity with the baSrtB substrate *Abz-DNPKTGDEK(Dnp)-NH₂* (data not shown). Reactions catalyzed by saSrtB appeared to progress more slowly than those catalyzed by baSrtB.

3.2.2.5: Refining Reaction Conditions

The sortase A reaction conditions presented in this work utilize a pH 7.5 buffer. The preliminary baSrtB reactions presented here also used a pH 7.5 buffer. However, Puorger et al. (2017) reported that baSrtB was most active between pH 8.0 and 9.0.[8] We therefore decided to investigate the effect of pH on baSrtB activity in this assay. We performed an activity assay using reaction buffers that ranged in pH from 7.0 to 9.0 (n = 2) (Figure 22). It appears that baSrtB is most active in pH 7.5 and 8.0 reaction buffers in a 1.8-hour activity assay. In this assay, a pH of 8.5 and 9.0 appear slightly less favorable for baSrtB activity and a pH of 7.0 appears to be the least favorable. Based on these results and the findings of Puorger et al. (2017), we chose to perform future baSrtB assays in a pH 8.0 buffer.[8]
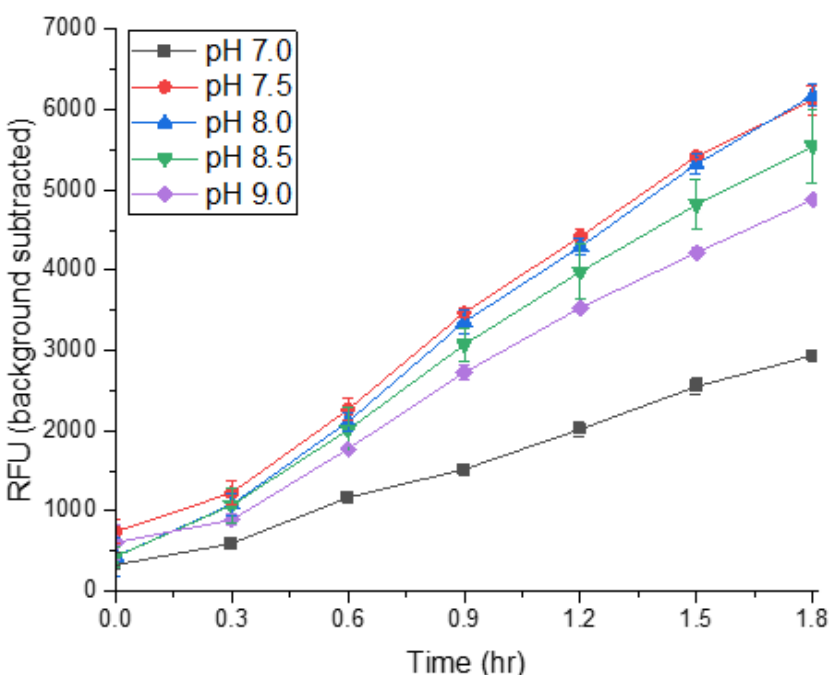
Figure 22. Investigating the effect of reaction pH on baSrtB activity. BaSrtB activity in a 1.8-hour assay is dependent on buffer pH. BaSrtB appears to be most active in reaction mixes using buffers at pH 7.5 (red circles) and pH 8.0 (blue triangles). BaSrtB is slightly less active in a pH of 8.5 (green triangles), followed by 9.0 (purple diamonds) and 7.0 (black squares).

Additionally, preliminary baSrtB, cdSrtB, and saSrtB reactions were performed using enzyme stocks made in a buffer containing 50 mM Tris pH 7.5 (baSrtB and cdSrtB) or 50 mM Tris pH 8.5 (saSrtB), 150 mM NaCl, and 1 mM TCEP. Sortase reaction mixtures contained an additional 1:10 dilution of a 10x sortase reaction buffer (500 mM Tris pH 7.5 or 8.5, 1500 mM NaCl). This is the method that was used for sortase A reactions described previously, but the high concentration of sortase B used in sortase B reactions meant that a large percent of the reaction mixture was made up of enzyme stock solution, sometimes in excess of 50% v/v. This resulted in higher concentrations of Tris and NaCl than may be desired for reactions, as well as a high concentration of residual TCEP, in preliminary assays. We therefore changed the sortase B stock solution to one that contained 100 mM Tris pH 8.0 (baSrtB, lmSrtB, cdSrtB) or pH 8.5 (saSrtB) and 300 mM NaCl. TCEP was omitted due to concerns that it may inhibit activity, although we did not see an effect for baSrtB (Figure A10). Sortase B reactions were then performed so that final reaction conditions contained 50 mM Tris, 150 mM NaCl, most commonly by diluting enzyme stock 1:2 in the creation of the reaction mixture. These conditions were adopted for preliminary assays using lmSrtB and for all non-preliminary baSrtB assays presented in this work.

3.2.2.5: Important Considerations for FRET-Based Activity Assays

A high concentration of fluorophore can suppress fluorescence, an effect known as "inner filter effects". As described in *Chapter 2: Structures of Streptococcus pyogenes Class A Sortase in Complex with Substrate and Product Mimics Provide Key Details of Target Recognition,* peptide mimetic substrates are synthesized with the fluorophore 2-aminobenzoyl (*Abz)* and the fluorescence of this chemical group is used to detect sortase activity in our FRET-based assays. Kruger et al. (2004) report that inner filter effects begin to affect 2-aminobenzoyl fluorescence at concentrations as low as 50 µM.[93] Our sortase B reaction conditions contain 50-200 µM of *Abz*-containing substrates. It is therefore likely that the high concentration of fluorophore present in these assays suppresses the fluorescence readings we obtain to some degree. Additionally, we

find that reaction mixtures are subject to evaporation when reactions are left uncovered in the Biotek plate reader during analysis. This effect is seen as early as 1.5 hours and becomes prominent after 6 hours (Figures A11-A12). As reaction volumes evaporate, the concentration of reactants, enzymes, and products increases. While the high starting concentrations of fluorophore likely decreased fluorescence in our reactions even prior to evaporation, evaporation does worsen fluorescence suppression over time. For this reason, data from preliminary assays was analyzed only for fluorescence readings taken within the first 2 hours. Later assays were analyzed with data taken within the first hour. It may be beneficial in the future to adjust sortase B experiment conditions to allow for monitoring FRET-based assays beyond this brief window.

3.3: Discussion

The four sortase B homologs we selected for study showed either no (lmSrtB) or extremely limited (baSrtB, cdSrtB, and saSrtB) *in vitro* activity. This finding is consistent with previous research performed using these enzymes.[8] Low *in vitro* activity makes studying these enzymes challenging in several ways. To compensate for low enzymatic activity, we increased enzyme and substrate concentrations. These high concentrations complicated both HPLC and FRET-based assays. The best separation of peptide substrates and products is achieved using a C18 HPLC column, however larger proteins, such as the approximately 25-30 kDa sortase B enzymes, are hard to clear from these columns and can result in significant carry-over at high concentrations. However, we find that reactions can be observed at lower concentrations of enzyme over longer reaction periods. Additionally, we find that high substrate concentration and evaporation over time can complicate the interpretation of FRET-based activity assays. This is an important finding that sheds light on the limitations of these assays and will guide future method development.

As expected based on previously published findings, we have observed clear *in vitro* activity for baSrtB, cdSrtB, and saSrtB. The apparent inactivity of lmSrtB with the substrate peptide *Abz*-T<u>NPKSS</u>DSK(Dnp)-*NH₂* is surprising given the enzyme's *in vivo* selectivity for the

NPKSS pentapeptide and the apparent suitability for this sequence for substrate binding (Figure 16).[94] Further investigations of lmSrtB *in vitro* activity may extend this peptide sequence to include additional upstream residues, e.g. *Abz*-**NKV**T<u>NPKSS</u>DSK(Dnp)-*NH₂*, as Bierne et al. (2004) report that these residues improve *in vivo* activity.[94] Additionally, we did not investigate the possibility of calcium dependence on lmSrtB activity. Class B sortases are generally thought to be calcium independent, however as very few sortases in this class have been investigated for *in vitro* activity and calcium dependence is well established in at least one sortase, this possibility merits exploration.[7,29] Additionally, lmSrtB *in vitro* activity should be tested using the alternative substrate sequence NAKTN. Although cdSrtB activity was not observable in the FRET-based assays attempted here, further study of cdSrtB could be conducted using HPLC-based methods. It will also likely be possible to modify the FRET-based assay protocol to allow for a longer incubation time between fluorescence readings, which may result in observable activity for cdSrtB. The investigations into saSrtB presented in this work are also particularly limited. Further studies of saSrtB were hindered by the relative insolubility of substrate peptides, which made purifying these peptides quite challenging. As a result, only a small amount of purified substrate could be obtained. Additionally, these peptides tended to precipitate out of solution after being stored at 4°C and could not be dissolved again. Further studies using this enzyme may improve substrate solubility by increasing the amount of DMSO used in stock solutions and allow for an increase in the residual DMSO percentage in reactions. Additionally, it may be possible to modify enzyme substrates to improve solubility.

Of the enzymes presented in this chapter, baSrtB appears to be the most active. For this reason, baSrtB was selected for further study and is the focus of chapters 4 and 5 of this work.

The substrate-bound models presented in this work may be useful tools for better understanding enzyme-substrate interactions. Co-crystallization experiments are time-consuming, cannot provide insight into enzyme-substrate interactions beyond the narrow scope of peptides used, and were unsuccessful in this work. Using AI-generated models is a

comparatively fast and cost-effective method. However, the reliability of these models requires careful scrutiny and experimental validation. Of the models obtained in this work, the model of lmSrtB bound to Lmo2186 appears to show substrate binding along the NPKSS sequence in good agreement with *in vivo* research, but seemingly at odds with the apparent inability of the enzyme to cleave this sequence *in vitro*.[94] Additionally, the substrate-bound cdSrtB model did not indicate successful substrate docking within the active site. However, the substrate-bound baSrtB and saSrtB models appear to be in good agreement with experimental data.[1,2,8,84,89,95] These models represent a useful starting point for better understanding substrate-enzyme interactions.

Very few sortase B enzymes have been experimentally characterized. Of the many putative sortase B genes identified through genetic screens, we believe only six have experimentally-derived structures and only five, including lmSrtB presented in this work, have been tested for *in vitro* activity.[1,3] There are many challenges in studying sortase B enzymes. These challenges include low *in vitro* activity, a generally high likelihood of spurious identifications in genetic screens, and lack of information about endogenous substrates.[3,8] In this work, we have presented several promising sortase B candidates for future studies, as well as a method for generating substrate-bound models that may lend further insights into how these enzymes function.

3.4: Methods

3.4.1: Protein Expression and Purification

The wild-type sequences used for preliminary assays were baSrtB$_{35\text{-}254}$ (UniProt ID A0A6L8PZR0), cdSrtB$_{26\text{-}225}$ (Uniprot ID Q183F3, N-terminus shortened by 25 residues), lmSrtB$_{26\text{-}246}$ (UniProt ID Q8Y588), and saSrtB$_{30\text{-}244}$ (UniProt ID Q2FZE3). These are referred to as baSrtB, cdSrtB, lmSrtB, and saSrtB in the text for simplicity. Protein genes were recombinantly expressed and proteins were purified as described previously (2.4.1: Expression and Purification of SpySrtA Protein).

3.4.2: Peptide Synthesis

Model peptide substrates were synthesized and purified as previously described (2.4.2: Peptide Synthesis).[20,22]

3.4.3: Substrate-Bound Models of Sortase B Enzymes

Full-length enzyme protein sequences were artificially linked to the C-terminus of substrate proteins using a glycine-serine linker sequence. The fused protein sequence was submitted to AlphaFold v.2.3.1 (Galaxy Version), an AI system that predicts protein tertiary structure.[96–98]. This system is referred to by the simplified name "AlphaFold2 (Galaxy)" in the text. Substrate protein sequences were found using an NCBI BLAST search. In the case of cdSrtB, the substrate protein was very large (>100 kDa) and a shortened version of the substrate was used to generate the model.[99] Protein sequences submitted to AlphaFold2 (Galaxy) are shown in the appendix of this work. Models were viewed using PyMOL (Shrödinger software). Figures of models and crystal structures were created using PyMOL (Shrödinger software).

3.4.4 Preliminary Fluorescence Assays

Reactions were performed in black, flat-bottom Costar 96-well plates. Reaction mixtures were 100 uL in volume and were started by the addition of enzyme. Reactions were monitored by measuring fluorescence intensity (λ_{ex} = 320 nm, λ_{em} = 420 nm) every 6-18 minutes for up to 18 hours using a Biotek Synergy H1 plate reader. All reactions were performed in triplicate unless otherwise indicated. Reaction conditions were adjusted while performing preliminary assays in response to data. Additionally, saSrtB was found to be unstable in the original Tris pH 7.5 buffer and was stored in pH 8.5 Tris buffer. SaSrtB reactions also used Tris pH 8.5 buffer. Background-subtracted fluorescence (measured in relative fluorescence units, RFU) was obtained by subtracting the RFU of negative controls from samples. RFU and background-subtracted RFU

over time was plotted using OriginPro (Learning Edition) 2023. The baSrtB activity assays shown in this chapter were performed using the first baSrtB prep (Figure A16).

10 µM baSrtB assay conditions: 10 µM sortase B, 50 µM Abz-DNPKTGDEK(Dnp)-*NH₂* peptide substrate, 5 mM NH₂OH, 1:10 dilution of 10x reaction buffer. Enzyme was prepared beforehand to 2x reaction concentration (100 µM) in a buffer containing 50 mM Tris pH 7.5, 150 mM NaCl, 1 mM TCEP. Abz-DNPKTGDEK(Dnp)-*NH₂* peptide stock contained DMSO to aid solubility; residual DMSO in the reaction was 0.2%. 10x sortase A reaction buffer was 500 mM Tris pH 7.5, 1500 mM NaCl.

50 µM baSrtB assay conditions: 50 µM sortase B, 175 µM Abz-DNPKTGDEK(Dnp)-*NH₂* peptide substrate, 5 mM NH₂OH, 1:10 dilution of 10x reaction buffer. Enzyme was prepared beforehand to 2x reaction concentration (100 µM) in a buffer containing 50 mM Tris pH 7.5, 150 mM NaCl, 1 mM TCEP. Abz-DNPKTGDEK(Dnp)-*NH₂* peptide stock contained DMSO to aid solubility; residual DMSO in the reaction was 0.7%. 10x sortase A reaction buffer was 500 mM Tris pH 7.5, 1500 mM NaCl.

BaSrtB pH assays: pH assays were performed as described above with the following alterations: reactions mixtures contained 30 µM enzyme and 55 µM peptide model substrate (Abz-DNPKTGDEK(Dnp)-*NH₂*). Enzyme was prepared beforehand to a concentration of 297 µM (the highest concentration possible) in a buffer containing 50 mM Tris pH 7.5, 150 mM NaCl, 1 mM TCEP. Reactions contained a 1:10 dilution of 10x sortase reaction buffer, which consisted of 500 mM Tris at the desired pH (7.0 - 9.0), 1500 mM NaCl. Reactions were performed in duplicate.

CdSrtB conditions: Several conditions were tested for cdSrtB. Briefly, enzyme concentrations attempted were 10, 20, 30, 35, 40, 50, 60 120 µM. Peptide concentrations attempted were 20, 35, 50, and 175 µM. Peptide stock did not contain DMSO in any percentage. Reaction buffers with and without calcium were used. A pH study was performed using reaction buffers ranging in pH from 7.0 - 9.0. Reactions utilized hydroxylamine (5 mM) as a nucleophile or water (no added nucleophile). Enzyme stock solutions were prepared in a buffer containing 50

mM Tris pH 7.5, 150 mM NaCl, 1 mM TCEP. Reactions contained 1:10 dilutions of 10x sortase reaction buffer (500 mM Tris pH 7.5, 1500 mM NaCl).

LmSrtB conditions: 50 µM sortase B, 200 µM Abz-TNPKSSDSK(Dnp)-*NH₂*, 5 mM hydroxylamine. Enzyme was prepared beforehand to a 2x concentration (100 µM) in a buffer containing 100 mM Tris pH 8.0, 300 mM NaCl. Peptide stock contained DMSO to aid solubility; residual DMSO was 1.1%.

SaSrtB with Abz-KVENPQTNAGK(Dnp)-NH₂ substrate assay conditions (n = 2): 50 µM sortase B, 175 µM Abz-KVENPQTNAGK(Dnp)-*NH₂*, 5 mM NH₂OH, 1:10 dilution of 10x saSrtB reaction buffer. Enzyme was prepared beforehand to a concentration of 95 µM in a buffer containing 50 µM Tris pH 8.5, 150 mM NaCl, 1 mM TCEP. Abz-KVENPQTNAGK(Dnp)-*NH₂* peptide stock contained DMSO to aid solubility; residue DMSO in the reaction was 2%. SaSrtB reaction buffer consisted of 500 mM Tris pH 8.5, 1500 mM NaCl.

3.4.4: HPLC-Based Assays

Reactions were started with addition of the enzyme and were monitored at 0 hr, 24 hr, and 48 hr by HPLC. Negative controls were included that substituted enzyme stock buffer (0 µM enzyme) in place of the enzyme. Reactions were monitored through RP-HPLC using a Dionex Ultimate 3000 HPLC system and a Phenomenex Kinetex 2.6 µM C18 100 Å column (100 x 2.1 mm) (aqueous [95% water, 5% MeCN, 0.1% formic acid]/MeCN [0.1% formic acid] mobile phase at 0.3 mL/min, elution gradient progressed from 10% mobile phase B to 90% mobile phase B back down to 10% mobile phase B over the course of an approximately 15 minute elution time). HPLC system was interfaced with an Advion CMS expression mass spectrometer in order to confirm substrate and product identity through electrospray ionization mass spectroscopy (ESI-MS). The N-terminal product was not observed for baSrtB until a later analysis was performed using a lower percentage organic phase elution method and a polar HPLC column for separation (Phenomenex Kinetex 2.6 µM Polar C18 100 Å column (100 x 2.1 mm) (aqueous [95% water, 5%

MeCN, 0.1% formic acid]/MeCN [0.1% formic acid] mobile phase at 0.3 mL/min, elution gradient progressed from 0% mobile phase B to 90% mobile phase B back down to 0% mobile phase B over the course of an approximately 15 minute elution time)).

CdSrtB: Reaction was 100 µL in volume and consisted of 20 µM cdSrtB, 100 µM Abz-PVPPKTGDSK(Dnp)-*NH₂*, 5 mM NH₂OH, 1:10 dilution of 10x sortase reaction buffer. CdSrtB stock solution was prepared beforehand to a concentration of 350 µM in 50 mM Tris pH 7.5, 150 mM NaCl, 1 mM TCEP buffer. Abz-PVPPKTGDSK(Dnp)-*NH₂* peptide stock did not have any DMSO; residual DMSO in reaction mixture was 0%.

BaSrtB: Reaction was 100 µL in volume and consisted of 10 µM baSrtB, 50 µM Abz-DNPKTGDEK(Dnp)-*NH₂*, 5 mM NH₂OH, 1:10 dilution of 10x sortase reaction buffer. BaSrtB stock solution was prepared beforehand to a concentration of 100 µM in 50 µM Tris pH 7.5, 150 mM NaCl, 1 mM TCEP buffer. Abz-DNPKTGDEK(Dnp)-*NH₂* stock solution contained DMSO to aid solubility; residual DMSO was 0.5%.

SaSrtB: Reaction was 100 µL in volume and consisted of 10 µM saSrtB, 50 µM Abz-KVENPQTNAGK(Dnp)-*NH₂*, 5 mM hydroxylamine, 1:10 dilution of 10x sortase reaction buffer. Reaction was incubated at room temperature for at least 20 hours before analysis.

LmSrtB: Reaction was 150 µL in volume and consisted of 50 µM lmSrtB, 200 µM Abz-TNPKSSDSK(Dnp)-*NH₂*, 5 mM hydroxylamine. LmSrtB stock was prepared beforehand to a concentration of 100 µM in a buffer containing 150 mM Tris pH 8.5, 300 mM NaCl. Peptide stock contained DMSO to aid solubility; residual DMSO was 1.1%. Reactions were quenched by mixing 45 µL of reaction mixture with 45 µL of LCMS-grade acetonitrile at 0 hr, 24 hr, and 48 hr. This caused lmSrtB to precipitate out of solution; protein precipitate removed prior to HPLC analysis by centrifugation and collection of the supernatant. Reactions were monitored using RP (0 hr and 24 hr) and polar (48 hr) C18 methods described above.

Chapter 4: The Structural and Biochemical Basis of BaSrtB Selectivity and Activity

Sophie N. Jackson, Jadon M. Blount, Justin W. Ibershof, Darren E. Lee, Kyle M. Whitham,

Kayla A. Croney, James McCarty, John M. Antos, Jeanine F. Amacher

Contributions by Thesis Author

I contributed to experiment design and selection of enzymatic mutants for study. I also contributed to sortase B protein expression and purification. I selected the sequences of peptide mimetic substrates and contributed to the synthesis, purification, and characterization of peptide substrates. I chose reaction conditions for assays and performed substrate and nucleophile selectivity assays for wild-type and mutant baSrtB (both FRET-based and HPLC) and LC-MS characterization of reactions. I contributed to N-terminal and R116A mutant FRET-based assays. I performed data analysis for the assays presented here. I generated substrate-bound loop chimera models using AlphaFold2 (Galaxy). I also performed *in silico* analyses of enzyme-substrate interactions using substrate-bound enzyme models and molecular dynamics simulations created and performed by others.

4.1: Introduction

Sortase enzymes display a high degree of specificity.[6] For sortase A, particularly saSrtA, this specificity is relatively well characterized.[1,6–8] However, the selectivity profiles of the other sortase classes are less well understood.[8] For class B sortases, selectivity studies are hindered by extremely low *in vitro* activity.[8] Specifically, baSrtB is estimated to have activity levels that are 1000-fold lower than the saSrtA pentamutant commonly used in SML.[8] However, understanding the biochemical basis for sortase selectivity is vital for SML and SML optimization. Here, we study the biochemical and structural features driving baSrtB selectivity.

The substrate-bound baSrtB model described in *Chapter 3: Literature Review and Preliminary Experiments with Sortase B Enzymes* provides important insights into enzyme-substrate interactions. We further studied these interactions by performing 1000 ns molecular dynamics simulations on the AlphaFold 2 (Galaxy) models. These models point towards several key interactions between the enzyme and substrate that we investigate here, including interactions between the P5, P2, and P3' residues of the substrate and several residues around the baSrtB binding pocket (Figure 23). We explore these interactions through a series of enzymatic mutants and alterations of peptide substrate sequences.

Here, we test baSrtB activity using the substrates Abz-**D**NPKTGD**E**K(Dnp)-*NH*$_2$, Abz-NPKTGD**E**K(Dnp)-*NH*$_{22}$, Abz-
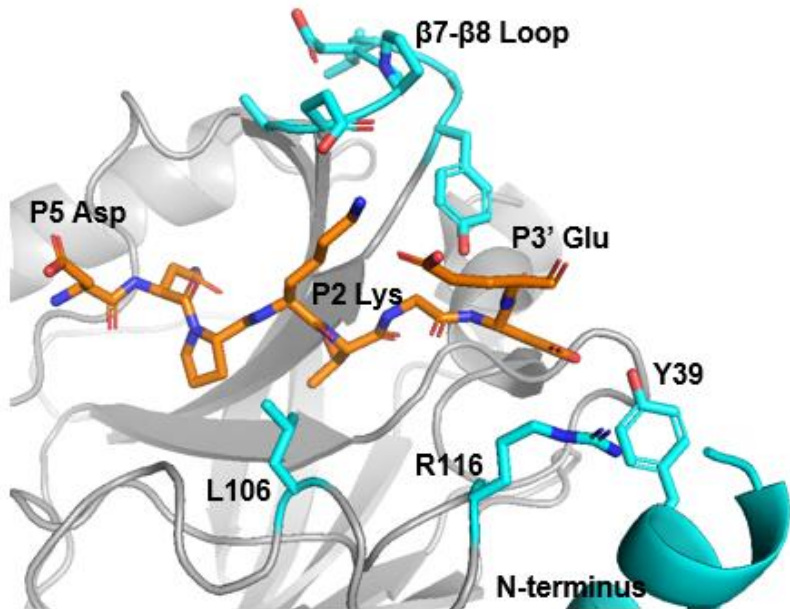

Figure 23. Investigating possible substrate-enzyme interactions. A substrate-bound baSrtB model and molecular dynamics simulations point to possible interactions between baSrtB and its substrate peptide. We investigate some of these interaction here.

NPKTGD**E**K(Dnp)-*NH₂*, and Abz-NPKTGDK(Dnp)-*NH₂* (pentapeptide motif underlined, P5 and

P3' residues bolded). Our findings show that the P5 and P3' residues improve baSrtB activity, as

has been previously reported, and that P5 and P3' each contribute to this improvement in activity.[8]

We also show that baSrtB is able to accommodate a P3' histidine residue as well as the

endogenous glutamic acid, indicating that selectivity is not absolute at this position. Additionally,

the molecular dynamics simulation indicates that an arginine residue on the β2-β3 loop, Arg116,

may interact with the P3' glutamic acid residue. We explore this possibility using an R116A

mutant.

We also wanted to study the effect of the second sortase substrate on baSrtB activity: the

nucleophile. Sortase reactions are completed when the second substrate nucleophilically attacks

the acyl-enzyme intermediate.[5] When a suitable nucleophile is present, then acylation (the original

cleavage reaction) is thought to be the rate-limiting step.[100] However, in the absence of a suitable

nucleophile, the apo-enzyme is restored through hydrolysis. In this case, hydrolysis becomes the

rate-limiting step and the reaction rate is reduced.[100] We therefore wondered if a poor nucleophile

might be a contributing factor to low baSrtB *in vitro* activity. The endogenous nucleophile for

baSrtB is thought to be *m*-diaminopimelic acid, though baSrtB may not recognize *m*-

diaminopimelic acid *in vitro* (Table 4A).[8] While baSrtB will accommodate an N-terminal glycine

nucleophile, it's unclear if this is an ideal nucleophile for this enzyme.[8] Here, we investigate baSrtB

nucleophile selectivity using the nucleophiles triglycine, Gly-*NH₂* (where -*NH₂* indicates a C-

terminal primary amide), Ala-*NH₂*, d-Ala-*NH₂*, and d-Ala, as well as the strong nucleophile

hydroxylamine (Table 4B-G). The nucleophiles d-Ala-*NH₂* and d-Ala were selected based on

structural and chemical similarity to *m*-diaminopimelic acid and Ala-*NH₂* was selected to

investigate any stereochemical preference that might be displayed by the enzyme. Glycine

nucleophiles were selected based on previously published experiments demonstrating the ability

of baSrtB to catalyze ligation reactions to the nucleophile GGGK-Biotin.[8] Hydroxylamine has

previously been shown to serve as a suitable nucleophile for several sortase homologs and was

used as the nucleophile in the sortase B-catalyzed reactions described in *Chapter 3: Literature Review and Preliminary Experiments with Sortase B Enzymes*.[6] To our surprise, reactions progressed at seemingly similar rates in a 0.9-hour activity assay regardless of nucleophile, and were only slightly elevated above the activity levels seen in a hydrolysis (no added nucleophile) control.
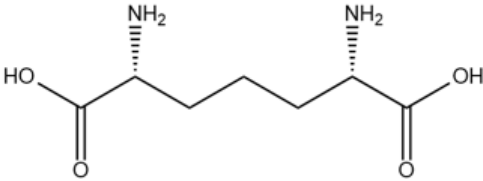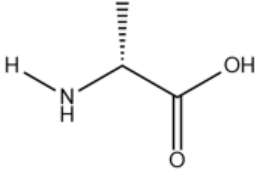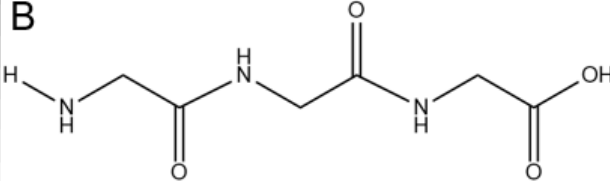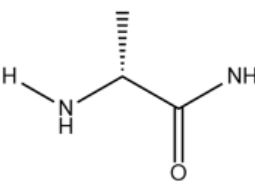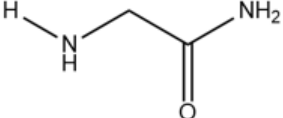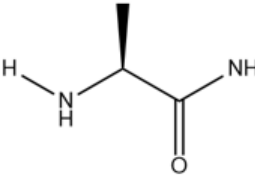
Table 4. Nucleophiles. (A) The endogenous baSrtB nucleophile is *m*-diaminopimelic acid. However, baSrtB *in vitro* activity has not been observed with *m*-diaminopimelic acid as the nucleophile. We compared baSrtB activity using the alternative nucleophiles triglycine **(B)**, Gly-NH$_2$ **(C)**, d-Ala **(E)**, d-Ala-NH$_2$ **(F)**, Ala-NH$_2$ **(G)**, and hydroxylamine **(D)**. In reactions where no suitable nucleophile is provided, acyl-enzyme intermediates are resolved by hydrolysis **(H)**.

We also sought to investigate the impact of two major structural features on baSrtB activity: the β7-β8 loop and the N-terminus. The β7-β8 loop plays an important role in sortase activity and selectivity.[20] Intriguingly, there is also notable primary sequence variability in the β7-β8 loop between sortase homologs.[20] Here, we create loop-swap chimeras by substituting the baSrtB β7-β8 loop primary sequence with that from cdSrtB, lmSrtB, and saSrtB and test the activity of these chimeras in a FRET-based assay. Chimeras created with the β7-β8 loop primary sequences from cdSrtB (baSrtB$_{difficile}$) and saSrtB (baSrtB$_{aureus}$) showed markedly less activity compared to wild-type. The chimera created with the β7-β8 loop primary sequence from lmSrtB (baSrtB$_{monocytogenes}$) showed far more activity than the wild-type. We further investigate these loop mutants by generating substrate-bound loop chimera models. We also test the effect of the lmSrtB β7-β8 loop primary sequence on baSrtB activity through a series of point mutants, each exchanging a baSrtB β7-β8 loop residue for the corresponding lmSrtB loop residue. Additionally, the impact of the β7-β8 loop primary sequence on baSrtB selectivity is investigated. Activity levels of baSrtB$_{monocytogenes}$ and the β7-β8 loop point mutants were compared to that of the wild-type using substrate sequences Abz-**D**NPKTGD**E**K(Dnp)-*NH*$_2$, Abz-NPKTGD**E**K(Dnp)- *NH*$_2$, Abz-NPKTGD**E**K(Dnp)-*NH*$_2$, and Abz-NPKTGDK(Dnp)-*NH*$_2$. We also investigate nucleophile selectivity using our most active β7-β8 loop mutant, A241K.

The class B sortases have additional N-terminal α-helices compared to class A sortases and the role of these helices in enzyme activity is unknown.[1] Our "wild-type" baSrtB construct is shortened to residue 35 to aid solubility, but retains the N-terminal helix (Figure 24). Removing hydrophobic N-terminal residues has long been part of solubilizing sortases and is typically understood to have little impact on activity.[9] The first solubilized saSrtA construct removed the N-terminal transmembrane domain up to residue 29 and a later construct removed up to residue 59.[9] The resulting mutants retained what is now considered the canonical sortase catalytic core (the β-barrel fold) and both displayed comparable *in vitro* activity.[9] Many other sortase A enzymes display similar N-terminal architecture and behavior, however a few notable exceptions have been

described.[1] In particular, the baSrtA N-terminal region interacts with the active site and is proposed to regulate enzyme activity.[1,30,101] Interestingly, a baSrtA construct with an N-terminus that was shortened by 8 residues displayed increased catalytic activity relative to the longer baSrtA protein.[30] We initially hypothesized that the baSrtB N-terminus, like that of baSrtA, may regulate activity. We therefore hypothesized that removing the N-terminal α-helix would improve baSrtB activity *in vitro*. We generated a truncated baSrtB construct without N-terminal residues up to residue 65
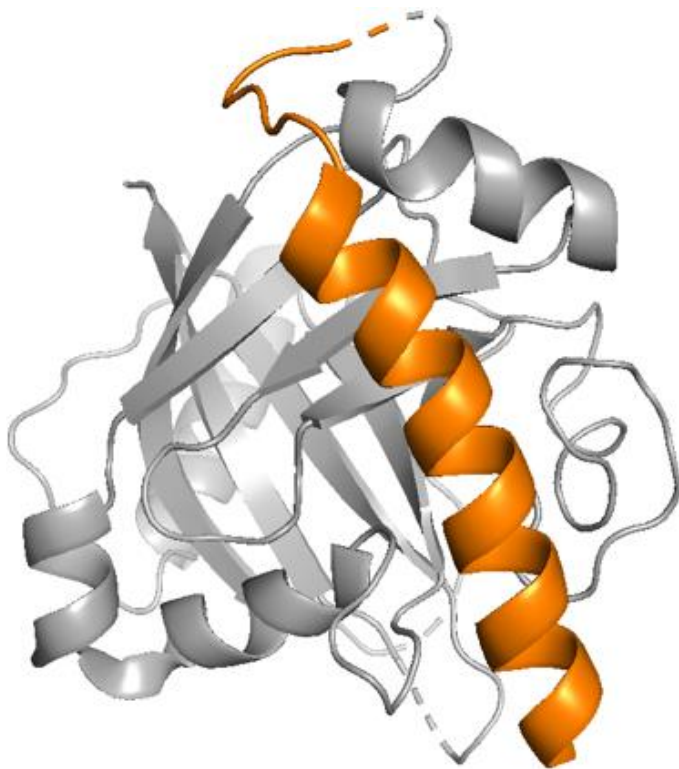
Figure 24. The baSrtB N-terminal α-helix. BaSrtB$_{35-254}$ crystal structure, shown in cartoon form, with the characteristic class B sortase N-terminal α-helix (orange) (PDB ID 2OQW).

(baSrtB$_{65-254}$). This removed the entire N-terminal α-helix (Figure 24, orange). To our surprise, baSrtB activity was markedly reduced in this truncated construct. We further explored the role of the N-terminus with an additional truncated mutant (baSrtB$_{42-254}$) and the point mutants D38A and Y39A.

Taken together, these investigations shed light on the structural and biochemical factors contributing to baSrtB selectivity and activity. We find that the β7-β8 loop primary sequence has a large impact on enzyme activity and that mutations in this region can increase baSrtB activity by more than six-fold in a 0.9-hour activity assay. We also find that the unique class B N-terminal α-helix plays an important role in baSrtB activity. An understanding of the structural and biochemical interactions that drive sortase activity is critical for continued improvement of SML techniques.

4.2: Results

4.2.1: Specificity of Wild-Type BaSrtB for Substrate P5 and P3' Residues

We performed a FRET-based baSrtB activity assay using four peptide substrates: *Abz-**D**NPKTGD**E**K(Dnp)-NH₂*, *Abz-**D**NPKTGDK(Dnp)-NH₂*, *Abz-NPKTGD**E**K(Dnp)-NH₂*, and *Abz-NPKTGDK(Dnp)- NH₂*

(pentapeptide motif underlined, P5 and P3' bolded). Results showed that baSrtB was the most active with the longest substrate sequence, in agreement with previously published results (Figure 25).[8] When normalized to baSrtB activity as measured in background-subtracted RFU value with *Abz-**D**NPKTGD**E**K(Dnp)* -
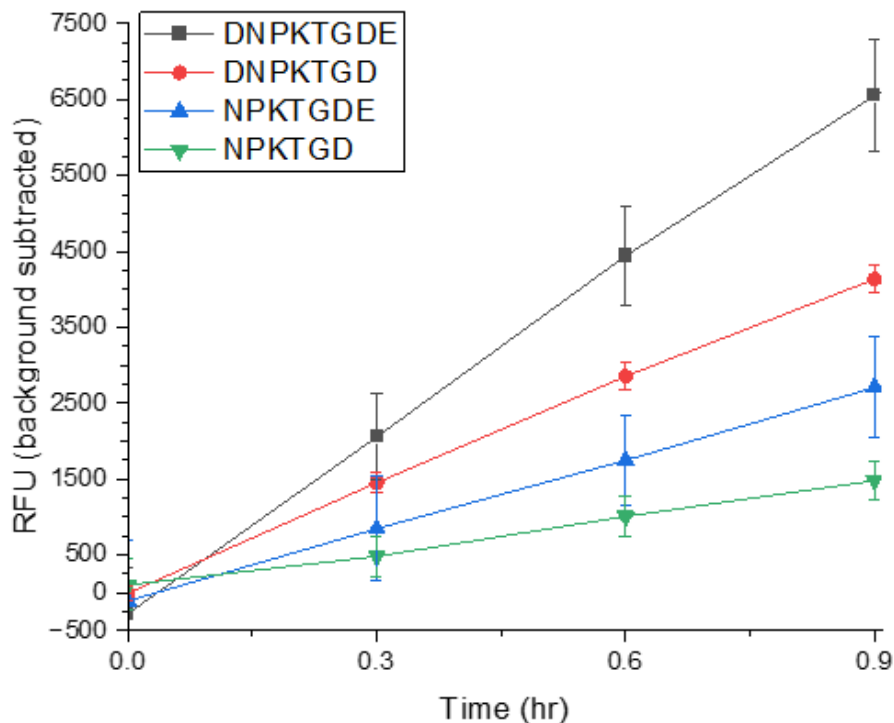
Figure 25. BaSrtB is most active when substrate peptides retain P5 and P3' residues. BaSrtB is most active with the substrate sequence *Abz*-DNPKTGDEK(Dnp)-*NH₂* (black squares) and least active with the substrate sequence *Abz*-NPKTGDK(Dnp)-*NH₂* (green triangles). It appears that baSrtB activity is more reduced by the removal of the P5 aspartic acid residue (blue triangles) and then P3' glutamic acid residue (red circles).

NH₂ peptide (1.0 ± 0.1), baSrtB activity was 0.63 ± 0.03 with *Abz-**D**NPKTGDK(Dnp)-NH₂*, 0.4 ± 0.1 with *Abz-NPKTGD**E**K(Dnp)-NH₂, and* 0.23 ± 0.04 with *Abz-NPKTGD**E**K(Dnp)-NH₂*. These results suggest that enzymatic activity is more adversely affected by the absence of the P5 aspartic acid residue than the P3' glutamic acid residue, although both contribute to activity.
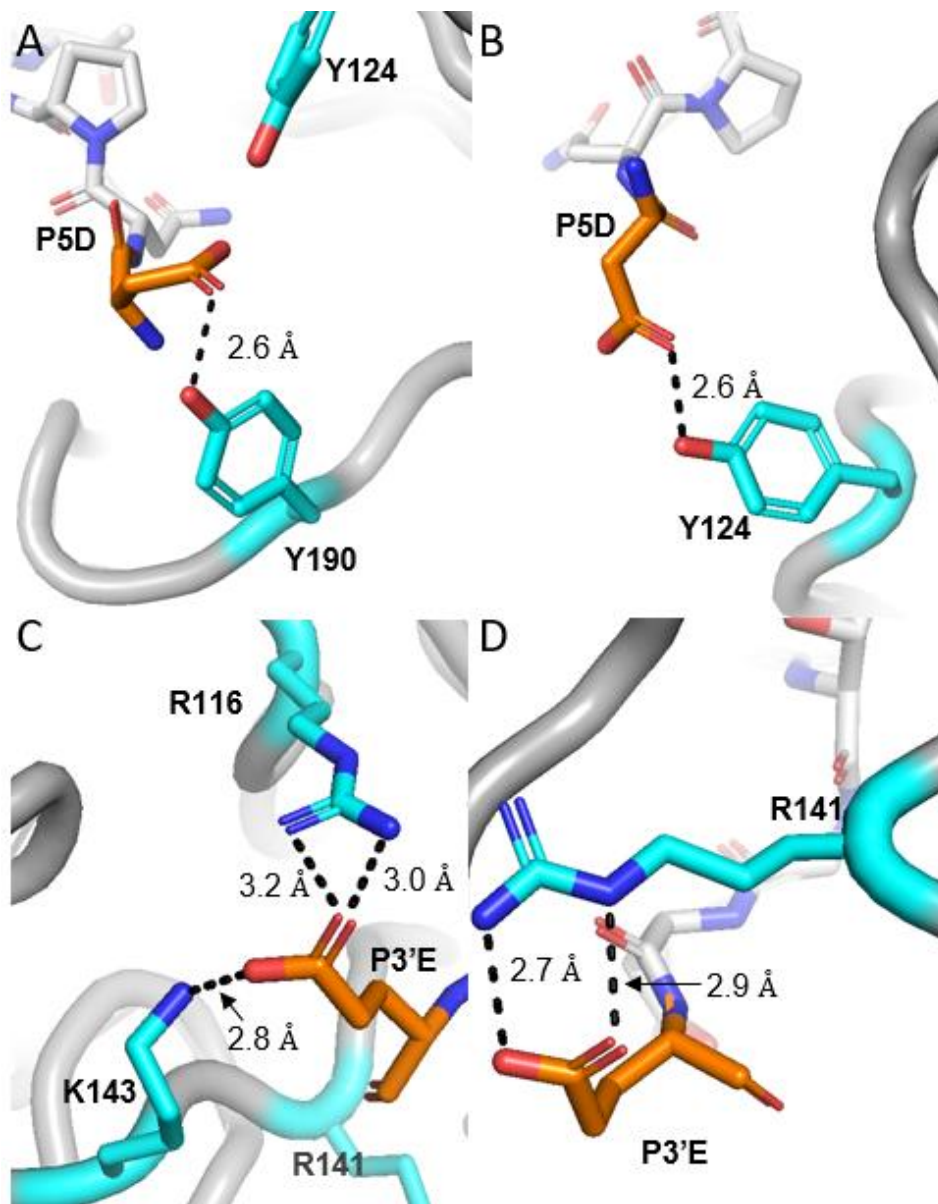
Figure 26. Substrate P5 asparagine and P3' glutamic acid form interactions with several baSrtB residues. A molecular dynamics simulation indicates that the P5 aspartic acid residue forms hydrogen bonds with tyrosine residues on the β3-β4 **(A)** and β6-β7 **(B)** loops. The P3' glutamic acid residue forms fleeting interactions with several residues, including Arg116 and Lys143 **(C)** and Arg141 **(D)**.

A 1000 ns molecular dynamics simulation suggested that the P5 aspartic acid residue forms hydrogen bonds with two tyrosine residues, Tyr124 and Tyr190 (Figure 26A-B). Tyr124 is

on the β3-β4 loop and Tyr190 is on the β6-β7 loop. These loops form walls around the sortase

binding pocket. The β6-β7 loop in particular is known to be important for sortase selectivity.[1,92]

The molecular dynamics simulation also suggested that the P3' glutamic acid forms

transient interactions with several residues, including Arg116, which is in the β2-β3 loop, Arg141

and Lys143, both of which are in the β4-β5 loop (Figure 26C-D). These loops also form part of

the sortase binding pocket.[1] These interactions are relatively fleeting and the P3' residue of the

peptide appeared to be quite mobile.

Interestingly, we also found that when the endogenous P3' Glu residue is replaced with a

histidine residue in the peptide *Abz*-D<u>NPKTG</u>D**H**K(Dnp)-*NH₂* (P3' His bolded), baSrtB activity is similar to that seen with the endogenous substrate Glu (Figure 27). This suggests that, while the P3' position is important for baSrtB activity, the enzyme is not specific for glutamic acid and may be able to accommodate other charged residues in this position.
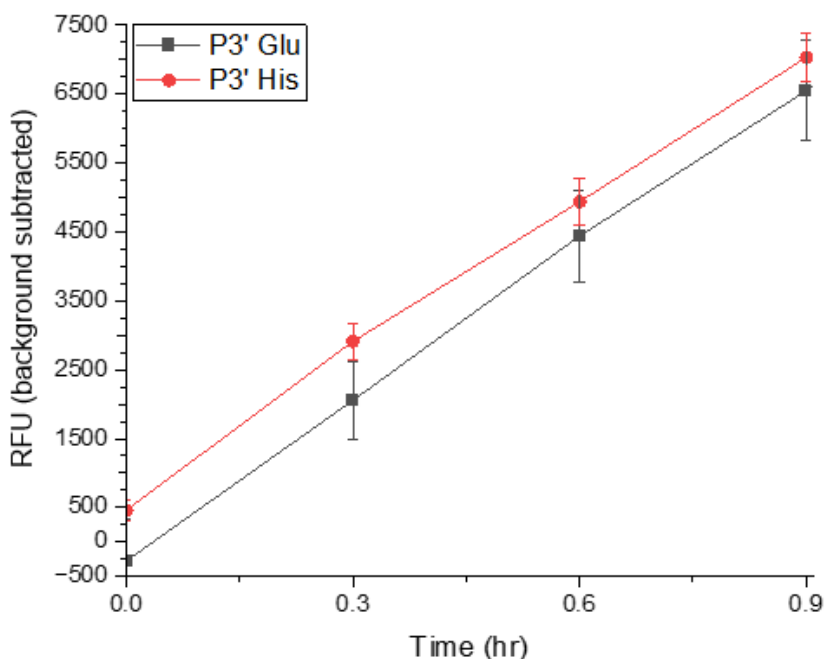
Figure 27. BaSrtB can accommodate P3' histidine residues. BaSrtB activity is similar when the P3' residue is glutamic acid (black circles) and histidine (red circles).

4.2.2: Substituting the P2 Lysine Residue for a Leucine Reduces BaSrtB Activity

Puorger et al. (2017) report that baSrtB is able to accommodate many amino acid residues

at the P2 position after a 24 hour incubation period, with no strong preference indicated.[8]

However, our substrate-bound model and molecular dynamics simulation suggested that this

residue may form hydrogen bonds with Glu240, part of the β7-β8 loop (Figure 28A). We therefore

decided to investigate the specificity of baSrtB at the P2 position by performing an activity assay with the substrate *Abz*-DNP**L**TGDEK(Dnp)-*NH₂*. Results showed that baSrtB is less active with a Leu in this position in a 0.9-hour activity assay than with the endogenous P2 Lys (Figure 28B).
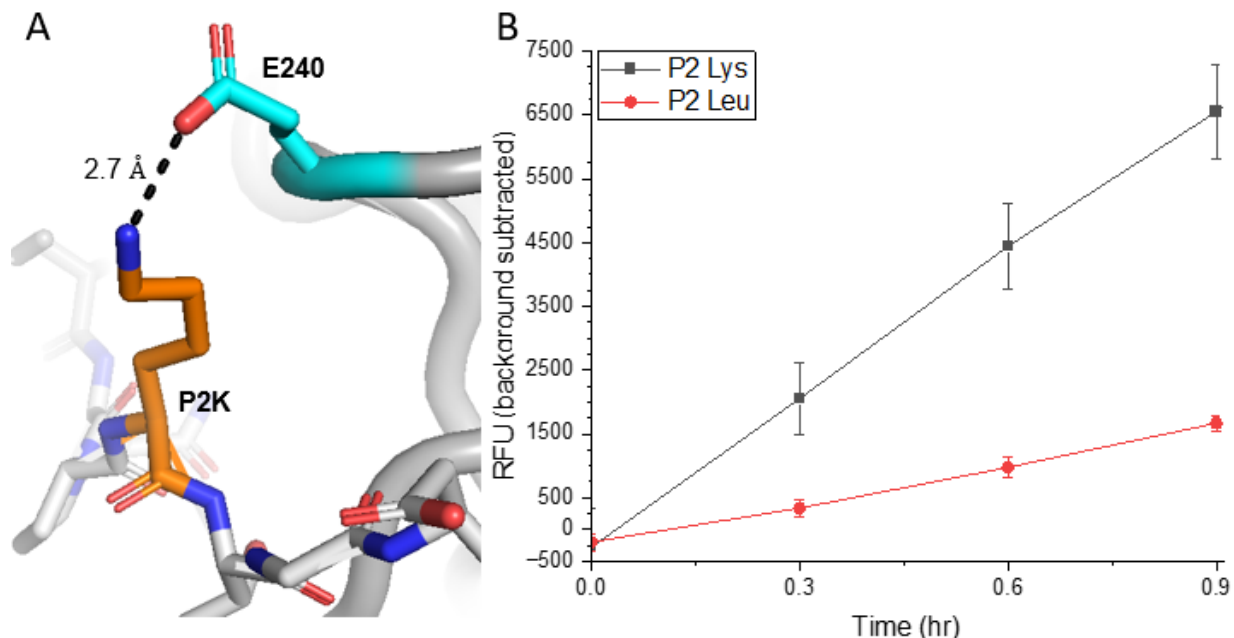
Figure 28. BaSrtB activity is reduced with P2 leucine. (A) Molecular dynamics simulations suggest that Glu240, part of the β7- β8 loop, forms stabilizing hydrogen bonds with the P2 lysine residue. **(B)** BaSrtB activity is markedly reduced when the endogenous P2 lysine residue (black squares) is replaced with a leucine residue (red circles).

4.2.3: BaSrtB is Able to Accommodate Glycine and Alanine Nucleophiles with Similar Affinity in a 0.9-hour Activity Assay

The endogenous nucleophile for baSrtB is believed to be *m*-diaminopimelic acid, However, baSrtB may not recognize this nucleophile *in vitro*.[8] Previous baSrtB ligations have been performed using a glycine nucleophile, although it's not clear if this is the most suitable nucleophile for the enzyme. We tested baSrtB activity using the nucleophiles hydroxylamine, triglycine, Gly-*NH₂,* d-Ala, d-Ala-*NH₂*, and Ala-*NH₂*. BaSrtB was able to accommodate all nucleophiles with apparently similar levels of activity in a 0.9 hr fluorescence assay (Figure 29). Interestingly, baSrtB activity progressed at only a slightly reduced rate in this time frame without

any added nucleophile. Reaction products were confirmed via LCMS after an approximately 18-24 hour incubation time (Table 5).

4.2.4: Arginine at Position 116 is Important for BaSrtB Activity

Computer modeling indicated a possible interaction between an arginine residue (Arg116) on the β2-β3 loop and the substrate P3' Glu (Figure 26C). To investigate this effect, we created an R116A mutant and tested this construct in an activity assay with *Abz*-DNPKTGD**E**K(Dnp)-*NH₂*, *Abz*-DNPKTGDK(Dnp)-*NH₂*, *Abz*-NPKTGD**E**K(Dnp)-*NH₂*, and *Abz*-NPKTGDK(Dnp)-*NH₂* (pentapeptide motif underlined, P3' Glu bolded). Activity dropped below background for all
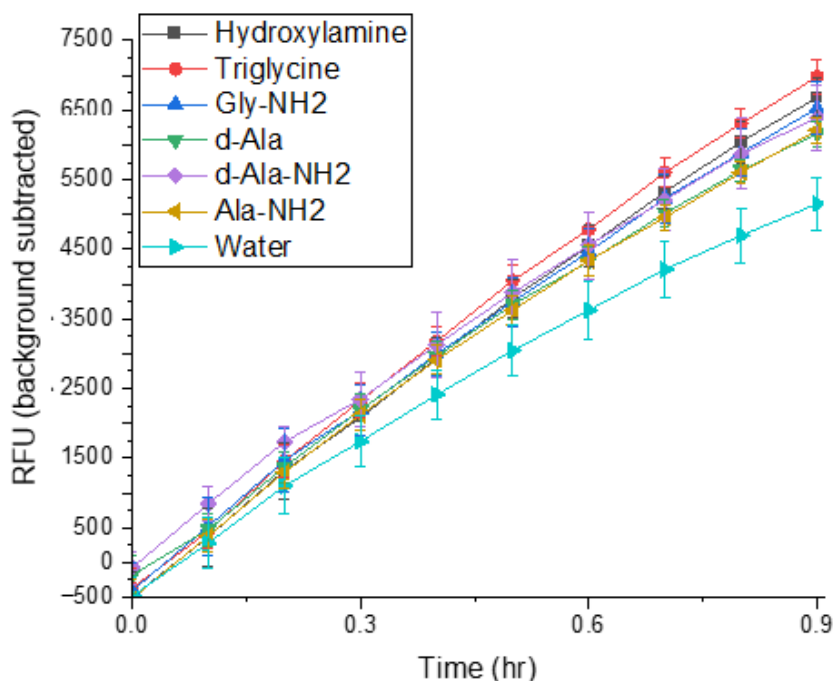
Figure 29. BaSrtB activity is similar in a 0.9 hr assay with different nucleophiles. BaSrtB activity is slightly reduced when water is the principal nucleophile available (blue triangles). Activity levels appear similar in a 0.9 hr assay when the principal nucleophile available is hydroxylamine (black squares), triglycine (red circles), Gly-*NH₂* (blue triangles), d-Ala (green triangles), d-Ala-*NH₂* (purple squares), and Ala-*NH₂* (gold triangles)

substrates that retained the P3' Glu residue and only rose minimally above background for *Abz*-NPKTGDK(Dnp)-*NH₂* (Figure 30A,C-D). However, R116A displayed measurable activity with the

substrate *Abz*-D<u>NPKTG</u>DK(Dnp)-*NH₂* in a 0.9-hour activity assay (Figure 30B) Though activity levels appear reduced relative to wild-type baSrtB in all assays, it appears that R116A performs more similarly to the wild-type construct when the P3' Glu is absent. These findings suggest that Arg116 is involved in substrate recognition at the P3' position for baSrtB. They also suggest that the R116A mutation impairs baSrtB activity, particularly when the P3' glutamic acid residue appears in substrate sequences.
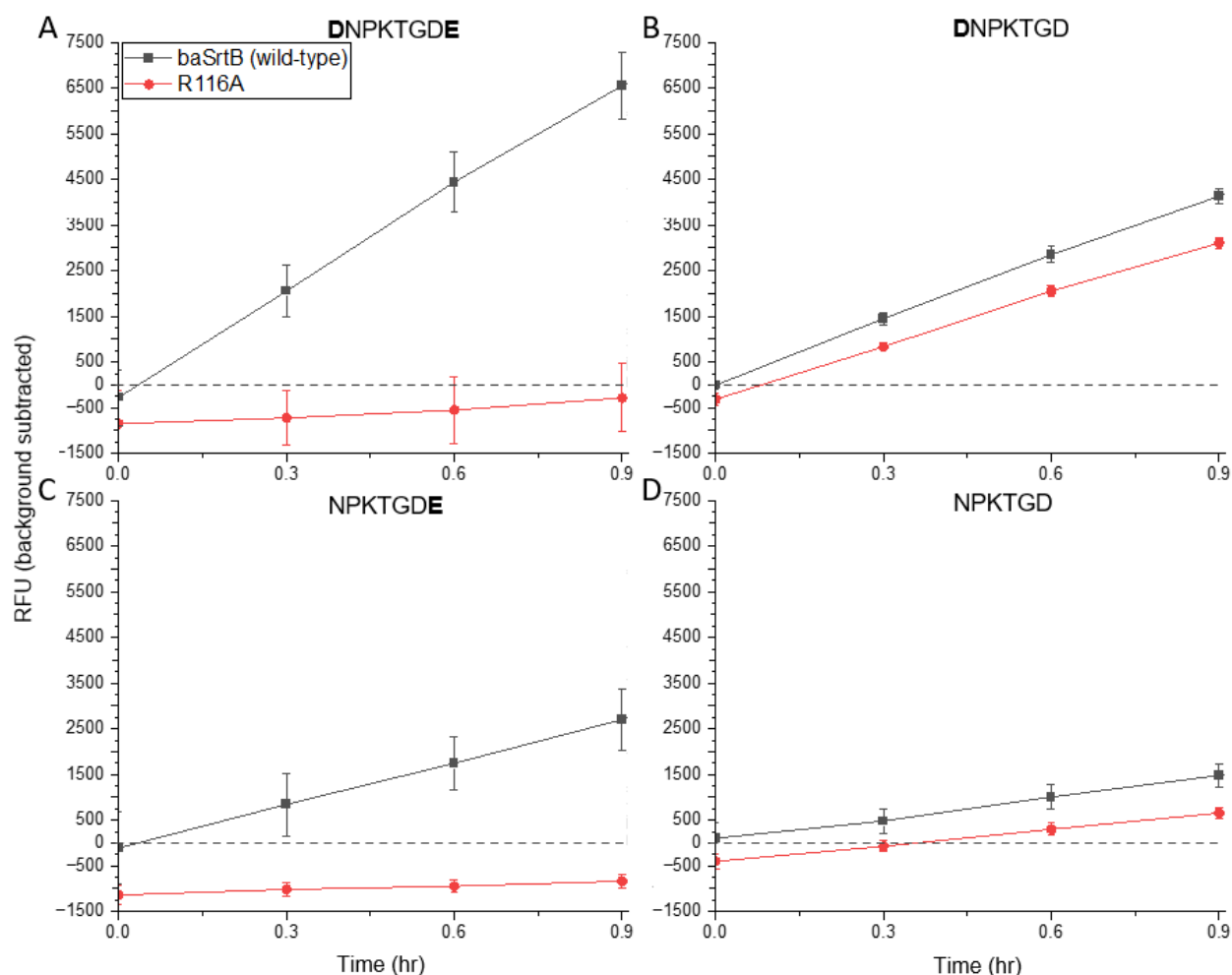
Figure 30. An R116A baSrtB mutant is less active when the P3' Glu residue is retained. BaSrtB activity is nearly abolished with substrates *Abz*-DNPKTGDEK(Dnp)-*NH₂* **(A)**, *Abz*-NPKTGDEK(Dnp)-*NH₂* **(C)**, and *Abz*-NPKTGDK(Dnp)-*NH₂* **(D)** when Arg-116 is mutated to an alanine residue (red circles). When substrate peptide is replaced with *Abz*-DNPKTGDK(Dnp)-*NH₂*, the R116A mutant displays activity that is only slightly reduced relative to the wild-type (black squares) **(B)**. Dashed lines indicate 0 RFU (background subtracted).

Molecular dynamics simulations indicated that, in addition to P3' Glu interactions, Arg116 forms hydrogen bonds with the P2' Asp residue (Figure 31A) and the N-terminal residue Asp38
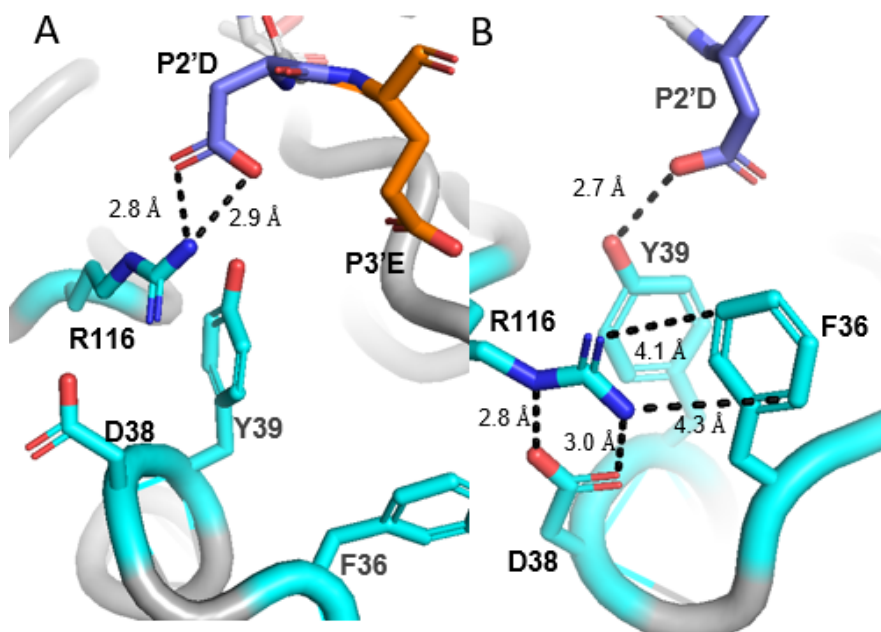
Figure 31. Arg116 interacts with substrate and enzyme residues. A molecular dynamics simulation indicates that Arg116 forms intermittent interactions with several residues, including the substrate P2' aspartic acid **(A)** and the N-terminal residues Asp38 and Phe36 **(B)**. Also visible in **(B)** is an interaction between the P2' aspartic acid residue and Tyr39, suggesting that this conformation is favorable for substrate binding.

(Figure 31B). Arg116 also forms cation-π interactions with the N-terminal Phe36 (Figure 31B). These interactions suggest that Arg116 may play a more general role in substrate binding and enzyme stabilization than merely interacting with the P3' residue, which it forms only fleeting associations with in our simulation.

4.2.5: The N-terminal α-Helix is Contributes to BaSrtB Activity

To better understand the role that the unique N-terminus plays in sortase B activity, we created a truncated baSrtB mutant (baSrtB$_{65-254}$) in which the N-terminal helix was removed (Figure 32A). Based on N-terminal truncations in class A sortases, we expected this mutation to either improve baSrtB activity or have little effect. To our surprise, baSrtB activity was noticeably
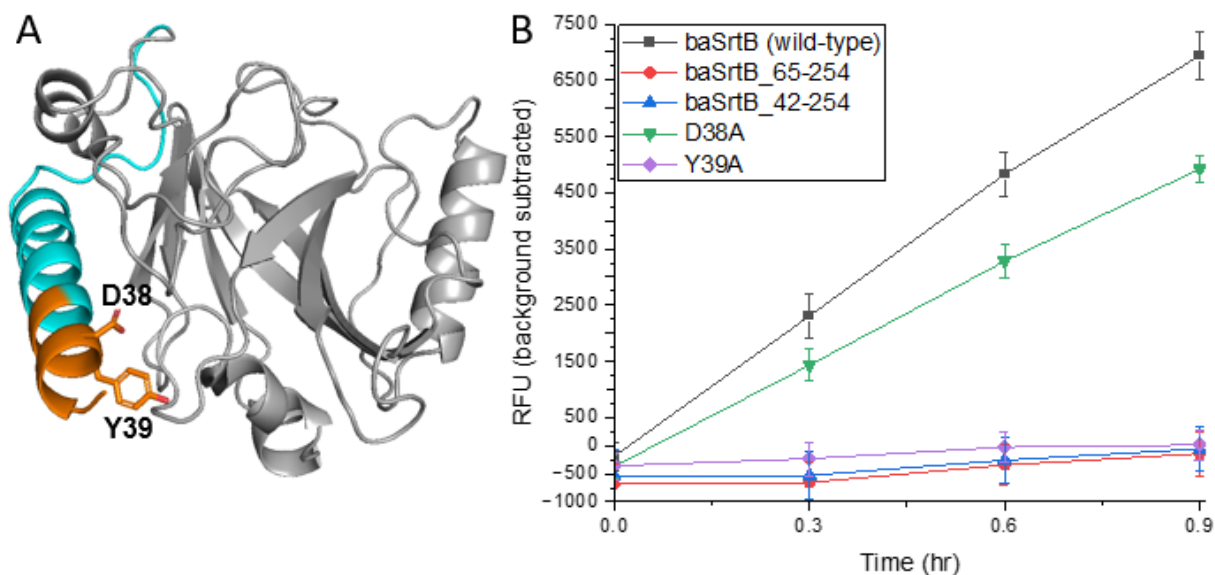
Figure 32. BaSrtB N-terminal truncations and point mutants. (A) *In vitro* activity assays were performed with the truncated baSrtB mutants baSrtB$_{65-254}$ (gray) and baSrtB$_{42-254}$ (cyan and gray). The activity of the point mutants Y39A and D38A were also investigated (labeled, side chains shown as sticks) (structure generated by AlphaFold2 (Galaxy)). **(B)** The truncated mutants baSrtB$_{42-254}$ (blue triangles) and baSrtB$_{65-254}$ (red circles) display greatly reduced activity compared to the "wild-type" baSrtB$_{35-254}$ (black squares). The point mutant Y39A displays activity at comparable levels to the truncated mutants (purple diamonds). The point mutant D38A also has reduced activity, although to a lesser extent (green triangles).

reduced relative to the wild-type (baSrtB$_{35-254}$) (Figure 32B). We created another truncated baSrtB construct (baSrtB$_{42-254}$) (Figure 32A) and again found a similar decrease in baSrtB activity (Figure 32B).

Molecular dynamics simulations pointed to a possible interaction between Asp38 and Arg116 as well as the possibility that Tyr39 participates in substrate binding (Figure 31B). Additionally, Tyr39 forms van der Waahls interactions with Asp144, a residue in the β4-β5 loop, potentially stabilizing both the N-terminus and the β4-β5 loop (Figure 33). We investigated the role these residues play in baSrtB activity with the mutants D38A
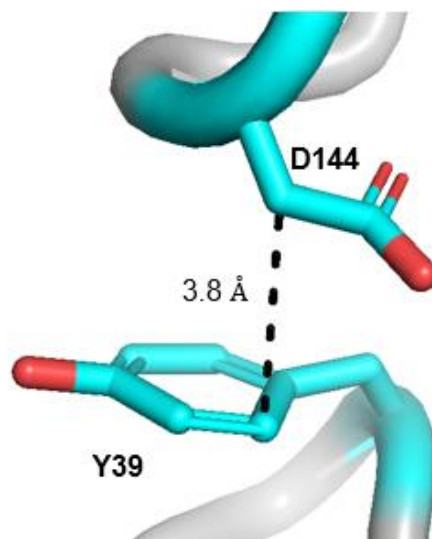
Figure 33. Tyr39 may stabilize the N-terminus and β4- β5 loop. Tyr39 appears to form van der Waals interactions with the side chain of Asp144, which may have the effect of stabilizing the N-terminus and the β4-β5 loop.

and Y39A (Figure 32A). D38A demonstrated reduced *in vitro* activity relative to the wild-type

enzyme, but improved activity relative to the truncated mutants (Figure 32B). Y39A activity levels

were comparable to those of the truncated mutants (Figure 32B). These findings suggest that

several residues of the N-terminus may be important for catalytic activity, and that Tyr39 plays a

particularly important role.

4.2.6 The β7-β8 Loop Primary Sequence Impacts BaSrtB Activity

We generated a series of baSrtB loop chimeras that replaced the baSrtB β7-β8 loop

primary sequence with that of cdSrtB (baSrtB$_{difficile}$), saSrtB (baSrtB$_{aureus}$), and lmSrtB

(baSrtB$_{monocytogenes}$). A partial sequence alignment showing the primary sequences of the β7-β8

loop is shown in Figure 34.
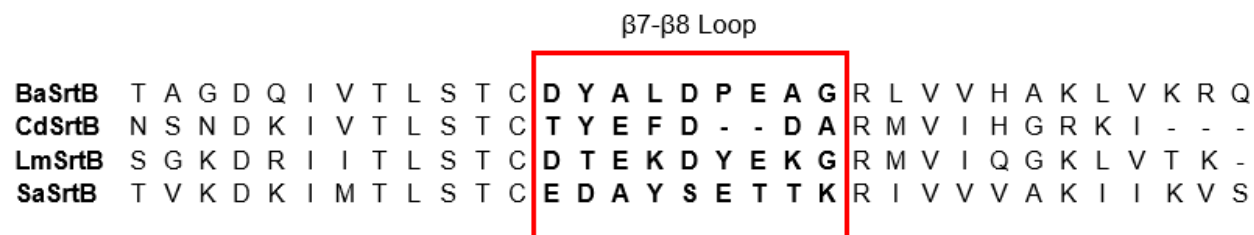
Figure 34. Primary sequence variability of the β7-β8 loop. A sequence alignment of the baSrtB, cdSrtB, lmSrtB, and saSrtB C-termini is shown. There is primary sequence variability in the β7-β8 loop (bold, boxed in red).

We found that the primary loop sequence has a dramatic impact on baSrtB activity (Figure

35). BaSrtB$_{monocytogenes}$ is far more active than wild-type baSrtB in a 0.9-hour FRET-based activity

assay and baSrtB$_{difficile}$ and baSrtB$_{aureus}$ are considerably less active. The finding that baSrtB

activity can be increased through mutations to the β7-β8 loop is particularly exciting given that

low *in vitro* activity has been among the primary obstacles for studying class B sortases and using

them in SML.

We generated IsdC-bound models of our loop chimeras using the method described in

Chapter 3: Literature Review and Preliminary Experiments with Sortase B Enzymes to gain

additional insights. The peptide sequence DNPKTGDE adopted a remarkably similar

conformation in all three models (Figure 36, model depicted with substrate shortened to the DNPKTGDE sequence). The β7-β8 loop appears markedly different, however. The baSrtB$_{monocytogenes}$ loop adopts a conformation similar to the wild-type. Conversely, the β7-β8 loop adopts very different conformations in the baSrtB$_{difficile}$ and baSrtB$_{aureus}$
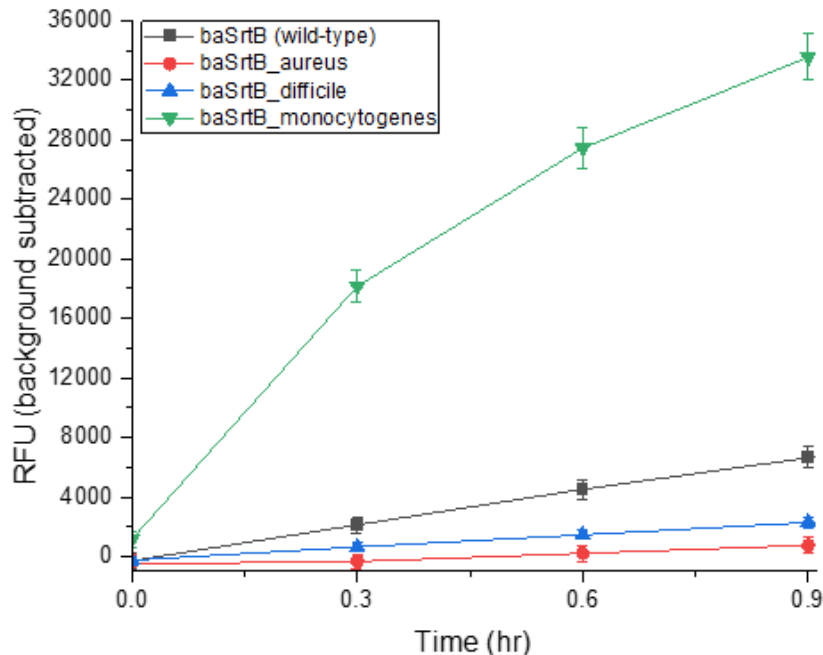
Figure 35. Sortase B β7- β8 loop chimeras display markedly different activity than wild-type baSrtB. BaSrtB$_{monocytogenes}$ shows drastically increased activity (green triangles) relative to wild-type baSrtB (black squares). Activity levels are markedly reduced in BaSrtB$_{aureus}$ (red circles) and BaSrtB$_{difficile}$ (blue triangles).

chimeras relative to the wild-type. Notably, the baSrtB$_{aureus}$ loop appears to stray the most from the baSrtB conformation and this construct is also the least active in our activity assay. These models suggest that the reduced activity of the baSrtB$_{difficile}$ and baSrtB$_{aureus}$ chimeras may be related in part to an unfavorable loop conformation.

There are other several notable features of the baSrtB$_{monocytogenes}$ chimera model (Figure 37). First, the chimera retains an interaction between the native residue Glu240 and the P2 lysine (Figure 37B). Of the three loop chimeras, only baSrtB$_{monocytogenes}$ retains this residue. Although the other two chimeras have residues that may interact with the P2 lysine, this was not indicated in the models. In addition, two mutations in the baSrtB$_{monocytogenes}$ chimera appear to form new interactions between the β7-β8 loop and the β4-β5 (Figure 36C) and β6-β7 loops (Figure 36D). These interactions may stabilize the β7-β8 loop, in a conformation that is favorable for substrate

binding and catalysis. Similarly, these interactions may stabilize the β4-β5 and β6-β7 loops in a

favorable conformation.
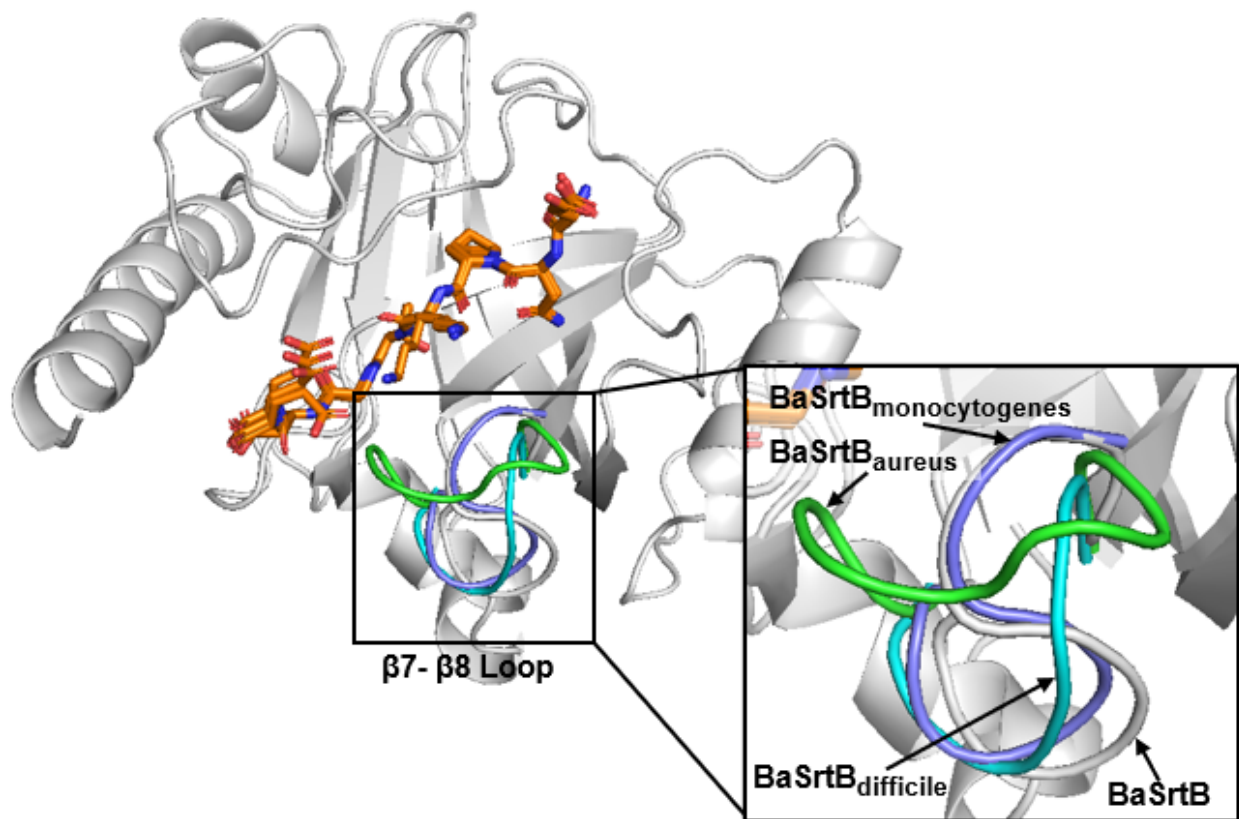
Figure 36. The β7- β8 loop primary sequence may affect loop conformation. Wild-type baSrtB shown in gray (AlphaFold2 (Galaxy) model). AlphaFold2 (Galaxy) models of substrate-bound loop chimeras show very similar substrate docking (orange, DNPKTGDE sequence for all loop chimeras and wild-type, shown in stick representation). However, the β7-β8 loop appears to adopt a different conformation in loop chimeras. The baSrtB$_{aurues}$ loop (green) appears to be the most distinct from the wild-type, while baSrtB$_{monocytogenes}$ (purple) seems to adopt a shape that is very similar to the wild-type loop.
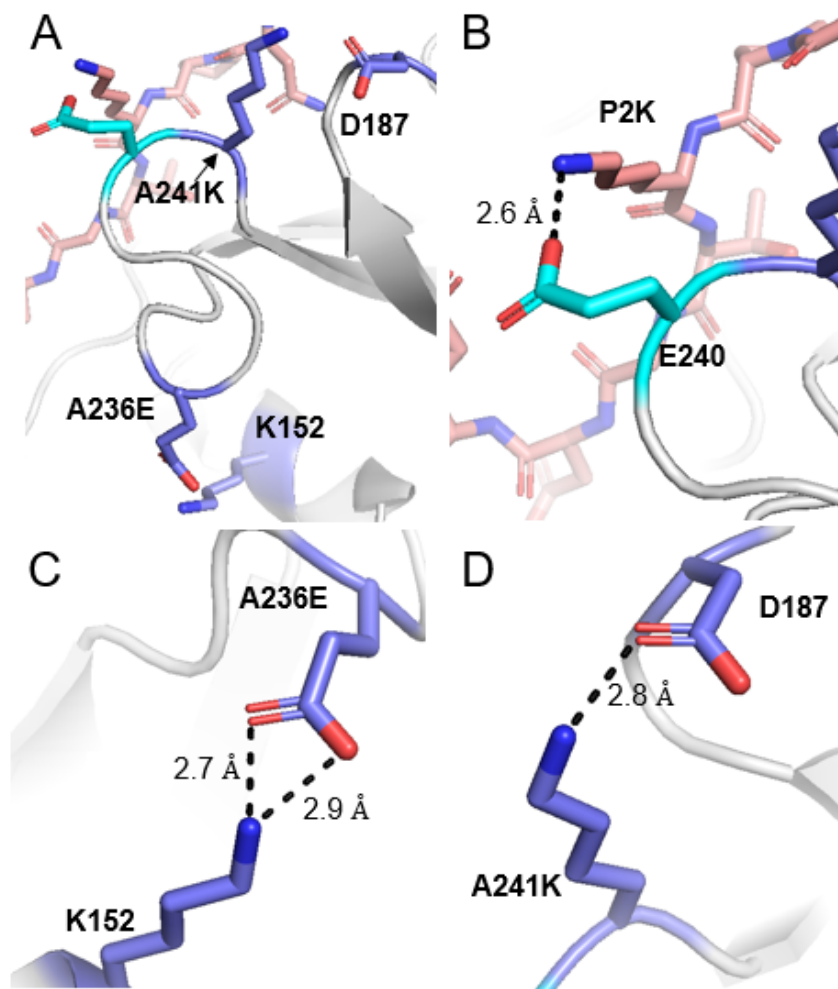
Figure 37. Some residues in the baSrtB_{monocytogenes} chimera form stabilizing interactions. (A) The baSrtB_{monocytogenes} chimera appears to be in a favorable conformation for substrate binding and catalysis (substrate depicted in pink as sticks). **(B)** The baSrtB_{monocytogenes} chimera retains the baSrtB E240 residue (cyan), which remains well-positioned to interact with the side chain of the substrate P2 lysine. A glutamic acid residue (A236E) in the baSrtB_{monocytogenes} chimera forms hydrogen bonds with Lys152 on the β4-β5 loop **(C)** and a lysine residue (A241K) forms hydrogen bonds with Asp187 on the β6-β7 loop **(D)**. These interactions may stabilize the loop in a conformation favorable for substrate binding.

4.2.7: Investigating the Impact of Individual Residues from the LmSrtB β7-β8 Loop on BaSrtB Activity

We further investigated the increase in activity associated observed in the baSrtB_{monocytogenes} chimera with a series of point mutants, each substituting a baSrtB β7-β8 loop

residue with the corresponding residue from lmSrtB (Figure 38A). We tested the activity of these point mutants in a 0.9-hour FRET-based assay with the substrate *Abz*-D<u>NPKTG</u>DEK(Dnp)-*NH₂* (Figure 38B). Two of these point mutants, Y235T and P239Y don't appear to impact activity. A236E, L237K, and A241K all increase baSrtB activity. A241K shows the largest increase in activity, and the activity of this mutant surpasses that of the chimera baSrtB_monocytogenes (Figures 39-40). These results suggest that A236E and A241K are indeed playing a role in the increase in activity seen in the baSrtB_monocytogenes, particularly

Figure 38. β7-β8 loop point mutants. (A) A partial alignment of baSrtB and lmSrtB (β7-β8 loop boxed in red, varied residues highlighted and bolded). Five baSrtB mutants were generated, each substituting a residue from the lmSrtB β7-β8 loop for the corresponding residue from the baSrtB loop. **(B)** The β7- β8 point mutant A241K baSrtB (gold triangles) displays far higher activity levels in a 0.9-hr activity assay than the wild-type baSrtB enzyme (black squares). A236E (blue triangles) and L237K (green triangles) also have markedly higher activity levels than the wild-type baSrtB enzyme. Y235T activity appears unchanged relative to wild-type (red circles) and P239Y appears to have slightly reduced activity.

A241K. Based on the position of these residues in our substrate-bound models (Figure 37A,B-D), it is unlikely that they participate directly in substrate binding. Instead, it seems most probable that these residues stabilize the β7-β8 loop and possibly other regions of the baSrtB catalytic pocket in a favorable conformation. These results also suggest that L237K is beneficial for baSrtB activity, although the reasons for this are not clear. Leu237 (wild-type) and L237K

(baSrtB_monocytogenes) appear to be solvent-exposed in our substrate-bound models. It's possible that the L237K mutation renders this conformation more favorable.

4.2.8: Selectivity Profiles of the BaSrtB_monocytogenes Chimera and β7-β8 Loop Point Mutants

We investigated the impact of our β7-β8 loop mutations on substrate specificity. The activity of baSrtB_monocytogenes and the β7-β8 loop point mutants described above was tested in FRET-based assays with the substrates *Abz*-**D**NPKTGD**E**K(Dnp)-*NH₂*, *Abz*-**D**NPKTGDK(Dnp)-*NH₂*, *Abz*-NPKTGD**E**K(Dnp)-*NH₂*, and *Abz*-NPKTGDK(Dnp)-*NH₂* (pentapeptide motif underlined,
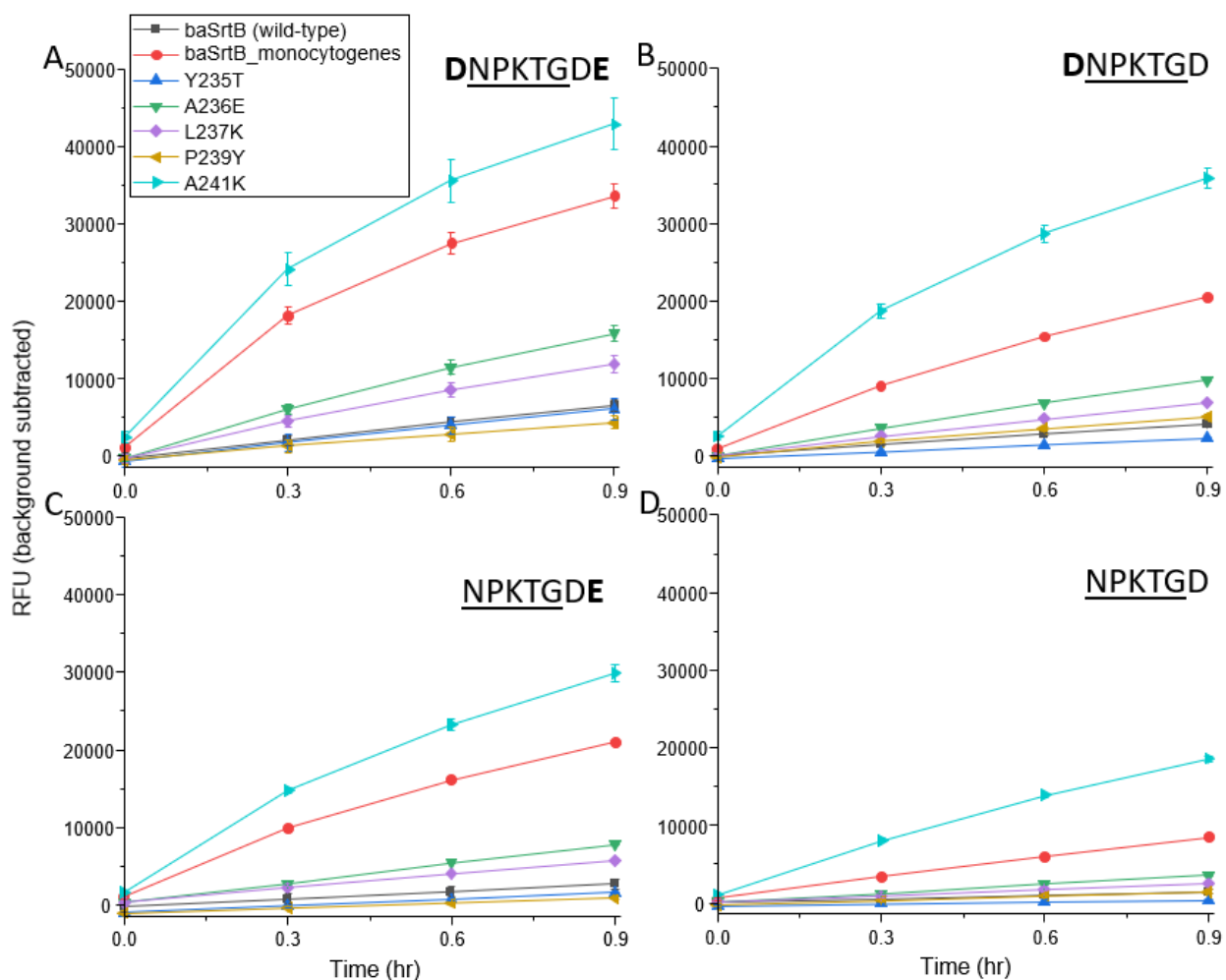
Figure 39. β7- β8 mutations don't affect baSrtB dependence on P5 and P3' substrate residues.
(A-D) The baSrtB_monocytogenes chimera and β7- β8 point mutant activity is highest with *Abz*-DNPKTGDEK(Dnp)-*NH₂* peptide and lowest with *Abz*-NPKTGDK(Dnp)-*NH₂*.

P5 and P3' bolded) (Figure 39A-D). A236E, L237K, and A241K display a similar activity pattern as the wild-type enzyme: the activity is highest with *Abz*-**D**NPKTGD**E**K(Dnp)-*NH₂*, followed by

Abz-**D**NPKTGDK(Dnp)-*NH₂*, then *Abz*-NPKTGD**E**K(Dnp)-*NH₂*, and activity is lowest with *Abz*-NPKTGDK(Dnp)-*NH₂*. Interestingly, baSrtB$_{monocytogenes}$ shows comparable activity with *Abz*-**D**NPKTGDK(Dnp)-*NH₂* and *Abz*-NPKTGD**E**K(Dnp)-*NH₂*, though otherwise activity patterns are consistent with the wild-type. Y235T activity may be more adversely affected by the absence of the P5 and P3' residues than the wild-type enzyme, although this effect was small. P239Y activity appeared to slightly improve in the absence of the P3' glutamic acid. The relative activity of each mutant is shown normalized to the wild-type for each substrate
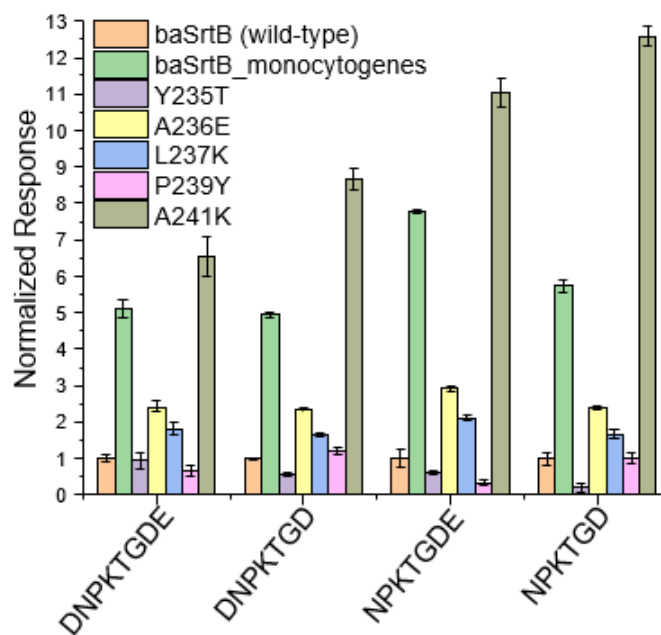
Figure 40. Relative activity of baSrtB β7- β8 loop mutants. BaSrtB mutant activity is shown normalized to the wild-type (RFU, background subtracted) with the indicated substrate at t = 0.9 hr. BaSrtB$_{monocytogenes}$ (light green), A236E (yellow), and L237K (blue) show consistent increases in activity relative to the wild-type with DNPKTGDE, DNPKTGD, and NPKTGD. Their relative activity appears to be slightly elevated with the NPKTGDE substrate, and this effect is particularly pronounced in BaSrtB$_{monocytogenes}$. Y235T (purple) and P239Y (pink) perform either comparably to the wild-type or slightly worse depending on substrate sequence. A241K (dark green, rightmost bars) shows improved activity compared to wild-type with all substrates, and the relative activity increases with decreasing substrate length.

sequence in Figure 40. A241K shows increasingly higher relative activity with the shortened substrate sequences. Although A214K activity, like wild-type activity, decreases when the P5 and/or P3' residues are absent, activity is not as adversely impacted as the wild-type. This suggests that A241K improves substrate binding within the pentapeptide motif or at the P2' position (NPKTG**D**). There is a slight improvement in relative activity for A236E and L237K with the *Abz*-NPKTGD**E**K(Dnp)-*NH₂*, suggesting that these mutations together may account for the

more dramatically improved relative activity in the baSrtB$_{monocytogenes}$ construct. Interestingly, the boost in relative activity is much more pronounced with *Abz*-NPKTGD**E**K(Dnp)-*NH$_2$* than in *Abz*-NPKTGDK(Dnp)-*NH$_2$*, suggesting that the P3' glutamic acid may be important for this effect in these mutants. Y235T and P239Y either display levels of activity that are comparable to the wild-type or slightly reduced compared to the wild-type, suggesting that these mutations are not beneficial for substrate binding and may hinder it in some cases. These data suggest that mutations in this region can have complicated effects on baSrtB selectivity for the P5 and P3' position.

4.2.9: The Identity of the Nucleophile Impacts A241K BaSrtB Activity

We decided to perform a nucleophile affinity assay using the more active A241K baSrtB mutant. The same nucleophiles were used as in the wild-type nucleophile affinity assay described in section 4.2.3. The results of this assay show that A241K, like wild-type baSrtB, is able to accommodate glycine and alanine nucleophiles (Table 5). Interestingly, A241K appears to be most active with a triglycine nucleophile
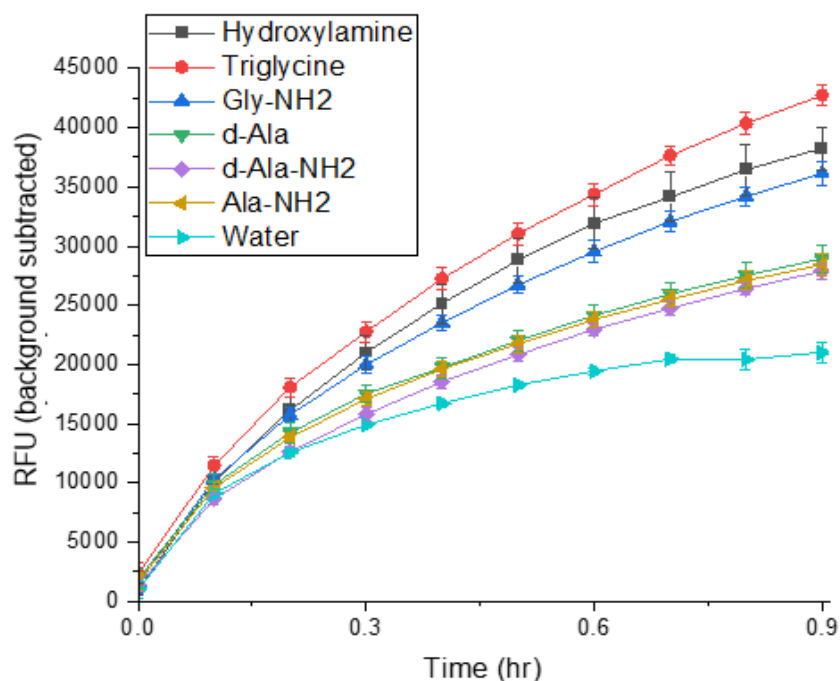
Figure 41. A241K baSrtB is most active with a triglycine nucleophile. A241K baSrtB activity is markedly reduced when water is the principal nucleophile available (blue triangles). Activity levels appear similar in a 0.9 hr assay when the principal nucleophile available is d-Ala (green triangles), d-Ala-*NH$_2$* (purple squares), and Ala-*NH$_2$* (gold triangles) and is lower than when the available nucleophile is hydroxylamine (black squares), triglycine (red circles), and Gly-*NH$_2$* (blue triangles). A241K activity is slightly improved with a triglycine nucleophile relative to hydroxylamine and Gly-*NH$_2$*.

in a 0.9 hour FRET-based activity assay (Figure 41). Activity levels are similar between the hydroxylamine and Gly-*NH₂* nucleophile reactions. A241K baSrtB appears to be similarly active with any alanine nucleophile tested, and this activity is reduced compared to that seen in the glycine and hydroxylamine reactions. A241K was least active in reactions that progressed without added nucleophile.

Nucleophile	Product	Calculated (m/z)	Wild-Type Observed (m/z)	A241K Observed (m/z)
H2O	Abz-DNPKT-*OH*	693.31	*not observed*	693.3
NH₂OH	Abz-DNPKT-*NHOH*	708.32	708.4	708.4
GGG	Abz-DNPKTGGG	864.38	864.5	864.4
G-*NH₂*	Abz-DNPKTG-*NH₂*	749.35	749.4	749.4
Ala-*NH₂*	Abz-DNPKTA-*NH₂*	763.37	763.5	763.4
d-Ala-*NH₂*	Abz-DNPKTA-*NH₂*	763.37	763.4	763.5
d-Ala	Abz-DNPKTA	764.35	764.4	764.4

Table 5. LCMS analysis of N-terminal products. N-terminal products from baSrtB (wild-type)-catalyzed reactions and A241K baSrtB-catalyzed reactions using different nucleophiles. Results show that both wild-type and A241K baSrtB can accommodate all nucleophiles tested. The hydrolysis product was not observed for wild-type baSrtB. Calculated m/z = [M + H⁺], measured in Daltons.

4.3 Discussion

We observe here a marked improvement in baSrtB activity when substrate residues are extended beyond the pentapeptide NPKTG. In a 0.9-hour FRET-based activity assay, we find that baSrtB activity improves when the native P5 aspartic acid residue and the native P3' glutamic acid residue are included in peptide substrates. We also find that this effect is additive, and that the most activity is seen when both residues are included in substrate sequences. Puorger et al. (2017) found a similar effect over a longer time period. In that experiment, baSrtB product formation rose by approximately 30% after a 24-hour incubation period when native P5 and P3' residues were included in peptide substrates (the individual contributions of P5 and P3' were not explored).[8] Interestingly, the authors also report that baSrtB could accommodate any amino acid tested at the P5 residue, although Met, Ile/Leu, Glu, Asn, Asp, and Valine resulted in the most

ligated product (selectivity for the P3' residue was not investigated). This may reflect the finding in our molecular dynamics simulations that the P5 residue interacts with tyrosine residues on the β3-β4 and β6-β7 loops, which can form hydrophobic interactions as well as hydrogen bonding. Similarly, we find that baSrtB can accommodate at P3' histidine residue as effectively as the native glutamic acid. These findings suggest that, while baSrtB activity is related to the P5 and P3' residues, it is not specific for a particular residue at these positions.

Little research has been devoted to elucidating sortase selectivity beyond the pentapeptide motif. Boyko et al. (2021) report that saSrtA catalytic efficiency is adversely affected by acidic residues N-terminal to the pentapeptide sequence.[102] In this experiment, enzyme efficiency was dramatically improved with the inclusion of a single glycine spacer in the P5 position to separate acidic residues from the pentapeptide.[102] Additionally, Pritz et al. (2007) report that saSrtA shows different levels of activity depending on the identity of the P2' residue (LPXTG**G**).[60] After 72 hours, the highest level of product formation was observed when P2' was a glycine residue and almost none was seen when P2' was a proline residue.[60] Additionally, Donahue et al. (2014) reported that cdSrtB *in vitro* activity doubled after 48 hours when peptide substrates included endogenous residues at the P2' and P3' positions rather than glycine substitutes.[13] These findings, as well as those presented in this work, suggest that SML techniques may be improved by substrate residues upstream and downstream of the canonical pentapeptide motif. They also suggest that employing native substrate residues in these positions is likely a useful strategy for studying sortase enzymes with low *in vitro* activity, as is true of the class B sortases.

Within the pentapeptide motif, we also find that the identity of the P2 residue has a large impact on baSrtB activity in a 0.9-hour FRET-based assay. Enzyme activity drops markedly when the native P2 lysine is substituted for a leucine residue. While baSrtB appears able to accommodate a leucine residue at this position, lysine is more favorable for baSrtB activity. Similarly, though Puorger et al. (2017) found no strong baSrtB amino acid preference at this position over a 24 hour-time period, the enzyme appeared to be most active with acidic P2

residues.[8] Our models suggest that this slight preference may be related to an interaction between the P2 residue and a glutamic acid (Glu240) in the β7-β8 loop.

We have also identified key enzyme structural features and residues that are important for *in vitro* activity. Arg116 contributes to baSrtB activity and may participate in substrate binding directly as well as enzyme stabilization. We also find that the N-terminus is very important for baSrtB activity. It appears that the majority of the impact of the N-terminus on baSrtB activity can be attributed to Tyr39. The exact role of this residue is unclear, although it may participate in substrate binding and enzyme stabilization. The unique class B N-terminus has not been previously investigated, and our findings with baSrtB show for the first time that this unique structural feature is important for baSrtB activity.[1] The role of the N-terminus in other class B sortase homologs should be studied, as this conserved feature may play a role in class B sortase activity in general.

The β7-β8 loop primary sequence has a large impact on baSrtB activity. We find that baSrtB activity drops when the β7-β8 loop primary sequence is replaced with that from cdSrtB and saSrtB. Conversely, baSrtB activity is improved by more than five-fold when the sequence is replaced with that from lmSrtB in a 0.9-hour activity assay. Substrate-bound chimera models suggest that this contrast may be related to conformational changes in the β7-β8 loop, with the conformation adopted by wild-type and the baSrtB_monocytogenes chimera being similar and likely most favorable for catalysis. This model also suggests that the baSrtB_monocytogenes chimera forms additional associations between the β7-β8 loop and the β4-β5 and β6-β7 loops. The β7-β8 loop point mutant A236E, which interacts with the β4-β5 loop in our model, more than doubles baSrtB activity and A241K, which interacts with the β6-β7 loop, increases activity by more than six-fold.

These findings show that baSrtB activity is very dependent on the primary sequence of the β7-β8 loop, which has been observed in other sortases. Two of the mutations in the saSrtA pentamutant described in section 1.4: SML Optimization occur in the β7-β8 loop.[20] Additionally, Piper et al. (2021) created sortase A loop-swap chimeras by substituting the primary sequence of

various sortase A β7-β8 loops onto a *Streptococcus pneumoniae* (spSrtA) scaffold.[20] The chimeras displayed altered activity and selectivity relative to the unmodified spSrtA.[20] Further investigations also showed that stabilizing interactions between the β7-β8 loop and other parts of spSrtA were important for enzyme activity.[20]

The β7-β8 loop may also be involved in substrate selectivity. The loop forms a wall of the catalytic pocket and our *in silico* model suggests that Glu240 interacts directly with the substrate P2 residue.[1] Additionally, Piper et al. (2021) report that the β7-β8 loop is particularly important for selectivity of the P1′ motif residue (e.g. LPXT<u>G</u>).[20] The most active β7-β8 loop mutants presented in this work generally did not change the wild-type baSrtB preference for substrates with included endogenous P5 and P3' residues, although they did in some cases reduce the adverse effect on activity when these residues are absent. This was particularly true for A241K, which shows a greater than 6-fold increase in activity relative to the wild-type with the substrate *Abz-***D**<u>NPKTG</u>**DE**K(Dnp)-*NH$_2$* and a more than twelve-fold increase in relative activity with the substrate *Abz-*<u>NPKTG</u>DK(Dnp)-*NH$_2$*. These findings suggest that these mutations, particularly A241K, improve substrate binding within the pentapeptide motif or at the P2' position (<u>NPKTG</u>**D**). We also investigated baSrtB nucleophile selectivity. Interestingly, baSrtB activity was comparable in a 0.9-hour assay with all tested nucleophiles, and activity was only slightly elevated above that seen when no nucleophile was added to reaction mixtures. When sortases cleave substrates, an acyl-enzyme intermediate is formed. This acyl-enzyme intermediate is resolved when the nucleophile attacks the P1 threonine carbonyl carbon, forming a new bond between the nucleophile and the P1 threonine (depicted in figure 42, *Chapter 5: The BaSrtB Catalytic Mechanism*). This restores the apo-enzyme, which is then able to catalyze a new reaction. If substrate cleavage progresses at a slow enough rate, however, the percentage of enzyme molecules in the population that have cleaved substrate to form acyl-enzyme intermediates will remain relatively low during early time points. In this circumstance, it would be possible to observe ongoing catalysis even without restoration of apo-enzyme via a nucleophilic attack. Thus, the lack

of a strong nucleophile effect in the 0.9-hour activity assay may be more indicative of the generally slow reaction rates of baSrtB rather than true nucleophile selectivity. Thus, it may be beneficial to adapt assay methods to study baSrtB activity over a longer period to assess baSrtB nucleophile selectivity. Interestingly, a similar nucleophile assay performed with A241K baSrtB did indicate nucleophile selectivity. The enzyme appears to be the most active with a triglycine nucleophile, followed by hydroxylamine and Gly-*NH*$_2$, and then the alanine nucleophiles. As expected, the lowest rates of activity were seen when no nucleophile was added to the reaction mixtures. These nucleophile selectivity patterns may have been introduced to baSrtB through the A241K mutation. The β7-β8 loop is thought to play a role in sortase nucleophile selectivity.[1] A saSrtB crystal structure bound to triglycine, a saSrtB nucleophile mimetic, was determined by Zong et al. (2004).[12] In this structure, the triglycine is held in place through interactions with a residue in the β7-β8 loop. Alternatively, the improved *in vitro* activity of A241K may simply have revealed baSrtB nucleophile selectivity patterns that were present in the wild-type enzyme but were not discernable in the 0.9-hour FRET-based assay.

The substrate modifications and enzyme mutations in this work demonstrate that baSrtB activity can be dramatically improved with relatively modest changes. The principal reason class B sortases have not been utilized for SML is their poor *in vitro* activity, a problem that plagues sortases generally. While the mutations introduced in this work do not improve baSrtB activity to a comparable level with saSrtA, we find that baSrtB activity can be markedly increased by including P5 and P3' residues in substrates and by modifying the β7-β8 loop. This finding suggests that further study of this and other class B sortases may lead to improved enzymes that can be more readily used for SML. Additionally, the findings presented in this work provide insight into class B sortase activity as well as general sortase activity and selectivity. These insights may prove generalizable to other sortases and may yield additional techniques that can be used to improve SML.

4.4 Materials & Methods

4.4.1: Expression and Purification of BaSrtB Protein

Genes were recombinantly expressed as described previously (2.4.1: Expression and Purification of SpySrtA Protein). Proteins were purified as described, with the following buffer modifications: wash buffer: 50 mM Tris pH 8.0, 150 mM NaCl, 20 mM imidazole pH 8.0 1 mM tris(2-carboxyethyl)phosphine (TCEP); elution buffer: 50 mM Tris pH 8.0, 150 mM NaCl, 300 mM imidazole pH 8.0, 1 mM TCEP; running buffer: 100 mM Tris pH 8.0, 300 mM NaCl. Assays presented in this chapter were performed with baSrtB first, second, and third preparations (Figure A16) and A241K baSrtB first and second preparations (Figure A17).

4.4.2: Peptide Synthesis

Model peptide substrates were synthesized and purified as previously described (2.4.2: Peptide Synthesis).[20,22]

4.4.3: FRET-Based Assay for Sortase Activity

Reactions were performed in black, flat-bottom Costar 96-well plates. Reaction mixtures were 100 uL in volume and consisted of 50 µM sortase B, 200 µM peptide model substrate, 5 mM hydroxylamine (nucleophile assays: 5 mM of indicated nucleophile or pure water), 50 mM Tris-HCl pH 8.0, 150 mM NaCl. Reactions were started by the addition of enzyme, which was prepared beforehand at 2x reaction concentration (100 µM) in a 2x reaction buffer (100 mM Tris pH 8.0, 150 mM NaCl). Peptide stocks often contained DMSO to aid solubility. Residual DMSO in reaction mixtures was ≤1.5%. DMSO concentrations were found to not adversely impact baSrtB activity at or below 2%, although some srtB inhibition was seen at 5% DMSO (Figure A13). Reactions were monitored by measuring fluorescence intensity (λ_{ex} = 320 nm, λ_{em} = 420 nm) every 6-18 minutes for up to 18 hours using a Biotek Synergy H1 plate reader. Data was not analyzed beyond 0.9-hours to eliminate any possibility of evaporation-driven effects (see appendix, "Fluorescence

suppression effects in FRET-based assays"). All reactions were performed at least in triplicate. Background-subtracted fluorescence (measured in relative fluorescence units, RFU) was obtained by subtracting the RFU of negative controls from samples. Background-subtracted RFU over time was plotted using OriginPro (Learning Edition) 2023.

4.4.5: LCMS characterization of reaction ligation products in nucleophile assays

Reaction solutions were collected from the 96-well plate after approximately 20 hours of incubation time. Protein was removed using Amicon Ultra 3K Centrifugal Filters and flow throw was analyzed by LCMS. Separation was achieved using a Dionex Ultimate 3000 HPLC system and a Phenomenex Kinetex 2.6 µM Polar C18 100 Å column (100 x 2.1 mm) (aqueous [95% water, 5% MeCN, 0.1% formic acid]/MeCN [0.1% formic acid] mobile phase at 0.3 mL/min, elution gradient progressed from 0% mobile phase B to 90% mobile phase B back down to 0% mobile phase B over the course of an approximately 15 minute elution time). The HPLC system was interfaced with an Advion CMS expression mass spectrometer utilizing electrospray ionization mass spectroscopy (ESI-MS).

4.4.6: Multiple Sequence Alignment

Multiple sequence alignments of sortase B proteins (baSrtB, cdSrtB, lmSrtB, and saSrtB) was performed using Uniprot Align. Full protein sequence alignment is shown in Figure A14.

4.4.7: Molecular Dynamics Simulations

Molecular dynamics simulations were performed using the substrate-bound baSrtB model described previously (*Chapter 3: Literature Review and Preliminary Experiments with Sortase B Enzymes*), with substrate sequence clipped to the DNPKTGDE sequence and the baSrtB N-terminus clipped to residue 35. The molecular dynamics simulation was performed using GROMACS 2022.4 with the AMBER99SB-ILDN protein, nucleic AMBER94 force fields.[53,68,69,]

[75,76,78,103–108] The system was solvated using spc216.gro, which is a generic equilibrated 3-point solvent model built into GROMACS and a TIP 3 point water model (TIP3P) within a cubic box with periodic boundary conditions. Ions were added to a 0.15M physiological ion concentration with a neutral net charge balanced with sodium cations and chloride ions. The steepest descent energy minimization was performed on the solvated system with a maximum force tolerance of 1000 kJ/mol/nm for all structures of the course of 100 picoseconds. Long-range electrostatic interactions were treated with the particle mesh Ewald (PME) algorithm and a 1.0 nm cutoff for Coulombic and Lennard-Jones interactions. All systems were equilibrated in an NVT ensemble for 100 picoseconds at a reference temperature of 300K with position restraints on all protein heavy atoms with velocities assigned from Maxwell distribution and V-rescale temperature coupling using modified Berendsen thermostat before moving on to NPT equilibration. All systems were equilibrated in an NPT ensemble for 5 nanoseconds at a reference pressure of 1.0 bar including constrained hydrogen bonds without position restraints with Isotropic Parrinello-Rahman pressure coupling. Sufficient equilibration was verified by protein RMSD and density analysis. All systems were simulated in an NVT ensemble for 1000 nanoseconds at a reference temperature of 300K. Molecular dynamics simulations were viewed using PyMOL (Shrödinger software), and related figures were generated using the same.

Chapter 5: The BaSrtB Catalytic Mechanism

Sophie N. Jackson, Justin W. Ibershof, Jadon M. Blount, Kayla A. Croney, Darren E. Lee, Kyle

M. Whitham, James McCarty, John M. Antos, Jeanine F. Amacher

Contributions by Thesis Author

I contributed to experiment design and selection of enzymatic mutants for study. I also contributed to sortase B protein expression and purification. I contributed to the synthesis, purification, and characterization of peptide substrates. I contributed to assays (both FRET-based and HPLC) and LC-MS characterization sortase B-catalyzed reactions. I performed data analysis for the assays presented here. I also performed LC-MS/MS analysis of purified proteins.

5.1: Introduction

The sortase catalytic mechanism is believed to be broadly conserved across sortase classes (Figure 42).[1] Sortase catalysis is dependent upon a catalytic triad consisting of histidine, cysteine, and arginine. These residues are found in all experimentally characterized sortases,
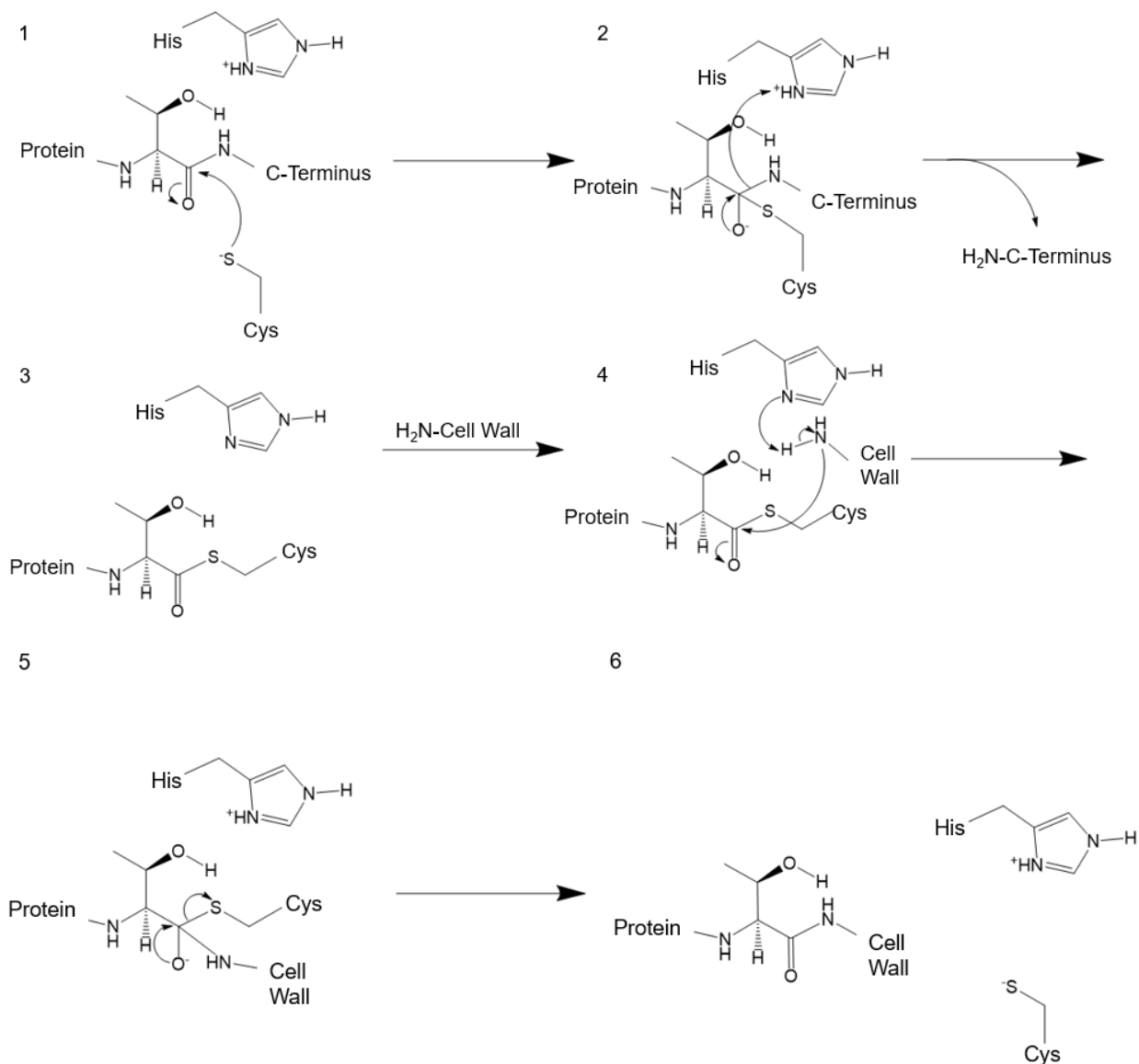
Figure 42. The sortase catalytic mechanism. Catalysis begins when the sortase catalytic cysteine nucleophilically attacks the threonine carbonyl carbon of the substrate protein (1). A tetrahedral intermediate forms and is resolved when the catalytic histidine protonates the leaving group (2). An acyl-enzyme intermediate forms (3). The nucleophilic substrate, drawn as a component of the cell wall, is activated by the catalytic histidine and attacks the acyl-enzyme intermediate (4). A second tetrahedral intermediate is formed (5), which quickly collapses, and the ligated product is released (6).

including baSrtB.[1–3] Substituting any one of these residues with alanine results in near total abrogation of enzyme activity, with a few notable exceptions.[2,22,86] Intriguingly, Kang et al. (2020) found that mutating the proposed catalytic histidine to alanine in cdSrtB had no effect on catalysis.[86] Jacobitz et al. (2014) performed a similar experiment in saSrtB and found that catalytic activity was abolished in the histidine-to-alanine mutant, as is more typical in sortase enzymes.[2] In order to better understand the role of the catalytic triad in baSrtB, we performed *in vitro* activity assays of a histidine-to-alanine (H140A) mutant and an arginine-to-alanine mutant (R243A). We find that Arg243 is vital for barSrtB catalysis. The H140A investigation was inconclusive and is ongoing. Additionally, saSrtA shows marked calcium dependence and Puorger et al. (2017) report that baSrtB activity improves when calcium is present.[8] We investigate the effect of Ca^{2+} on baSrtB activity in this work.

We also performed a multiple sequence alignment with baSrtB, saSrtB, lmSrtB, and cdSrtB. This revealed several conserved residues. We investigated two of these, a serine located two positions N-terminal to the catalytic cysteine (Ser231 using baSrtB numbering) and a leucine in the β2-β3 loop (Leu106 using baSrtB numbering). Like our investigation into His140, the Ser231 investigation was inconclusive. We find, however, that Leu106 is very important for baSrtB activity and that catalysis is virtually eliminated in an L106A mutant.

BaSrtB cleaves the peptide bond between the P1 threonine and P1' glycine. Like many sortases, baSrtB can recognize an N-terminal glycine residue as a suitable nucleophile (*Chapter 4: The Structural and Biochemical Basis of BaSrtB Selectivity and Activity)*.[8] This means that the leaving group from the original cleavage reaction can serve as the substrate for the ligation reaction, effectively reforming the starting substrate. The reversible nature of sortase-catalyzed reactions is a well-known issue in SML.[7] Previous work has demonstrated that this can be avoided when the leaving group contains a histidine residue in the P3' position in the presence of nickel cations.[109] This results from the nickel cations forming complexes with the histidine residues, rendering the leaving group an unsuitable nucleophile. We investigated the impact of a reverse

reaction on baSrtB activity by synthesizing a peptide substrate mimetic with a P3' histidine and testing baSrtB activity with and without nickel cations present.

5.2: Results

5.2.1: BaSrtB Activity is not Improved with the Presence of Ca^{2+}

Though sortase B enzymes are considered calcium-independent, Puorger et al. (2017) report that baSrtB product formation is approximately doubled when bivalent ions such as Ca^{2+} and Mg^{2+} are present.[8] To investigate this effect, we conducted a 0.9-hour FRET-based activity assay comparing baSrtB activity in the presence and absence of 5 mM $CaCl_2$ (Figure 43). We find that the presence of calcium ions does not impact baSrtB activity under these conditions. $CaCl_2$ was not included in our other baSrtB activity assays.
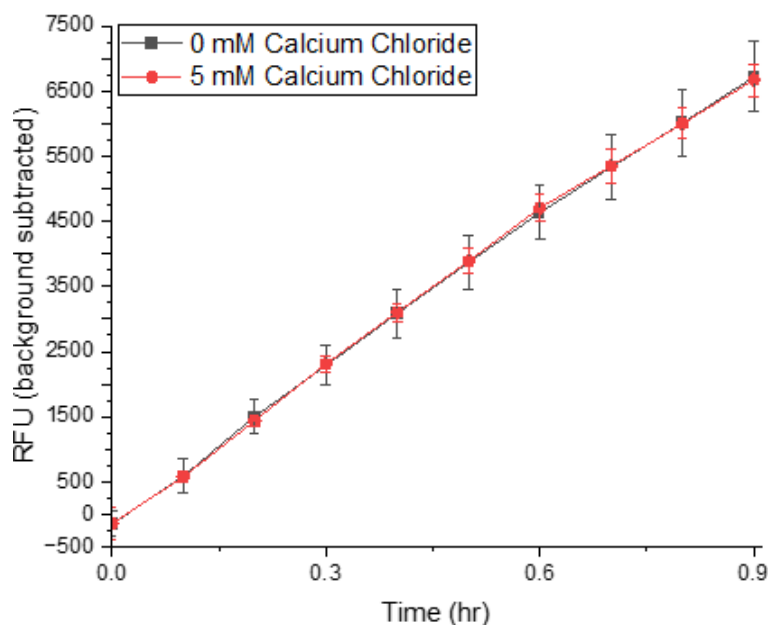
Figure 43. BaSrtB activity is unchanged by the addition of calcium. BaSrtB reactions appears to progress similarly when calcium cations are included in the reaction mix (5 mM $CaCl_2$, red circles) compared to when calcium is absent (black squares).

5.2.2: Catalytic Triad Residue R243 is Necessary for BaSrtB activity

To our knowledge, the importance of the presumed catalytic arginine (Arg243) has not yet been investigated in baSrtB. This residue, along with cysteine (Cys233) and histidine (His140, discussed in 5.2.3 below), form the sortase "catalytic triad." We investigated the importance of this residue to enzyme activity by performing activity assays with an R243A mutant. R243A was

99

inactive (Figure 44A). A molecular dynamics simulation indicates that R243 interacts with the carbonyl carbons of the substrate P4 Asn and P3 Pro residues (Figure 44B).
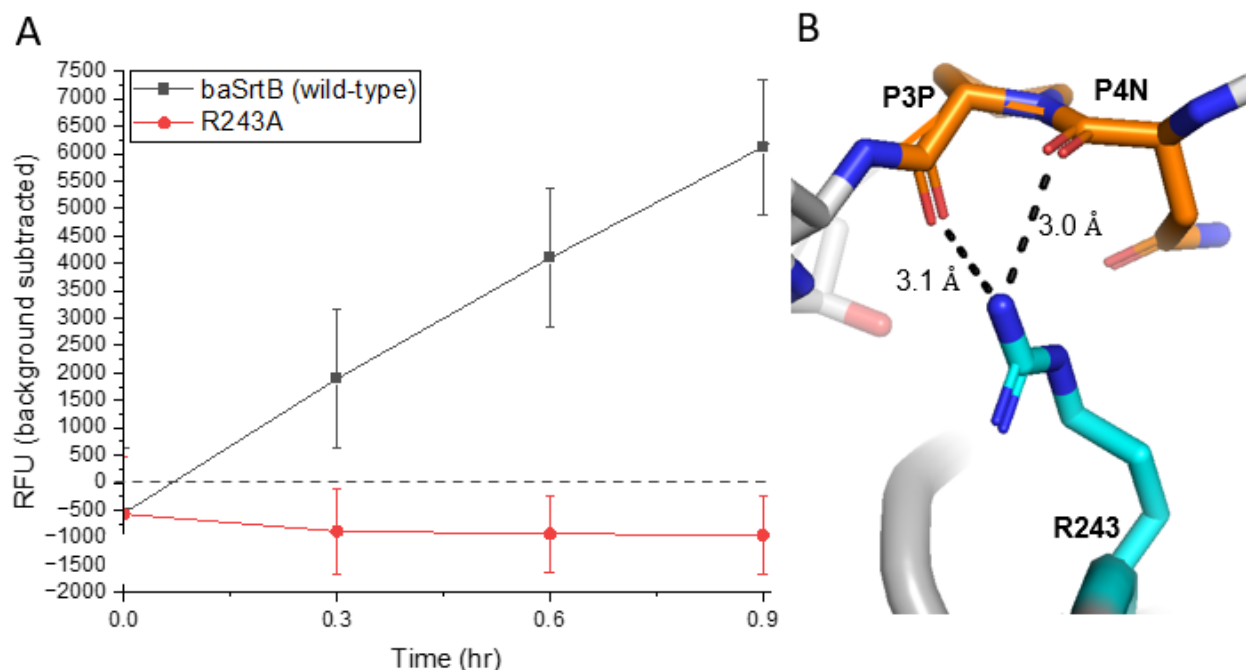
Figure 44. Arg243 is important for baSrtB activity. (A) BaSrtB activity is abolished in an R243A mutant (red circles). The dashed line indicates 0 RFU (background subtracted). **(B)** A molecular dynamics simulation indicates the Arg243 (cyan, side chain shown in stick representation) stabilizes the substrate within the catalytic pocket by interacting with the carbonyl carbons of the P4 Asn and P3 Pro residues (orange, main and side chains shown in stick representation).

5.2.3: H140A May be Necessary for BaSrtB Catalysis

To our knowledge, the role of His140 in baSrtB activity has also not been tested. We investigated the importance of this residue by generating an H140A mutant. However, H140A was not stable enough to be concentrated beyond 33 µM and had to be tested at a final concentration of 16 µM in an activity assay. In the assay (n = 1), this construct was found to be inactive compared to 16 µM wild-type baSrtB (data not shown). However, subsequent analysis of our H140A mutant showed that its mass was 31.10 Da higher than the expected molecular weight (observed mass = 28124.44 Da, calculated mass = 28093.34 Da). This mass shift could potentially indicate oxidation of the catalytic cysteine to a sulfinic acid derivative, which would also

be expected to eliminate baSrtB activity. We were unable to chemically test for the presence of a reduced cysteine due to the instability of the protein, which did not remain in solution after storage at -80° C and subsequent thawing. The role of His140 will be the subject of future work.

5.2.4: A Conserved Serine Residue May be Important for BaSrtB Activity

A multiple sequence alignment baSrtB, cdSrtB, saSrtB, and lmSrtB, indicated several conserved residues (Figure A14). We decided to investigate the impact of one of these residues, Ser231, on baSrtB activity. However, like H140A, S231A was not stable enough to be concentrated to the normal baSrtB stock concentration of 100 μM and had to be tested at a final concentration of 40 μM in an activity assay. In the assay (n = 1), this construct was found to be inactive compared to 40 μM wild-type baSrtB (data not shown). However, subsequent analysis of our S231A mutant showed that its mass was 31.34 Da higher than the expected molecular weight (observed mass = 28174.74 Da, calculated mass = 28143.40 Da). As with H140A, this mass shift could be explained by oxidation of the catalytic cysteine, which would also be expected to eliminate baSrtB activity. We were again unable to chemically test for the presence of a reduced cysteine due to the instability of the protein, which did not remain in solution after storage at -80° C and subsequent thawing. The role of Ser231 will be the subject of future work.

5.2.5: Conserved Leucine Residue is Important for BaSrtB Activity

We next turned our attention to a conserved leucine residue, Leu106 (baSrtB numbering). We tested the activity of an L106A baSrtB mutant, and found that activity was virtually abolished in this construct (Figure 45A). A molecular dynamics simulation indicated that this residue may stabilize the substrate within the catalytic pocket by forming hydrophobic interactions with the side chains of the P3 Pro and P1 Thr residues (Figure 45B).
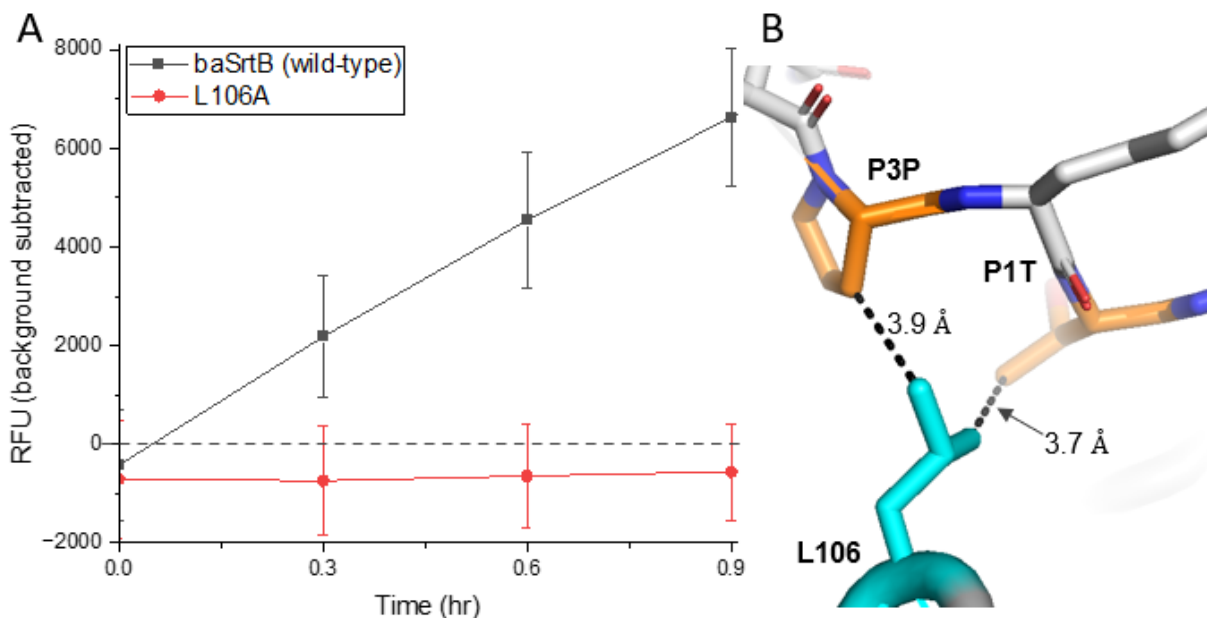
Figure 45. BaSrtB activity is largely dependent on Leu106. BaSrtB activity is nearly abolished in a 0.9-hour activity assay when Leu106 is mutated to alanine **(A)**. The dashed line indicates 0 RFU (background subtracted). **(B)** Molecular dynamics simulations suggest that Leu106 forms stabilizing interactions with the side chains of the P3 proline and P1 threonine residues.

3.2.8: Role of the Reverse Reaction in BaSrtB Activity

Interestingly, baSrtB activity with the substrate *Abz*-D<u>NPKTG</u>D**H**K(Dnp)-*NH₂* did not improve in the presence of Ni^{2+} (Figure 46). In reactions without a strong nucleophile added to the reaction mixtures (hydrolysis treatments), the presence of Ni^{2+} appears to slightly inhibit baSrtB activity. This result may reflect the tendency of Ni^{2+} to suppress baSrtB activity, which we found occurs at a concentration of 200 µM $NiSO_4$ (Figure A15). The results of this assay are consequently hard to interpret. While it's possible that baSrtB activity was inhibited by the presence of nickel ions, this effect does not appear in the reactions performed in the presence of hydroxylamine. This may suggest that the benefit of rendering the leaving group an unsuitable nucleophile balanced out the detrimental effects of the Ni^{2+}, although there is not enough evidence presented here to come to that conclusion. Additionally, the nucleophile activity assay performed in *Chapter 4: The Structural and Biochemical Basis of BaSrtB Selectivity and Activity* suggests

that the assay conditions may need to be modified in order to assess the impact of any nucleophile, whether leaving group or otherwise, on baSrtB activity.

3.3: Discussion

These findings show that baSrtB catalysis relies upon the canonical sortase catalytic triad residue arginine. This residue likely performs a similar function in baSrtB as in other sortases. Historically, the

Figure 46. The reverse reaction in baSrtB activity. BaSrtB reactions appears to progress similarly in the presence of hydroxylamine when the leaving group can serve as a nucleophile (0 µM Nickel Sulfate; red circles) as when the leaving group cannot serve as a nucleophile (400 µM Nickel Sulfate; black squares). When hydroxylamine is not present, however, it appears that activity is slightly higher when the leaving group can serve as a nucleophile (0 µM Nickel Sulfate, hydrolysis; green triangles) than when it cannot (400 µM Nickel Sulfate, hydrolysis; blue triangles).

function of the catalytic arginine has been controversial, though there is evidence suggesting a role in substrate binding and intermediate stabilization.[2,22,46] Substrate-bound crystal structures for spySrtA as well as *in silico* models for lmSrtA and saSrtA indicate that the catalytic arginine forms stabilizing hydrogen bonds with the P4 and P3 carbonyl carbons (*Chapter 2: Structures of Streptococcus pyogenes Class A Sortase in Complex with Substrate and Product Mimics Provide Key Details of Target Recognition*).[22,46]. This is in good agreement with the interactions between R243 and the P4 asparagine and P3 proline residues in our molecular dynamics simulation (Figure 44B). This suggests that Arg243 is primarily important for substrate binding. Interestingly, Tian and Eriksson (2011) report that the catalytic arginine in lmSrtA and saSrtA undergoes a confirmation change *in silico* when the catalytic cysteine is in the active (thiolate) form to interact
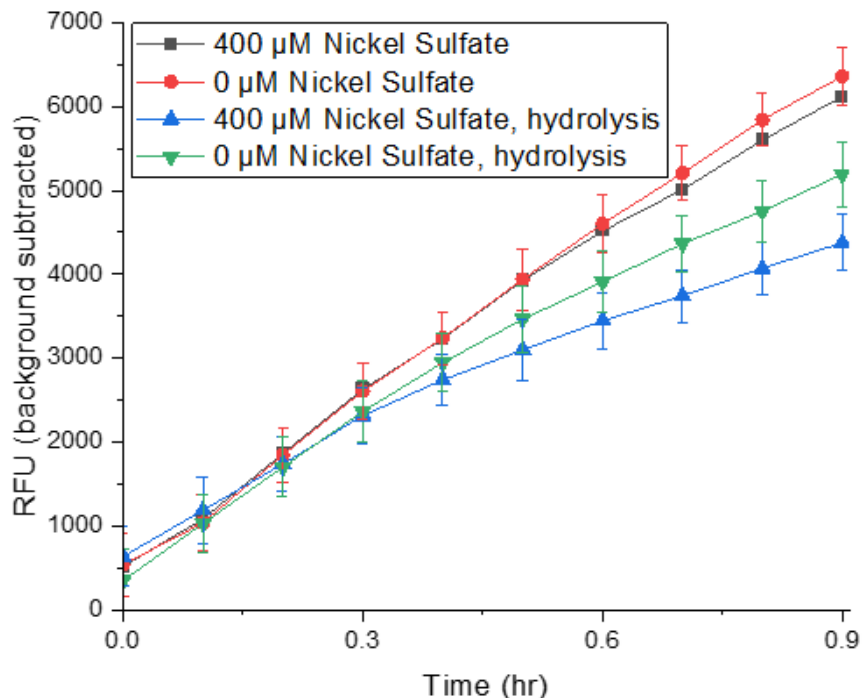
with the P1 threonine and P1' glycine peptide bonds. Additionally, a saSrtB tetrahedral mimetic crystal structure indicated that the catalytic arginine may form a hydrogen bond with the P1' hydroxyl group to stabilize tetrahedral reaction intermediates, although this interaction has not been observed in other sortase models.[1,2,22] It's possible that R243 performs similar functions in baSrtB, however we did not create the relevant models to explore these possibilities.

Though the initial attempt to investigate the importance of His140 and Ser231 in baSrtB activity presented in this work was unsuccessful, further attempts will be made. In sortases, the catalytic histidine likely acts as a general acid to protonate the leaving group, as shown in Figure 42.[2,46,100] In our substrate-bound model, the histidine appears well positioned to perform this function. Interestingly, *in silico* models by Tian and Erikson (2011) suggest the proton transfer from histidine may occur prior to nucleophilic attack by the catalytic cysteine. Their findings suggest that this proton transfer may actually be an important step in driving the nucleophilic attack forward.[46] Ser231 may also be important for baSrtB activity. The Ser231 hydroxyl group is 3.7 Å from the P4 asparagine amino group in our substrate-bound baSrtB model (Figure 47). This positioning is suggestive of a possible role in stabilizing the substrate P4 residue. Interestingly, the corresponding residue (located two positions N-terminal to the catalytic cysteine) in spySrtA is valine. Substrate-bound crystal structures indicate

Figure 47. Ser231 may interact with the P4 Asn. A molecular dynamics simulation shows Ser231 (cyan, side chain shown in stick representation) forming hydrogen bonds with the substrate P4 Asn residue (orange, main and side chains shown in stick representation).

that this valine residue contributes to a hydrophobic pocket that stabilizes the P4 leucine residue recognized by spySrtA (*Chapter 2: Structures of Streptococcus pyogenes Class A Sortase in Complex with Substrate and Product Mimics Provide Key Details of Target Recognition*).[22] It

seems plausible that Ser231 may serve an analogous role in baSrtB. A serine residue located two positions N-terminal to the catalytic cysteine is conserved in four sortase B homologues, baSrtB, lmSrtB, saSrtB, and cdSrtB. Of these, lmSrtB and saSrtB also recognize a P4 asparagine residue.[14,89] The substrate-bound models described in *Chapter 3: Literature Review and Preliminary Experiments with Sortase B Enzymes* indicate a similar orientation of the serine and P4 asparagine, with distances ranging from 3.5 Å to 4.1 Å (Figure 48). In all these models and in molecular dynamics simulations, the stronger interaction is between the P4 Asn and a tyrosine residue on the β6-β7 loop, however it appears that the conserved serine residue is well-positioned to further stabilize P4 Asn within the catalytic pocket.

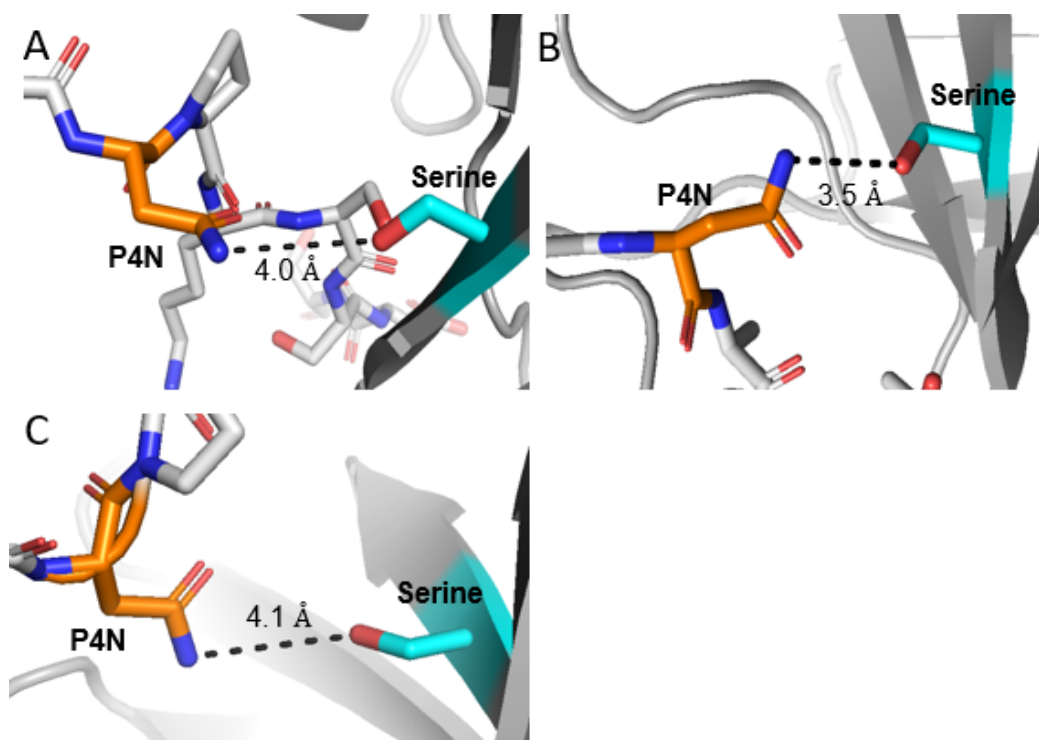**Figure 48. A conserved serine may interact with the P4 Asn.** Substrate-bound lmSrtB **(A-B)** and saSrtB **(C)** models position the conserved serine residue (found two residues N-terminal to the catalytic cysteine) near the asparagine side chain. While these are not depicted within hydrogen bonding distance, only a small displacement would be necessary to position these residues favorably for stabilizing interactions.

Leu106 is also important for baSrtB catalysis. Our molecular dynamics simulation indicates a possible role in stabilizing the P2 proline and P1 threonine side groups (Figure 45B) Like Ser231, this residue is conserved in sortase B homologues. The lmSrtB and saSrtB substrate-bound models generated in *Chapter 3: Literature Review and Preliminary Experiments with Sortase B Enzymes* suggest a similar interaction between the leucine residue and the methyl group of the P1 threonine, although this residue is positioned slightly farther away in these models (not shown).

Our attempts to characterize the importance of the reverse reaction in baSrtB activity were inconclusive. More success may be found by using the more active mutant A241K baSrtB, described in *Chapter 4: The Structural and Biochemical Basis of BaSrtB Selectivity and Activity*. Additionally, monitoring the baSrtB reaction over a longer time frame may provide more insight.

The studies presented in this work indicate that baSrtB activity is calcium-independent in a 0.9-hour FRET-based activity assay. We also find that the catalytic Arg is a vital residue for baSrtB function and identify a residue, Leu106, that is similarly important for baSrtB activity. This residue appears to be conserved in the sortase B homlogs studied in this work, suggesting that it may play a general role in class B sortase catalysis. The roles of Ser231, His140, and the reverse reaction on baSrtB catalytic activity should be the focus of future work.

5.4: Materials and Methods

5.4.1: Protein expression and purification

Proteins were expressed and purified as described in 2.4.1: Expression and Purification of SpySrtA Protein with the modifications described in 4.4.1: Expression and Purification of BaSrtB Protein. BaSrtB used in assays in this chapter were performed using protein from the second and third baSrtB preps (Figure A16).

5.4.2: Model peptide synthesis and purification

Model peptide substrates were synthesized and purified as previously described (2.4.2: Peptide Synthesis).[20,22]

5.4.3: Fluorescence Resonance Energy Transfer- (FRET-) based activity assays

Reactions were performed in black, flat-bottom Costar 96-well plates. Reaction mixtures were 100 uL in volume and most commonly consisted of 50 μM sortase B, 200 μM peptide model substrate (Abz-DNPKTGDEK(Dnp)-NH_2 or Abz-DNPKTGDHK(Dnp)-NH_2), 5 mM hydroxylamine, 50 mM Tris-HCl pH 8.0, 150 mM NaCl, with exceptions outlined below for the pH, S231A, and H140A assays. 400 μM $NiSO_4$ was added to the reaction mixture for nickel binding assays and 5 mM $CaCl_2$ was added to the mixture for the calcium test. Reactions were started by the addition of enzyme, which was prepared beforehand at 2x reaction concentration (100 μM) in a 2x reaction buffer (100 mM Tris pH 8.0, 150 mM NaCl). Peptide stocks often contained DMSO to aid solubility. Residual DMSO in reaction mixtures was ≤1.5%. Reactions were monitored by measuring fluorescence intensity (λ_{ex} = 320 nm, λ_{em} = 420 nm) every 6-18 minutes for up to 18 hours using a Biotek Synergy H1 plate reader. All reactions were performed in triplicate, with exceptions outlined below for the pH, S231A, and H140A assays.

H140A Assay: Assay (n = 1) was performed as described above with the following modification: reaction mixtures contained 16 μM enzyme and 100 μM peptide model substrate (Abz-DNPKTGDEK(Dnp)-NH_2).

S231A Assay: Assay (n = 1) was performed as described above with the following modification: reaction mixtures contained 40 μM enzyme and 100 μM peptide model substrate (Abz-DNPKTGDEK(Dnp)-NH_2).

5.4.4: Multiple Sequence Alignment

Multiple sequence alignments of sortase B proteins (baSrtB, cdSrtB, lmSrtB, and saSrtB) were performed using Uniprot Align.

5.4.5: Molecular Dynamics Simulation and Substrate-Bound Model

The MD simulation and substrate-bound model methods were described in 4.4.7: Molecular Dynamics Simulations and 3.4.3: AlphaFold2 (Galaxy) Models of Substrate-bound Enzymes.

Chapter 6: Concluding Remarks and Future Directions

In this work, we have identified substrate-enzyme interactions that contribute to the activity and selectivity of spySrtA and baSrtB. We have shown that baSrtB activity can be improved by extending the substrate sequence beyond the pentapeptide motif to the P5 and P3' position. These findings agree well with previously published literature in baSrtB and other sortases, suggesting that sortase selectivity preferences in general extend beyond the canonical pentapeptide motif in a way that is highly relevant for SML.[8,13,60,102] Interestingly, it appears that baSrtB can effectively accommodate different amino acids at these positions, and so selectivity may be more flexible beyond the motif than for other substrate residues within the motif, particularly P4, P3, and P1.[8] Further elucidating sortase selectivity beyond the pentapeptide motif should be a subject of future work in baSrtB and other sortases.

We have also shown that baSrtB activity can be greatly improved through modifications to the β7-β8 loop. Low *in vitro* activity is perhaps the single largest reason class B sortases have not been extensively researched or utilized in SML. Further research may be done to better understand why these mutations, particularly the A241K point mutation, have such a dramatic effect on baSrtB activity and to determine if similar mutations might improve the *in vitro* activity of other sortases. Additionally, the β7-β8 loop mutants identified in this work may be further developed to improve baSrtB activity. In particular, assay data and *in silico* modeling suggest that an A236E/A241K baSrtB mutant may further stabilize the β7-β8 loop and promote catalysis. We have also found that the A241K baSrtB point mutant displays nucleophile selectivity preferences that either the wild-type baSrtB does not have or that were not revealed in the time frame studied. Additional research should be done to better understand this effect, and to understand baSrtB nucleophile selectivity in general.

We have also found that the unique sortase B N-terminus, which has not previously been investigated, is important for baSrtB function. This appears to be largely attributable to a single

tyrosine residue, Tyr39. The role of the N-terminus on baSrtB activity may be further explored in future work. For example, it may be possible to discern how Tyr39 is contributing to activity by performing a Y39F mutation, which would be expected to retain van der Waals interactions with Asp144 but not hydrogen binding with substrate residues. Another implication of this finding is that the N-terminus may be a productive region to explore through chimera mutations.

Sortase enzymes have attracted interest because of their role in bacterial pathogenicity and because they are valuable protein engineering tools.[1,5] We have found that the class B sortases, and baSrtB in particular, represent a promising avenue for study. The findings presented in this work build upon previous research in sortase enzymes and provide a more complete picture of activity and selectivity determinants in this important enzyme superfamily.

References

(1) Jacobitz, A. W.; Kattke, M. D.; Wereszczynski, J.; Clubb, R. T. Sortase Transpeptidases: Structural Biology and Catalytic Mechanism. *Adv. Protein Chem. Struct. Biol.* **2017**, *109*, 223–264.

(2) Jacobitz, A. W.; Wereszczynski, J.; Yi, S. W.; Amer, B. R.; Huang, G. L.; Nguyen, A. V.; Sawaya, M. R.; Jung, M. E.; McCammon, J. A.; Clubb, R. T. Structural and Computational Studies of the Staphylococcus Aureus Sortase B-Substrate Complex Reveal a Substrate-Stabilized Oxyanion Hole. *J. Biol. Chem.* **2014**, *289* (13), 8891–8902.

(3) Malik, A.; Kim, S. B. A Comprehensive in Silico Analysis of Sortase Superfamily. *J. Microbiol.* **2019**, *57* (6), 431–443.

(4) Bradshaw, W. J.; Davies, A. H.; Chambers, C. J.; Roberts, A. K.; Shone, C. C.; Acharya, K. R. Molecular Features of the Sortase Enzyme Family. *FEBS J.* **2015**, *282* (11), 2097–2114.

(5) Spirig, T.; Weiner, E. M.; Clubb, R. T. Sortase Enzymes in Gram-Positive Bacteria. *Mol. Microbiol.* **2011**, *82* (5), 1044–1059.

(6) Nikghalb, K. D.; Horvath, N. M.; Prelesnik, J. L.; Banks, O. G. B.; Filipov, P. A.; Row, R. D.; Roark, T. J.; Antos, J. M. Expanding the Scope of Sortase-Mediated Ligations by Using Sortase Homologues. *Chembiochem* **2018**, *19* (2), 185–195.

(7) Antos, J. M.; Truttmann, M. C.; Ploegh, H. L. Recent Advances in Sortase-Catalyzed Ligation Methodology. *Curr. Opin. Struct. Biol.* **2016**, *38*, 111–118.

(8) Puorger, C.; Di Girolamo, S.; Lipps, G. Elucidation of the Recognition Sequence of Sortase B from Bacillus Anthracis by Using a Newly Developed Liquid Chromatography-Mass Spectrometry-Based Method. *Biochemistry* **2017**, *56* (21), 2641–2650.

(9) Ilangovan, U.; Ton-That, H.; Iwahara, J.; Schneewind, O.; Clubb, R. T. Structure of Sortase, the Transpeptidase That Anchors Proteins to the Cell Wall of Staphylococcus Aureus. *Proc Natl Acad Sci USA* **2001**, *98* (11), 6056–6061.

(10) Dramsi, S.; Magnet, S.; Davison, S.; Arthur, M. Covalent Attachment of Proteins to Peptidoglycan. *FEMS Microbiol. Rev.* **2008**, *32* (2), 307–320.

(11) Ton-That, H.; Faull, K. F.; Schneewind, O. Anchor Structure of Staphylococcal Surface Proteins. *Journal of Biological Chemistry* **1997**, *272* (35), 22285–22292.

(12) Zong, Y.; Mazmanian, S. K.; Schneewind, O.; Narayana, S. V. L. The Structure of Sortase B, a Cysteine Transpeptidase That Tethers Surface Protein to the Staphylococcus Aureus Cell Wall. *Structure* **2004**, *12* (1), 105–112.

(13) Donahue, E. H.; Dawson, L. F.; Valiente, E.; Firth-Clark, S.; Major, M. R.; Littler, E.; Perrior, T. R.; Wren, B. W. Clostridium Difficile Has a Single Sortase, SrtB, That Can Be Inhibited by Small-Molecule Inhibitors. *BMC Microbiol.* **2014**, *14*, 219.

(14) Mariscotti, J. F.; García-del Portillo, F.; Pucciarelli, M. G. The Listeria Monocytogenes Sortase-B Recognizes Varied Amino Acids at Position 2 of the Sorting Motif. *J. Biol. Chem.* **2009**, *284* (10), 6140–6146.

(15) Freund, C.; Schwarzer, D. Engineered Sortases in Peptide and Protein Chemistry. *Chembiochem* **2021**, *22* (8), 1347–1356.

(16) Race, P. R.; Bentley, M. L.; Melvin, J. A.; Crow, A.; Hughes, R. K.; Smith, W. D.; Sessions, R. B.; Kehoe, M. A.; McCafferty, D. G.; Banfield, M. J. Crystal Structure of Streptococcus

Pyogenes Sortase A: Implications for Sortase Mechanism. *J. Biol. Chem.* **2009**, *284* (11), 6924–6933.

(17) Suree, N.; Liew, C. K.; Villareal, V. A.; Thieu, W.; Fadeev, E. A.; Clemens, J. J.; Jung, M. E.; Clubb, R. T. The Structure of the Staphylococcus Aureus Sortase-Substrate Complex Reveals How the Universally Conserved LPXTG Sorting Signal Is Recognized. *J. Biol. Chem.* **2009**, *284* (36), 24465–24477.

(18) Ritzefeld, M. Sortagging: A Robust and Efficient Chemoenzymatic Ligation Strategy. *Chem. Eur. J* **2014**, *20* (28), 8516–8529.

(19) Tsukiji, S.; Nagamune, T. Sortase-Mediated Ligation: A Gift from Gram-Positive Bacteria to Protein Engineering. *Chembiochem* **2009**, *10* (5), 787–798.

(20) Piper, I. M.; Struyvenberg, S. A.; Valgardson, J. D.; Johnson, D. A.; Gao, M.; Johnston, K.; Svendsen, J. E.; Kodama, H. M.; Hvorecny, K. L.; Antos, J. M.; et al. Sequence Variation in the B7-B8 Loop of Bacterial Class A Sortase Enzymes Alters Substrate Selectivity. *J. Biol. Chem.* **2021**, *297* (2), 100981.

(21) Antos, J. M.; Chew, G.-L.; Guimaraes, C. P.; Yoder, N. C.; Grotenbreg, G. M.; Popp, M. W.-L.; Ploegh, H. L. Site-Specific N- and C-Terminal Labeling of a Single Polypeptide Using Sortases of Different Specificity. *J. Am. Chem. Soc.* **2009**, *131* (31), 10800–10801.

(22) Johnson, D. A.; Piper, I. M.; Vogel, B. A.; Jackson, S. N.; Svendsen, J. E.; Kodama, H. M.; Lee, D. E.; Lindblom, K. M.; McCarty, J.; Antos, J. M.; et al. Structures of Streptococcus Pyogenes Class A Sortase in Complex with Substrate and Product Mimics Provide Key Details of Target Recognition. *J. Biol. Chem.* **2022**, *298* (10), 102446.

(23) Ton-That, H.; Schneewind, O. Anchor Structure of Staphylococcal Surface Proteins. IV. Inhibitors of the Cell Wall Sorting Reaction. *J. Biol. Chem.* **1999**, *274* (34), 24316–24320.

(24) Mazmanian, S. K.; Liu, G.; Ton-That, H.; Schneewind, O. Staphylococcus Aureus Sortase, an Enzyme That Anchors Surface Proteins to the Cell Wall. *Science* **1999**, *285* (5428), 760–763.

(25) Ton-That, H.; Liu, G.; Mazmanian, S. K.; Faull, K. F.; Schneewind, O. Purification and Characterization of Sortase, the Transpeptidase That Cleaves Surface Proteins of Staphylococcus Aureus at the LPXTG Motif. *Proc Natl Acad Sci USA* **1999**, *96* (22), 12424–12429.

(26) Zhang, J.; Liu, H.; Zhu, K.; Gong, S.; Dramsi, S.; Wang, Y.-T.; Li, J.; Chen, F.; Zhang, R.; Zhou, L.; et al. Antiinfective Therapy with a Small Molecule Inhibitor of Staphylococcus Aureus Sortase. *Proc Natl Acad Sci USA* **2014**, *111* (37), 13517–13522.

(27) Hendrickx, A. P. A.; Budzik, J. M.; Oh, S.-Y.; Schneewind, O. Architects at the Bacterial Surface - Sortases and the Assembly of Pili with Isopeptide Bonds. *Nat. Rev. Microbiol.* **2011**, *9* (3), 166–176.

(28) Perry, A. M.; Ton-That, H.; Mazmanian, S. K.; Schneewind, O. Anchoring of Surface Proteins to the Cell Wall of Staphylococcus Aureus. III. Lipid II Is an in Vivo Peptidoglycan Substrate for Sortase-Catalyzed Surface Protein Anchoring. *J. Biol. Chem.* **2002**, *277* (18), 16241–16248.

(29) Clancy, K. W.; Melvin, J. A.; McCafferty, D. G. Sortase Transpeptidases: Insights into Mechanism, Substrate Specificity, and Inhibition. *Biopolymers* **2010**, *94* (4), 385–396.

(30) Chan, A. H.; Yi, S. W.; Terwilliger, A. L.; Maresso, A. W.; Jung, M. E.; Clubb, R. T. Structure of the Bacillus Anthracis Sortase A Enzyme Bound to Its Sorting Signal: A

FLEXIBLE AMINO-TERMINAL APPENDAGE MODULATES SUBSTRATE ACCESS. *J. Biol. Chem.* **2015**, *290* (42), 25461–25474.

(31)   Zong, Y.; Bice, T. W.; Ton-That, H.; Schneewind, O.; Narayana, S. V. L. Crystal Structures of Staphylococcus Aureus Sortase A and Its Substrate Complex. *J. Biol. Chem.* **2004**, *279* (30), 31383–31389.

(32)   Popp, M. W.; Dougan, S. K.; Chuang, T.-Y.; Spooner, E.; Ploegh, H. L. Sortase-Catalyzed Transformations That Improve the Properties of Cytokines. *Proc Natl Acad Sci USA* **2011**, *108* (8), 3169–3174.

(33)   Guimaraes, C. P.; Witte, M. D.; Theile, C. S.; Bozkurt, G.; Kundrat, L.; Blom, A. E. M.; Ploegh, H. L. Site-Specific C-Terminal and Internal Loop Labeling of Proteins Using Sortase-Mediated Reactions. *Nat. Protoc.* **2013**, *8* (9), 1787–1799.

(34)   Dai, X.; Böker, A.; Glebe, U. Broadening the Scope of Sortagging. *RSC Adv.* **2019**, *9* (9), 4700–4721.

(35)   Gao, M.; Johnson, D. A.; Piper, I. M.; Kodama, H. M.; Svendsen, J. E.; Tahti, E.; Longshore-Neate, F.; Vogel, B.; Antos, J. M.; Amacher, J. F. Structural and Biochemical Analyses of Selectivity Determinants in Chimeric Streptococcus Class A Sortase Enzymes. *Protein Sci.* **2022**, *31* (3), 701–715.

(36)   Raz, A.; Tanasescu, A.-M.; Zhao, A. M.; Serrano, A.; Alston, T.; Sol, A.; Bachrach, G.; Fischetti, V. A. Streptococcus Pyogenes Sortase Mutants Are Highly Susceptible to Killing by Host Factors Due to Aberrant Envelope Physiology. *PLoS ONE* **2015**, *10* (10), e0140784.

(37)   Litou, Z. I.; Bagos, P. G.; Tsirigos, K. D.; Liakopoulos, T. D.; Hamodrakas, S. J. Prediction of Cell Wall Sorting Signals in Gram-Positive Bacteria with a Hidden Markov Model: Application to Complete Genomes. *J. Bioinform. Comput. Biol.* **2008**, *6* (2), 387–401.

(38)   Theile, C. S.; Witte, M. D.; Blom, A. E. M.; Kundrat, L.; Ploegh, H. L.; Guimaraes, C. P. Site-Specific N-Terminal Labeling of Proteins Using Sortase-Mediated Reactions. *Nat. Protoc.* **2013**, *8* (9), 1800–1807.

(39)   Schmohl, L.; Bierlmeier, J.; von Kügelgen, N.; Kurz, L.; Reis, P.; Barthels, F.; Mach, P.; Schutkowski, M.; Freund, C.; Schwarzer, D. Identification of Sortase Substrates by Specificity Profiling. *Bioorg. Med. Chem.* **2017**, *25* (18), 5002–5007.

(40)   Kruger, R. G.; Otvos, B.; Frankel, B. A.; Bentley, M.; Dostal, P.; McCafferty, D. G. Analysis of the Substrate Specificity of the Staphylococcus Aureus Sortase Transpeptidase SrtA. *Biochemistry* **2004**, *43* (6), 1541–1551.

(41)   Wójcik, M.; Vázquez Torres, S.; Quax, W. J.; Boersma, Y. L. Sortase Mutants with Improved Protein Thermostability and Enzymatic Activity Obtained by Consensus Design. *Protein Eng. Des. Sel.* **2019**, *32* (12), 555–564.

(42)   Zhang, Y.; Huang, D.; Zhang, C.; Meng, J.; Tan, B.; Deng, Z. IQF Characterization of a Cathepsin B-Responsive Nanoprobe for Report of Differentiation of HL60 Cells into Macrophages. *RSC Adv.* **2021**, *11* (27), 16522–16529.

(43)   Dorr, B. M.; Ham, H. O.; An, C.; Chaikof, E. L.; Liu, D. R. Reprogramming the Specificity of Sortase Enzymes. *Proc Natl Acad Sci USA* **2014**, *111* (37), 13343–13348.

(44)   Zwart, P. H.; Grosse-Kunstleve, R. W.; Lebedev, A. A.; Murshudov, G. N.; Adams, P. D. Surprises and Pitfalls Arising from (Pseudo)Symmetry. *Acta Crystallogr. D Biol. Crystallogr.* **2008**, *64* (Pt 1), 99–107.

(45)   Marraffini, L. A.; DeDent, A. C.; Schneewind, O. Sortases and the Art of Anchoring Proteins to the Envelopes of Gram-Positive Bacteria. *Microbiol. Mol. Biol. Rev.* **2006**, *70* (1), 192–221.

(46)   Tian, B.-X.; Eriksson, L. A. Catalytic Mechanism and Roles of Arg197 and Thr183 in the Staphylococcus Aureus Sortase A Enzyme. *J. Phys. Chem. B* **2011**, *115* (44), 13003–13011.

(47)   Frankel, B. A.; Tong, Y.; Bentley, M. L.; Fitzgerald, M. C.; McCafferty, D. G. Mutational Analysis of Active Site Residues in the Staphylococcus Aureus Transpeptidase SrtA. *Biochemistry* **2007**, *46* (24), 7269–7278.

(48)   Schleifer, K. H.; Kandler, O. Peptidoglycan Types of Bacterial Cell Walls and Their Taxonomic Implications. *Bacteriol. Rev.* **1972**, *36* (4), 407–477.

(49)   Münch, D.; Sahl, H.-G. Structural Variations of the Cell Wall Precursor Lipid II in Gram-Positive Bacteria - Impact on Binding and Efficacy of Antimicrobial Peptides. *Biochim. Biophys. Acta* **2015**, *1848* (11 Pt B), 3062–3071.

(50)   Hadi, T.; Hazra, S.; Tanner, M. E.; Blanchard, J. S. Structure of MurNAc 6-Phosphate Hydrolase (MurQ) from Haemophilus Influenzae with a Bound Inhibitor. *Biochemistry* **2013**, *52* (51), 9358–9366.

(51)   Vollmer, W.; Blanot, D.; de Pedro, M. A. Peptidoglycan Structure and Architecture. *FEMS Microbiol. Rev.* **2008**, *32* (2), 149–167.

(52)   Navarre, W. W.; Schneewind, O. Surface Proteins of Gram-Positive Bacteria and Mechanisms of Their Targeting to the Cell Wall Envelope. *Microbiol. Mol. Biol. Rev.* **1999**, *63* (1), 174–229.

(53)   Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber Ff99SB Protein Force Field. *Proteins* **2010**, *78* (8), 1950–1958.

(54)   Ruzin, A.; Severin, A.; Ritacco, F.; Tabei, K.; Singh, G.; Bradford, P. A.; Siegel, M. M.; Projan, S. J.; Shlaes, D. M. Further Evidence That a Cell Wall Precursor [C(55)-MurNAc-(Peptide)-GlcNAc] Serves as an Acceptor in a Sorting Reaction. *J. Bacteriol.* **2002**, *184* (8), 2141–2147.

(55)   Biswas, T.; Pawale, V. S.; Choudhury, D.; Roy, R. P. Sorting of LPXTG Peptides by Archetypal Sortase A: Role of Invariant Substrate Residues in Modulating the Enzyme Dynamics and Conformational Signature of a Productive Substrate. *Biochemistry* **2014**, *53* (15), 2515–2524.

(56)   Podracky, C. J.; An, C.; DeSousa, A.; Dorr, B. M.; Walsh, D. M.; Liu, D. R. Laboratory Evolution of a Sortase Enzyme That Modifies Amyloid-β Protein. *Nat. Chem. Biol.* **2021**, *17* (3), 317–325.

(57)   Schmohl, L.; Bierlmeier, J.; Gerth, F.; Freund, C.; Schwarzer, D. Engineering Sortase A by Screening a Second-Generation Library Using Phage Display. *J. Pept. Sci.* **2017**, *23* (7–8), 631–635.

(58)   Zou, Z.; Nöth, M.; Jakob, F.; Schwaneberg, U. Designed Streptococcus Pyogenes Sortase A Accepts Branched Amines as Nucleophiles in Sortagging. *Bioconjug. Chem.* **2020**, *31* (11), 2476–2481.

(59)   Hirakawa, H.; Ishikawa, S.; Nagamune, T. Design of Ca2+-Independent Staphylococcus Aureus Sortase A Mutants. *Biotechnol. Bioeng.* **2012**, *109* (12), 2955–2961.

(60)   Pritz, S.; Wolf, Y.; Kraetke, O.; Klose, J.; Bienert, M.; Beyermann, M. Synthesis of Biologically Active Peptide Nucleic Acid-Peptide Conjugates by Sortase-Mediated Ligation. *J. Org. Chem.* **2007**, *72* (10), 3909–3912.

(61)   Wilkins, M. R.; Gasteiger, E.; Bairoch, A.; Sanchez, J. C.; Williams, K. L.; Appel, R. D.; Hochstrasser, D. F. Protein Identification and Analysis Tools in the ExPASy Server. *Methods Mol. Biol.* **1999**, *112*, 531–552.

(62)   Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (Pt 2), 125–132.

(63)   Kabsch, W. Integration, Scaling, Space-Group Assignment and Post-Refinement. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (Pt 2), 133–144.

(64)   Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (Pt 1), 12–21.

(65)   Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (Pt 4), 486–501.

(66)   Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; et al. PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (Pt 2), 213–221.

(67)   Moriarty, N. W.; Grosse-Kunstleve, R. W.; Adams, P. D. Electronic Ligand Builder and Optimization Workbench (ELBOW): A Tool for Ligand Coordinate and Restraint Generation. *Acta Crystallogr. D Biol. Crystallogr.* **2009**, *65* (Pt 10), 1074–1080.

(68)   Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26* (16), 1701–1718.

(69)   Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.

(70)   Lindahl; Hess; Spoel, V. D. GROMACS 2020.4 Manual. *Zenodo* **2020**.

(71)   Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174.

(72)   Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollmann, P. A. Application of RESP Charges to Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation. *J. Am. Chem. Soc.* **1993**, *115* (21), 9620–9631.

(73)   Barca, G. M. J.; Bertoni, C.; Carrington, L.; Datta, D.; De Silva, N.; Deustua, J. E.; Fedorov, D. G.; Gour, J. R.; Gunina, A. O.; Guidez, E.; et al. Recent Developments in the General Atomic and Molecular Electronic Structure System. *J. Chem. Phys.* **2020**, *152* (15), 154102.

(74)   Lu, T.; Chen, F. Multiwfn: A Multifunctional Wavefunction Analyzer. *J. Comput. Chem.* **2012**, *33* (5), 580–592.

(75)   Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577.

(76)   Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101.

(77)   Parrinello, M. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics

Method. *J. Appl. Phys.* **1981**, *52* (12), 7182.

(78)   Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463–1472.

(79)   Towns, J.; Cockerill, T.; Dahan, M.; Foster, I.; Gaither, K.; Grimshaw, A.; Hazlewood, V.; Lathrop, S.; Lifka, D.; Peterson, G. D.; et al. XSEDE: Accelerating Scientific Discovery. *Comput. Sci. Eng.* **2014**, *16* (5), 62–74.

(80)   Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185* (2), 604–613.

(81)   Kang, H. J.; Coulibaly, F.; Proft, T.; Baker, E. N. Crystal Structure of Spy0129, a Streptococcus Pyogenes Class B Sortase Involved in Pilus Assembly. *PLoS ONE* **2011**, *6* (1), e15969.

(82)   Young, P. G.; Proft, T.; Harris, P. W. R.; Brimble, M. A.; Baker, E. N. Structure and Activity of Streptococcus Pyogenes SipA: A Signal Peptidase-like Protein Essential for Pilus Polymerisation. *PLoS ONE* **2014**, *9* (6), e99135.

(83)   Tamai, E.; Sekiya, H.; Maki, J.; Nariya, H.; Yoshida, H.; Kamitori, S. X-Ray Structure of Clostridium Perfringens Sortase B Cysteine Transpeptidase. *Biochem. Biophys. Res. Commun.* **2017**, *493* (3), 1267–1272.

(84)   Maresso, A. W.; Chapa, T. J.; Schneewind, O. Surface Protein IsdC and Sortase B Are Required for Heme-Iron Scavenging of Bacillus Anthracis. *J. Bacteriol.* **2006**, *188* (23), 8145–8152.

(85)   Chambers, C. J.; Roberts, A. K.; Shone, C. C.; Acharya, K. R. Structure and Function of a Clostridium Difficile Sortase Enzyme. *Sci. Rep.* **2015**, *5*, 9449.

(86)   Kang, C.-Y.; Huang, I.-H.; Chou, C.-C.; Wu, T.-Y.; Chang, J.-C.; Hsiao, Y.-Y.; Cheng, C.-H.; Tsai, W.-J.; Hsu, K.-C.; Wang, S. Functional Analysis of Clostridium Difficile Sortase B Reveals Key Residues for Catalytic Activity and Substrate Specificity. *J. Biol. Chem.* **2020**, *295* (11), 3734–3745.

(87)   Yin, J.-C.; Fei, C.-H.; Lo, Y.-C.; Hsiao, Y.-Y.; Chang, J.-C.; Nix, J. C.; Chang, Y.-Y.; Yang, L.-W.; Huang, I.-H.; Wang, S. Structural Insights into Substrate Recognition by Clostridium Difficile Sortase. *Front. Cell. Infect. Microbiol.* **2016**, *6*, 160.

(88)   van Leeuwen, H. C.; Klychnikov, O. I.; Menks, M. A. C.; Kuijper, E. J.; Drijfhout, J. W.; Hensbergen, P. J. Clostridium Difficile Sortase Recognizes a (S/P)PXTG Sequence Motif and Can Accommodate Diaminopimelic Acid as a Substrate for Transpeptidation. *FEBS Lett.* **2014**, *588* (23), 4325–4333.

(89)   Mazmanian, S. K.; Ton-That, H.; Su, K.; Schneewind, O. An Iron-Regulated Sortase Anchors a Class of Surface Protein during Staphylococcus Aureus Pathogenesis. *Proc Natl Acad Sci USA* **2002**, *99* (4), 2293–2298.

(90)   Pucciarelli, M. G.; Calvo, E.; Sabet, C.; Bierne, H.; Cossart, P.; García-del Portillo, F. Identification of Substrates of the Listeria Monocytogenes Sortases A and B by a Non-Gel Proteomic Analysis. *Proteomics* **2005**, *5* (18), 4808–4817.

(91)   Zhang, R.; Wu, R.; Joachimiak, G.; Mazmanian, S. K.; Missiakas, D. M.; Gornicki, P.; Schneewind, O.; Joachimiak, A. Structures of Sortase B from Staphylococcus Aureus and Bacillus Anthracis Reveal Catalytic Amino Acid Triad in the Active Site. *Structure* **2004**, *12* (7), 1147–1156.

(92)   Bentley, M. L.; Gaweska, H.; Kielec, J. M.; McCafferty, D. G. Engineering the Substrate

Specificity of Staphylococcus Aureus Sortase A. The Beta6/Beta7 Loop from SrtB Confers NPQTN Recognition to SrtA. *J. Biol. Chem.* **2007**, *282* (9), 6571–6581.

(93) Kruger, R. G.; Dostal, P.; McCafferty, D. G. Development of a High-Performance Liquid Chromatography Assay and Revision of Kinetic Parameters for the Staphylococcus Aureus Sortase Transpeptidase SrtA. *Anal. Biochem.* **2004**, *326* (1), 42–48.

(94) Bierne, H.; Garandeau, C.; Pucciarelli, M. G.; Sabet, C.; Newton, S.; Garcia-del Portillo, F.; Cossart, P.; Charbit, A. Sortase B, a New Class of Sortase in Listeria Monocytogenes. *J. Bacteriol.* **2004**, *186* (7), 1972–1982.

(95) Kang, S. S.; Kim, J.-G.; Lee, T.-H.; Oh, K.-B. Flavonols Inhibit Sortases and Sortase-Mediated Staphylococcus Aureus Clumping to Fibrinogen. *Biol. Pharm. Bull.* **2006**, *29* (8), 1751–1755.

(96) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589.

(97) Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models. *Nucleic Acids Res.* **2022**, *50* (D1), D439–D444.

(98) Correction to "The Galaxy Platform for Accessible, Reproducible and Collaborative Biomedical Analyses: 2022 Update". *Nucleic Acids Res.* **2022**, *50* (15), 8999.

(99) Arato, V.; Gasperini, G.; Giusti, F.; Ferlenghi, I.; Scarselli, M.; Leuzzi, R. Dual Role of the Colonization Factor CD2831 in Clostridium Difficile Pathogenesis. *Sci. Rep.* **2019**, *9* (1), 5554.

(100) Frankel, B. A.; Kruger, R. G.; Robinson, D. E.; Kelleher, N. L.; McCafferty, D. G. Staphylococcus Aureus Sortase Transpeptidase SrtA: Insight into the Kinetic Mechanism and Evidence for a Reverse Protonation Catalytic Mechanism. *Biochemistry* **2005**, *44* (33), 11188–11200.

(101) Weiner, E. M.; Robson, S.; Marohn, M.; Clubb, R. T. The Sortase A Enzyme That Attaches Proteins to the Cell Wall of Bacillus Anthracis Contains an Unusual Active Site Architecture. *J. Biol. Chem.* **2010**, *285* (30), 23433–23443.

(102) Boyko, K. V.; Rosenkranz, E. A.; Smith, D. M.; Miears, H. L.; Oueld Es Cheikh, M.; Lund, M. Z.; Young, J. C.; Reardon, P. N.; Okon, M.; Smirnov, S. L.; et al. Sortase-Mediated Segmental Labeling: A Method for Segmental Assignment of Intrinsically Disordered Regions in Proteins. *PLoS ONE* **2021**, *16* (10), e0258531.

(103) Páll, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E. Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. In *Solving Software Challenges for Exascale*; Markidis, S., Laure, E., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, 2015; Vol. 8759, pp 3–27.

(104) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; et al. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29* (7), 845–854.

(105) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.*

2008, *4* (3), 435–447.

(106) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: A Package for Molecular Simulation and Trajectory Analysis. *J. Mol. Model.* **2001**, *7* (8), 306–317.

(107) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91* (1–3), 43–56.

(108) Miyamoto, S.; Kollman, P. A. Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13* (8), 952–962.

(109) David Row, R.; Roark, T. J.; Philip, M. C.; Perkins, L. L.; Antos, J. M. Enhancing the Efficiency of Sortase-Mediated Ligations through Nickel-Peptide Complex Formation. *Chem. Commun.* **2015**, *51* (63), 12548–12551.

Appendix

Supplemental Methods for Chapter 2: Structures of *Streptococcus pyogenes* Class A Sortase in Complex with Substrate and Product Mimics Provide Key Details of Target Recognition[22]

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).[71] 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.[72] Quantum chemistry calculations were performed using GAMESS with RESP fitting performed using Multiwfn.[73,74] 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.[75] 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.[76] This was followed by a 1 ns equilibration in the NPT ensemble at 1 bar without position restraints using a Parrinello-Rahman barostat.[77] We used an integration time step of 2 fs. All bonds to hydrogen atoms were constrained using the LINCS algorithm.[78] 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.[79] To monitor distances between atoms of interest, we used the PLUMED2 plugin.[80]

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 CH_2Cl_2/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 partioned 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 CH_2Cl_2 (3x). A 5 mL solution of 95:2.5:2.5 TFA/TIPS/H_2O 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_2O 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_2O/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 identity and purity of LPAT-LII in this stock solution were confirmed by LC-ESI-MS (Figure A8A) and RP-HPLC (Figure A8B). These analyses were conducted on a Dionex UltiMate 3000 HPLC system interfaced with an Advion CMS expression[L] mass spectrometer. Separations were achieved with a Phenomenex Kinetix® 2.6 mM C18 100 Å column (100 x 2.1 mm) [aqueous (95% H_2O, 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)].

Figures A1-A8 and Table A1 are supplemental figures for Chapter 2: Structures of *Streptococcus pyogenes* Class A Sortase in Complex with Substrate and Product Mimics Provide Key Details of Target Recognition[22]

A

spySrtA *(wild-type with N-terminal His₆, calculated average MW = 20657.5 Da)*

MESSHHHHHHENLYFQSSVLQAQMAAQQLPVIGGIAIPELGINLPIFKGLGNTELIYGAGTMKEEQVMGGENNY
SLASHHIFGITGSSQMLFSPLERAQNGMSIYLTDKEKIYEYIIKDVFTVAPERVDVIDDTAGLKEVTLVTCTDI
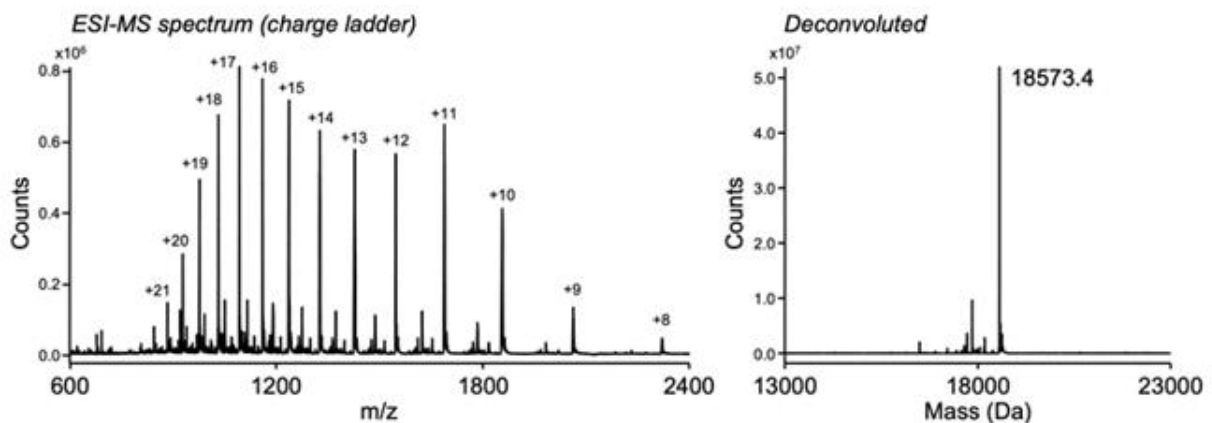EATERIIVKGELKTEYDFDKAPADVLKAFNHSYNQVST
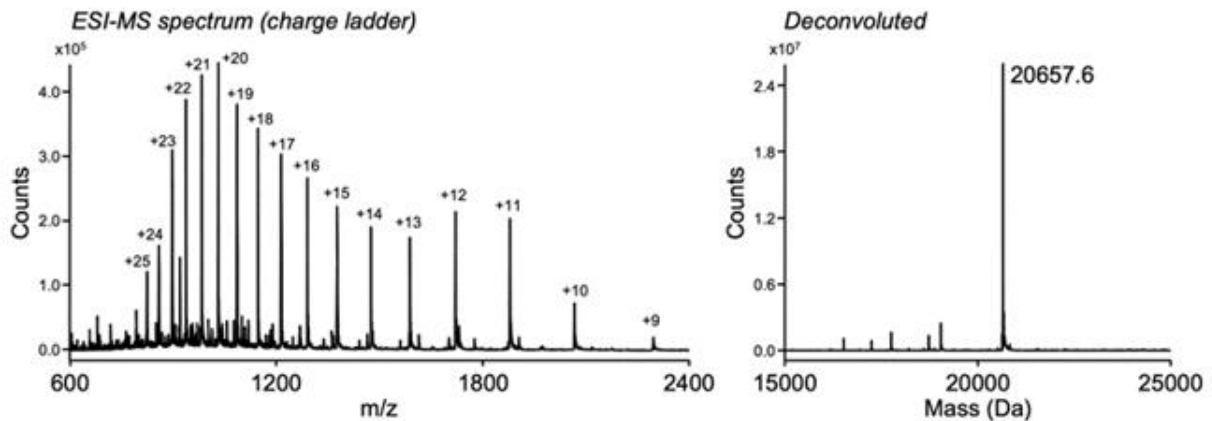
B

C208A spySrtA *(His₆ tag removed via TEV cleavage, calculated average MW = 18573.3 Da)*

SSVLQAQMAAQQLPVIGGIAIPELGINLPIFKGLGNTELIYGAGTMKEEQVMGGENNYSLASHHIFGITGSSQM
LFSPLERAQNGMSIYLTDKEKIYEYIIKDVFTVAPERVDVIDDTAGLKEVTLVTATDIEATERIIVKGELKTEY
DFDKAPADVLKAFNHSYNQVST

Figure A1. 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.
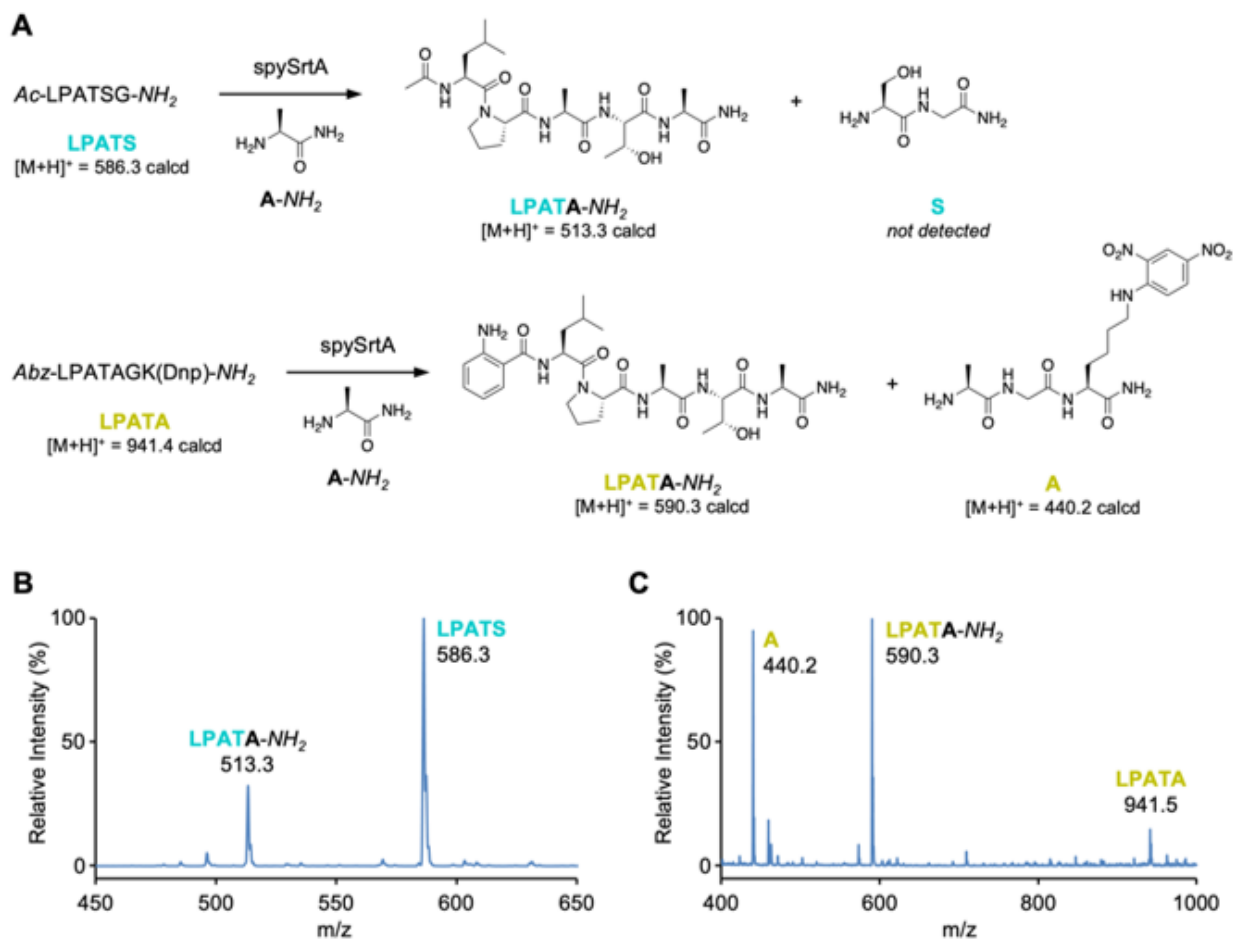
Figure A2. Model transacylation reactions of spySrtA with LPATA and LPATS substrates. (A) Schemes for the reaction of model peptides (*Ac*-LPATSG-*NH₂* or *Abz*-LPATAGK(Dnp)-*NH₂*) with excess alanine amide (A-*NH₂*) in the presence of spySrtA. Conditions: 50 μM LPATS/LPATA substrate, 5 mM A-*NH₂*, 1 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-*NH₂* (5 μM spySrtA, 180 min time point). **(C)** ESI-MS spectrum of crude reaction mixture with *Abz*-LPATAGK(Dnp)-*NH₂* (1 μM spySrtA, 260 min time point).
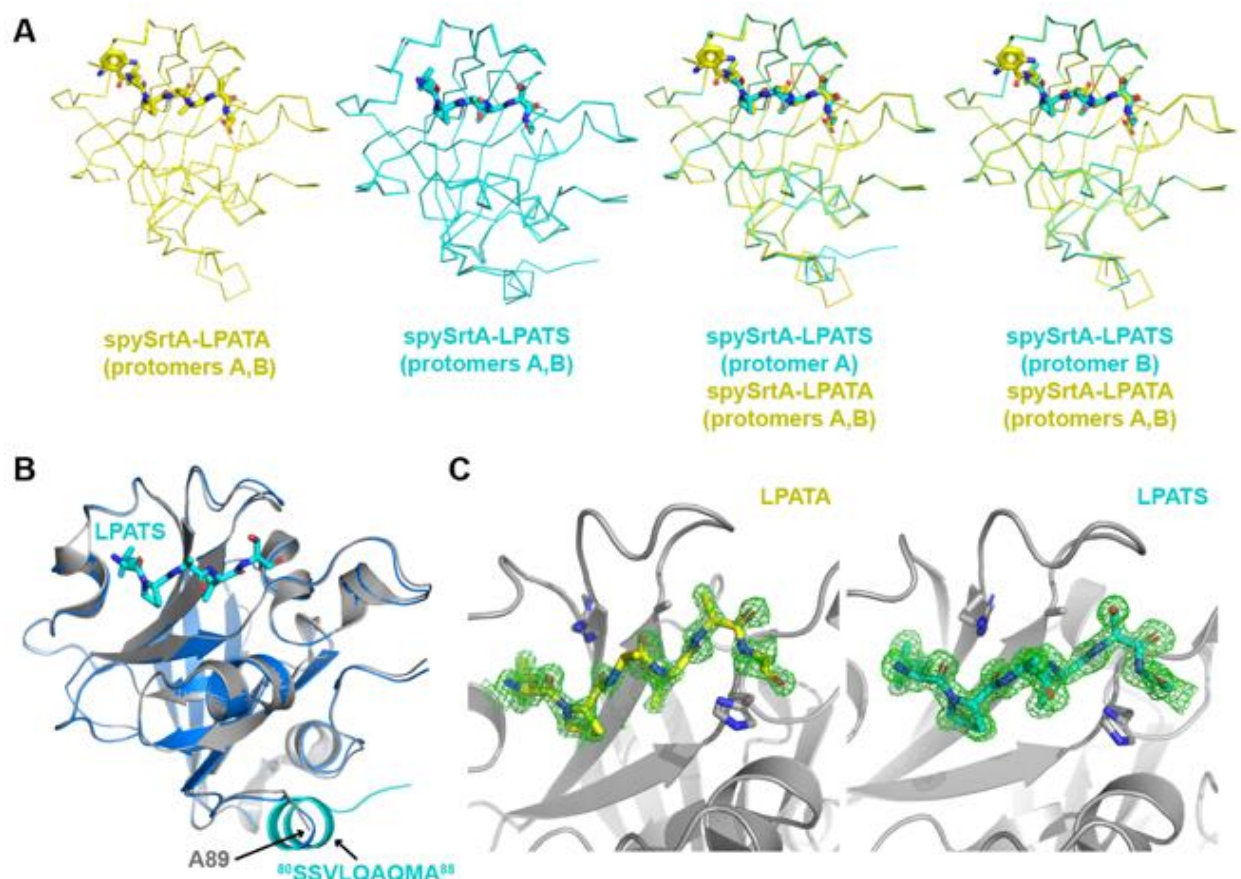
Figure A3. Structural alignments of spySrtA complex structures. (A) Pairwise alignments for all protomers of both structures reveal that the main-chain atoms of spySrtA-LPATS align to protomer A or protomer B of spySrtA-LPATA with RMSD = 0.124 Å (544 atoms) and 0.114 Å (556), respectively. Alignment 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.
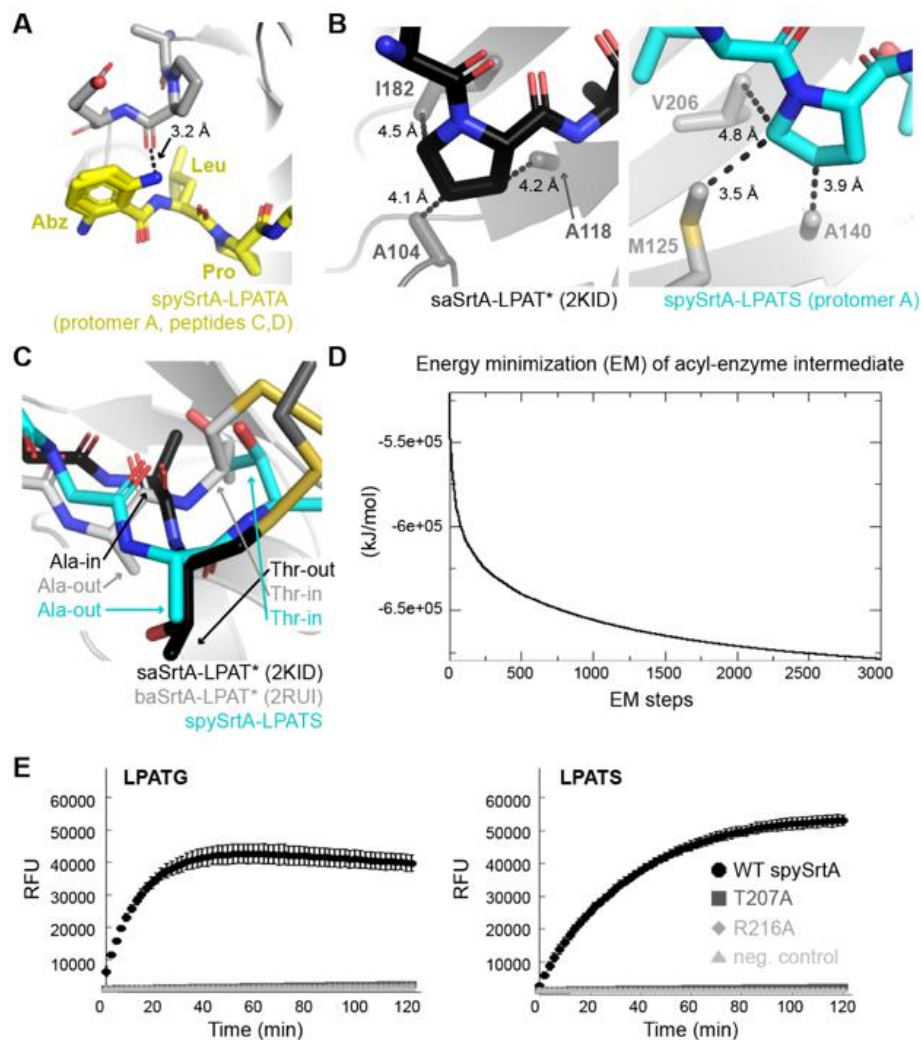
Figure A4. The spySrtA LPATA- and LPATS-complex structures. (A) The Abz moiety at the N-terminus of the *Abz*-LPATAGK(Dnp)-*NH₂* peptide is in two conformations in the peptides bound to protomer A (chain C) and B (chain D) in the spySrtA-LPATA structure. The peptides are in stick representation and colored 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*-LPATGGK(Dnp)-*NH₂*, *left*, or *Abz*-LPATSGK(Dnp)-*NH₂*, *right*, and H_2NOH in the presence of WT (black circles), T207A (dark gray squares), and R216A (gray diamonds) spySrtA protein, and a no protein control (light gray triangles). The key is the same for both graphs. Data for the *Abz*-LPATAGK(Dnp)-*NH₂* peptide is in figure 9D.
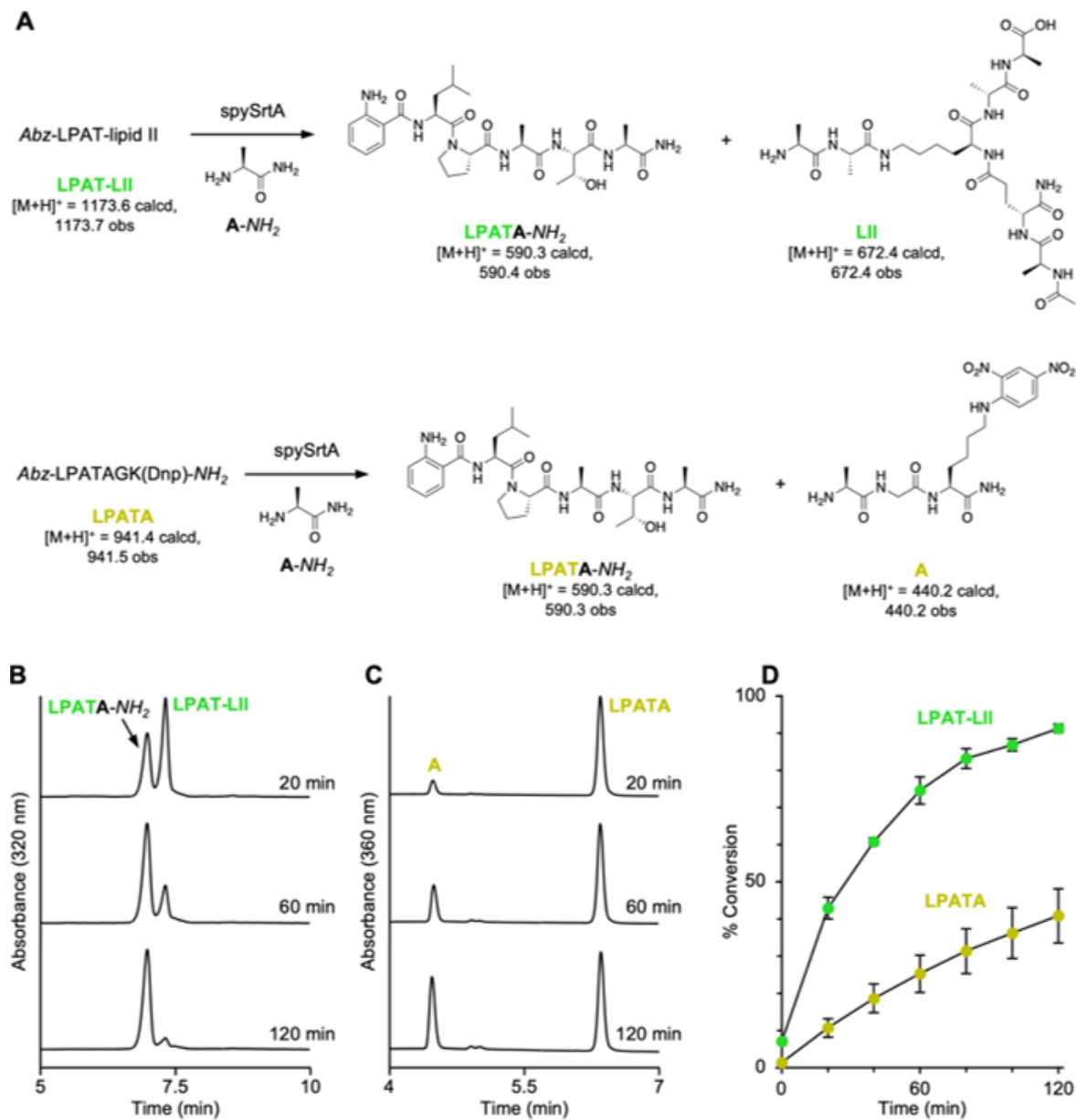
126

Figure A5. Comparison of transacylation reactions with LPAT-LII and LPATA. (A) Model transacylation reactions of *Abz*-LPAT-lipid II (LPAT-LII) or *Abz*-LPATAGK(Dnp)-*NH₂* (LPATA) with excess alanine amide (A-*NH₂*) in the presence of spySrtA. Conditions: 50 µM LPAT-LII/LPATA peptide, 5 mM A-*NH₂*, 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).
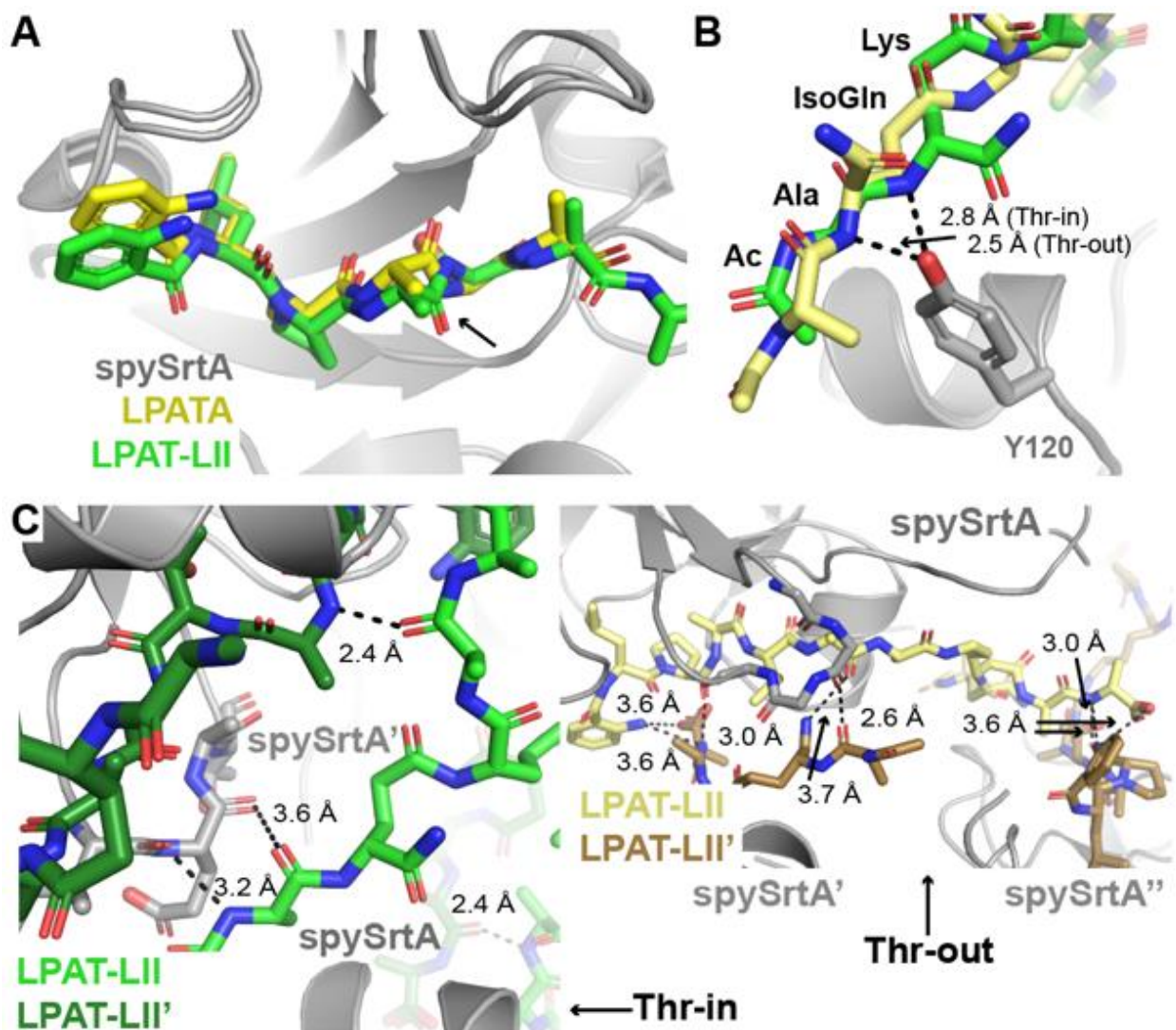
Figure A6. Structure of spySrtA bound to LPAT-LII peptide. The structures in these figures are rendered similarly: the spySrtA protein is shown as gray cartoon, with any side chains or highlighted areas as sticks and colored by heteroatom (N=blue, O=red). The LPAT-LII molecules are in green sticks and 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 most significant 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)**.
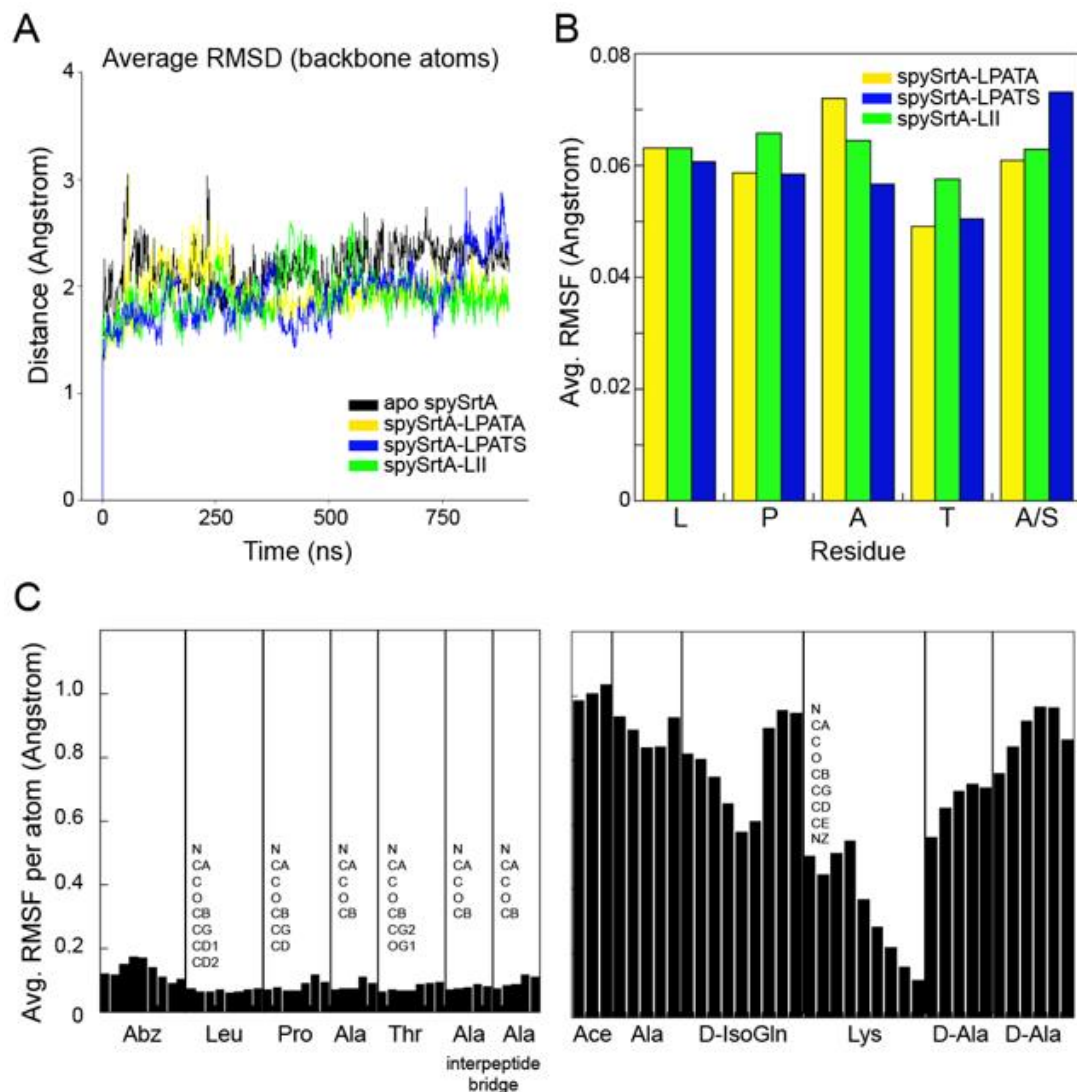
128

Figure A7. 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 Cε atom of the sidechain.
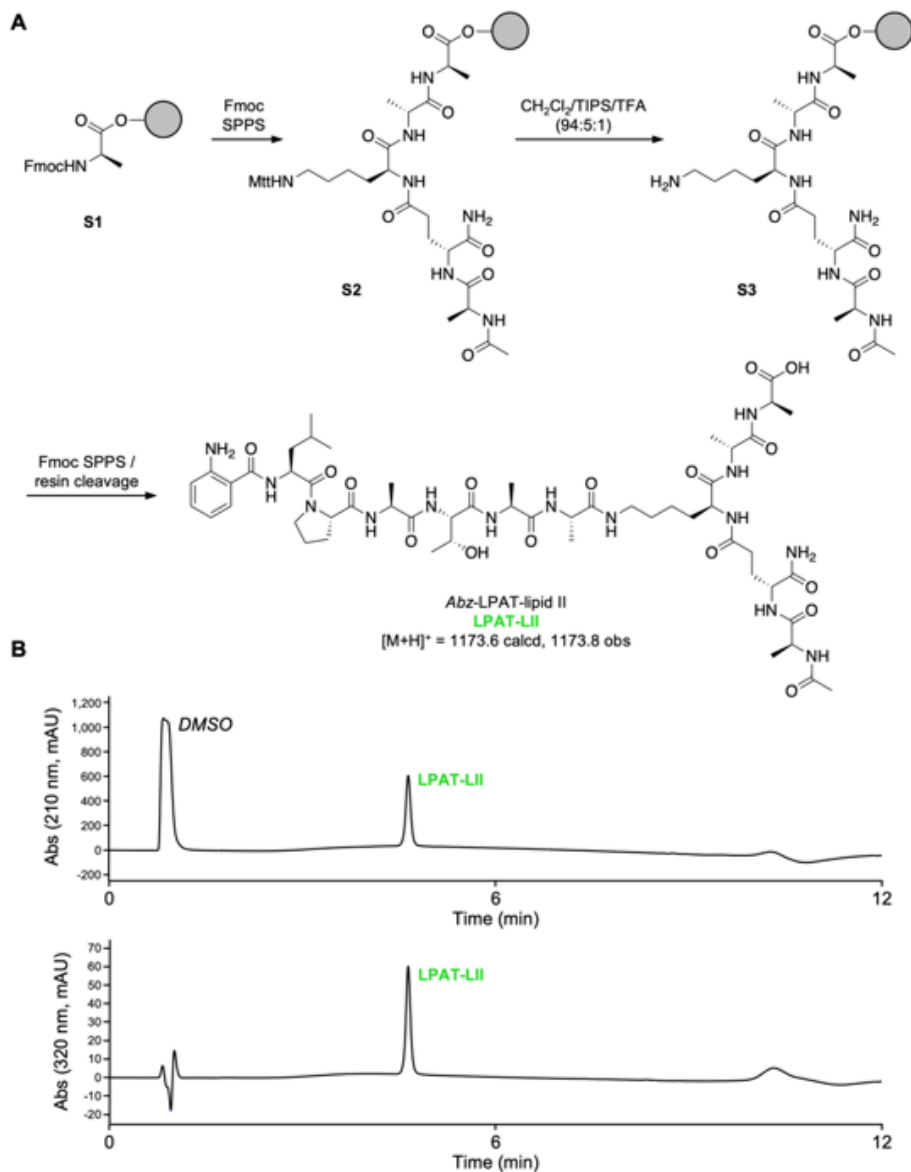
129

Figure A8. 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 A1. Details of the molecular dynamics simulation size.

System	Total number of atoms	Cubic box dimensions [nm]	Simulation time [ns]
SpySrtA (Apo)[a]	34123	7.0224	896.6
SpySrtA-LPATA[b]	34118	7.0186	905.0
SpySrtA-LPATS[c]	32523	6.9350	905.0
SpySrtA-LPAT-LII[d]	35836	7.1559	1005.0

[a]PDB ID 3FN5
[b]PDB ID 7S51
[c]PDB ID 7S40
[d]PDB ID 7T8Y

Table A2, part A: Summary of Sortase B *in vitro* Activity Assays

#	SrtB	Paper	[Enzyme]	Substrate (Motif in Bold)	[Substrate]	Nucleophile	[Nucleophile]
1	baSrtB	Maresso et al. (2006)	10 µM	2-aminobenzoyl-KTD**NPKTG**DEA-diaminoproprionic acid-dinitrophenyl-NH2	50 µM	water (hydrolysis)	NA
2	baSrtB	Puorger et al. (2017)	50 µM	FITC-D**NPNTG**DE FITC-**NPNTG**D *Performed substitutions at: P5, P4, P3, P2, and P1	1 mM	GGGK-biotin (biotin attached to Lys amino group)	250 µM
3	cdSrtB	Chambers et al. (2015)	unspecified	Abz-**SPKTG**-Dap(Dnp); Abz-**NVQTG**-Dap(Dnp)	0-40 nmol	m-diaminopimelic acid	equimolar to substrate
4	cdSrtB	Donahue et al. (2014)	10 µM	dabcyl-PV**PPKTG**DS-edans dabcyl-SD**SPKTG**DN-edans dabcyl-IH**SPQTG**DV-edans	20 µM	water (hydrolysis)	NA
5	cdSrtB	Kang et al. (2020)	120 µM	dabcyl-PV**PPKTG**DSTTIIGE-edans	20 µM	unspecified	
6	cdSrtB	Yin et al. (2016)	240 µM (optimal)	5-[(2-aminoethyl) amino] napthalene-1-sulfonic acid-**PPKTG**-4-([4-(dimethylamino) phenyl]azo benzoic acid	20 µM (optimal)	unspecified	
7	cdSrtB	van Leeuwen et al. (2014)	3 µg	KIVK**SPKTG**DETQLMK KPPV**PPKTG**DSTTIGK KPPV**SPKTG**DSTTIGK KIV**SPXTG**DETQLMK; X = Pro, Gln, Met, Phe, Arg	50 µM	tetraglycine pentaglycine Ala-D-Glu-2,6-diaminopimelic acid	1 mM
8	saSrtB	Jacobitz et al. (2014)	100 µM	SNKDKVE**NPQTN**AGT	200 µM	GGGGG	2 mM
9	saSrtB	Kang et al. (2006)	5 µM	dabcyl-**NPQTN**-Edans	10 µM	unspecified	
10	saSrtB	Mazmanian et al. (2002)	6 µM	dabcyl-KVE**NPQTN**AGT-edans	3 µM	unspecified	

Table A2, part A: Published sortase B *in vitro* activity assay conditions and results. Substrates testing for cross-reactivity with sortase A sequences did not indicate catalysis and were excluded.

Table A2, part B: Summary of Sortase B *in vitro* Activity Assays

#	Observations (Method)	Reaction Time	Temperature	Buffer Conditions
1	Substrate cleavage (fluorescence; HPLC; MALDI-MS)	18 hr	37° C	150 mM NaCl, 5 mM CaCl2, 50 mM Tris-HCl, pH 7.5
2	Substrate cleavage and ligation (HPLC; LCMS); additional observations for varied amino acid residues at investigated positions and substrate length; most product formation with P2 Asn	24 hr	25° C	50 mM Tris-HCl (pH 8.2), 150 mM NaCl, 10 mM CaCl2
3	SPKTG substrate: Substrate cleavage and ligation (HPLC; LCMS) NVQTG substrate: No activity	*unspecified*	*unspecified*	25 mM HEPES pH 7.5, 150 mM NaCl
4	All substrates: Substrate cleavage (fluorescence; MALDI-MS)	48 hr	37° C	5 mM CaCl2, 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM DTT
5	Substrate cleavage (fluorescence; N-terminal product LCMS)	24 hr	37° C	10 mM HEPES pH 7.4, 150 mM NaCl
6	Substrate cleavage (fluorescence)	48 hr	37° C	10 mM HEPES pH 7.4, 150 mM NaCl
7	KIVK**SPKTG**DETQLMK + tetraglycine: Substrate cleavage and ligation (MALDI-TOF) KIVK**SPKTG**DETQLMK + pentaglycine: Substrate cleavage and ligation (MALDI-TOF) KIVK**SPKTG**DETQLMK + Ala-D-Glu-2,6-diaminopimelic acid: Substrate cleavage and ligation (MS/MS, consistent with ligation of P1 Thr to the amino group of 2,6-diaminopimelic acid) KPPV**PPKTG**DSTTIGK + tetraglycine: Substrate cleavage and ligation (MALDI-TOF) KPPV**SPKTG**DSTTIGK + tetraglycine: Substrate cleavage and ligation (MALDI-TOF) KIV**SPXTG**DETQLMK (all variants) + tetraglycine: Substrate cleavage and ligation (MS/MS)	overnight	37° C	50 mM Tris-HCl (pH7.5), 2 mM DTT
8	Substrate cleavage and ligation (HPLC; MALDI-TOF)	24 hr	37° C	300 mM Tris-HCl, 150 mM NaCl
9	Substrate cleavage (fluorescence)	1 hr	37° C	50 mM Tris-HCl (pH 7.5), 5 mM CaCl2, 150 mM NaCl
10	Substrate cleavage (fluorescence)	2 hr	37° C	50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM [2-(trimethylammonium)ethyl]methanethiol-sulfonate or 10 mM DTT

Table A2, part B: Published sortase B *in vitro* activity assay conditions and results. Substrates testing for cross-reactivity with sortase A sequences did not indicate catalysis and were excluded.

133

CdSrtB Concentration (µM)	Substrate Concentration (µM)	pH	Calcium
10	50	7.5	Tested
50	175	7.5	-
35	35	7.0, 7.5, 8.0, 8.5, 9.0	-
120	20	7.5	-
20	20	7.5	-
40	20	7.5	-
80	20	7.5	-
140	20	7.5	-

Table A3. Reaction conditions for cdSrtB FRET-based assays. The cdSrtB reaction conditions tested in FRET-based assays. CdSrtB activity was not observed under any conditions.
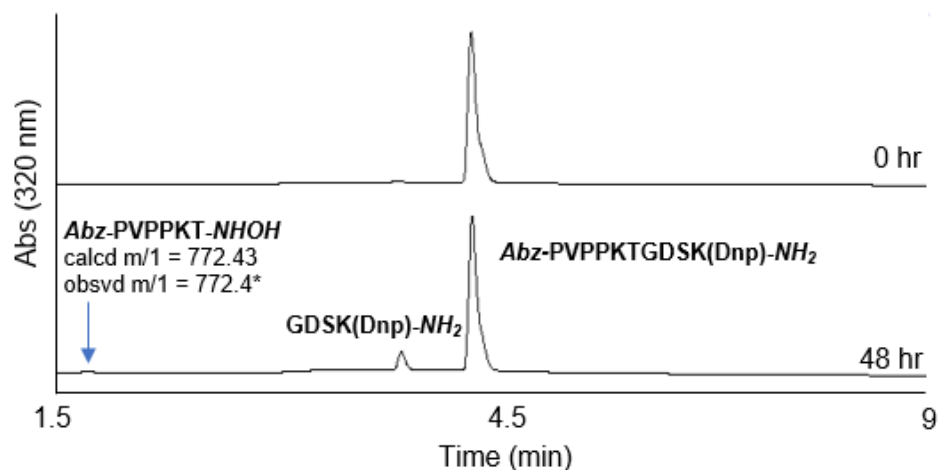

Figure A9. Indication of the N-terminal ligation product in a 48-hour cdSrtB HPLC assay. A small peak is discernable in the 320 nm chromatogram after 48 hours of incubation, and the mass spectra of this peak indicated a mass charge of 772.4, in good agreement with the calculated value for the N-terminal product. (*) However, the signal was very low and did not rise far above background levels. The identity of ligation products from cdSrtB activity should be verified.

The Effect of Residual TCEP on baSrtB Activity. Original sortase B stocks were made in buffers containing 1 mM TCEP. The high concentrations of enzyme used in reaction mixtures, however, meant that residual TCEP in reaction mixtures was high, often approximately 0.5 mM. An activity assay was performed to investigate the effect of 0.5 mM TCEP in reaction mixtures.

TCEP was removed from baSrtB stocks (100 µM baSrtB, 50 Tris pH 7.5, 150 mM NaCl) by dialysis. An assay comparing the activity baSrtB pulled from a stock containing TCEP and the activity of a stock from which TCEP was removed by dialysis was performed. The conditions of the reactions were 50 µM baSrtB, 100 µM Abz-DNPKTGDE-KDnp, 5 mM NH_2OH, 1:10 dilution of 10x sortase reaction buffer (500 mM Tris pH 7.5, 1500 mM NaCl). Reactions (n = 2) with and without TCEP progressed at similar rates (Figure A10). Although the results of this experiment indicate that the residual TCEP in sortase B reactions may not have impacted reaction rates, all enzyme stocks used in reactions outside of the preliminary sortase B reactions were made without using TCEP.
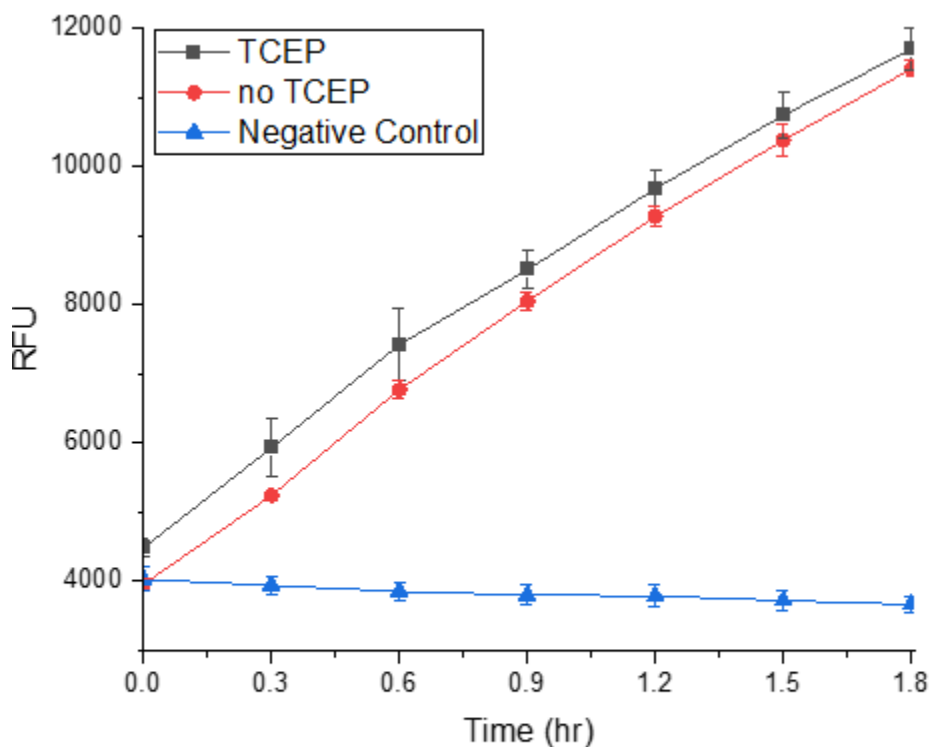
Figure A10. BaSrtB reactions with and without TCEP appear to progress at similar rates. A FRET-based baSrtB activity assay shows similar levels of activity with 0.5 mM residual TCEP (black squares) and without (red circles).

Fluorescence suppression in FRET-based assays. We noticed that reaction solutions evaporated when FRET-based reactions were monitored over long periods of time (>2 hours) in uncovered 96-well reaction plates. We became concerned that this may over time change the measured fluorescence as high concentrations of fluorophore can suppress fluorescence readings. To investigate this, we performed a series of experiments over 2 time frames using wild-type (WT) baSrtB and the more active baSrtB mutant A241K. Reaction solutions (50 µM enzyme, 200 µM Abz-DNPKTGDE-KDnp, 5 mM hydroxylamine, 50 mM Tris pH 8.0, 150 mM NaCl) we mixed in 2 mL microcentrifuge tubes (n = 1). A 100 µL aliquot of each reaction was applied to a well of a black, flat-bottom Costar 96-well plate immediately after reactions were started (t = 0 hr) and was monitored as normal. The remaining reaction volume was left in the microcentrifuge tube, which was capped and left at room temperature. After ½ of the studied time frame had passed, a fresh 100 µL aliquot of reaction solution was added to a clean well of the 96-well plate. Reactions were then monitored for the remainder of the reaction period. After which, reactions were diluted 1:2, 1:4, and, for the longer time frame, 1:8 to study. The results show that fluorescence is suppressed slightly (within 15%) at 1.5 hours (Figure A11A) and that after 3 hours diluting samples 1:2 does not reduce fluorescence readings (Figure A11B). Fluorescence is suppressed >30% after 6 hours (Figure A12) and after 1:2 and 1:4 dilutions after 11 hours result in fluorescence increases, indicating a strong fluorescence suppression effect related to high concentration of fluorophore.
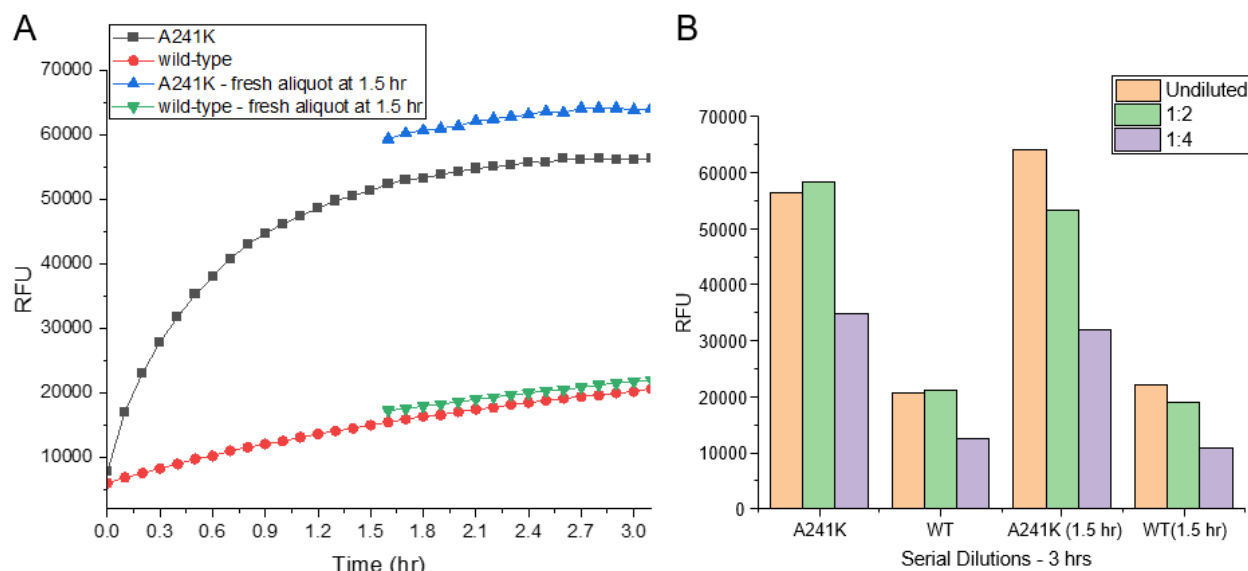
Figure A11. Fluorescence suppression effects emerge within 3 hours. (A) The fluorescence of reactions monitored in uncovered reaction wells appears to be suppressed after 1.5 hours when compared to fresh aliquots of reaction mixtures applied at 1.5 hours. **(B)** After 3 hours, 1:2 dilutions of uncovered reactions do not show reduced fluorescence. After 1.5 hours, fluorescence is reduced when a 1:2 dilution is performed.
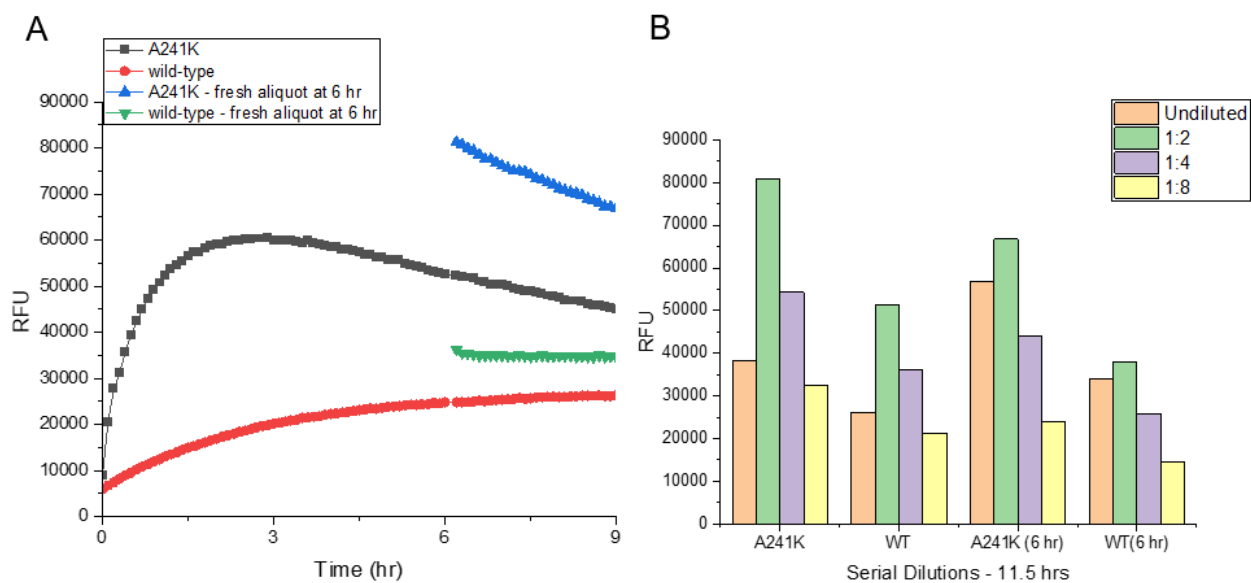

Figure A12. Fluorescence suppression effects emerge within 3 hours. (A) The fluorescence of reactions monitored in uncovered reaction wells is markedly suppressed after 6 hours when compared to fresh aliquots of reaction mixtures applied at 6 hours. **(B)** After more than 11 hours, 1:2 and 1:4 dilutions of uncovered reactions increase in fluorescence. After 6 hours, fluorescence is increased when a 1:2 dilution is performed.

137

Protein sequences

>baSrtB_35-254

MESSHHHHHHENLYFQSIFMDYYENRKVMAEAQNIYEKSPMEEQSQDGEV

RKQFKALQQINQEIVGWITMDDTQINYPIVQAKDNDYYLFRNYKGEDMRA

GSIFMDYRNDVKSQNRNTILYGHRMKDGSMFGSLKKMLDEEFFMSHRKLY

YDTLFEGYDLEVFSVYTTTTDFYYIETDFSSDTEYTSFLEKIQEKSLYKT

DTTVTAGDQIVTLSTCDYALDPEAGRLVVHAKLVKRQ

>saSrtB_30-244

MESSHHHHHHENLYFQSIEDKQERANYEKLQQKFQMLMSKHQEHVRPQFE

SLEKINKDIVGWIKLSGTSLNYPVLQGKTNHDYLNLDFEREHRRKGSIFM

DFRNELKNLNHNTILYGHHVGDNTMFDVLEDYLKQSFYEKHKIIEFDNKY

GKYQLQVFSAYKTTTKDNYIRTDFENDQDYQQFLDETKRKSVINSDVNVT

VKDKIMTLSTCEDAYSETTKRIVVVAKIIKVS

>cdSrtB_26-225

MESSHHHHHHENLYFQSSKLTKYNHDTKISSELQKKEYKKEDLSKINSDF

KFWLSVENTNINYPVVQSKDNSYYLDKDFYKKDSISGTLFMDYRNKSIDD

KNIIIYGHNMKNKTMFNNLNKFKDADFFKKNNKIKITLNGKEFLYDVFSA

YIVESDYDYLKTNFNNESDYQNYINDITSKSLYKSPIKVNSNDKIVTLST

CTYEFDDARMVIHGRLI

>lmSrtB_26-246

MESSHHHHHHENLYFQSIGMELYENKHNQTILDDAKAVYTKDAATTNVNG

EVRDELRDLQKLNKDMVGWLTIIDTEIDYPILQSKDNDYYLHHNYKNEKA

RAGSIFKDYRNTNEFLDKNTIIYGHNMKDGSMFADLRKYLDKDFLVAHPT

138

FSYESGLTNYEVEIFAVYETTTDFYYIETEFPETTDFEDYLQKVKQQSVY

TSNVKVSGKDRIITLSTCDTEKDYEKGRMVIQGKLVTK

>**C233A_baSrtB_NPTKG** **(N-terminal** **extension;** **DNPKTGDEARI)**

MESSHHHHHHENLYFQSDNPKTGDEARIYYENRKVMAEAQNIYEKSPMEE

QSQDGEVRKQFKALQQINQEIVGWITMDDTQINYPIVQAKDNDYYLFRNY

KGEDMRAGSIFMDYRNDVKSQNRNTILYGHRMKDGSMFGSLKKMLDEEFF

MSHRKLYYDTLFEGYDLEVFSVYTTTTDFYYIETDFSSDTEYTSFLEKIQ

EKSLYKTDTTVTAGDQIVTLSTADYALDPEAGRLVVHAKLVKRQ

>**C233A_baSrtB_NPTKG_C233** **(N-terminal** **extension;** **DNPKTGDEARI),** **Ala-to-Cys**
revertant

MESSHHHHHHENLYFQSDNPKTGDEARIYYENRKVMAEAQNIYEKSPMEE

QSQDGEVRKQFKALQQINQEIVGWITMDDTQINYPIVQAKDNDYYLFRNY

KGEDMRAGSIFMDYRNDVKSQNRNTILYGHRMKDGSMFGSLKKMLDEEFF

MSHRKLYYDTLFEGYDLEVFSVYTTTTDFYYIETDFSSDTEYTSFLEKIQ

EKSLYKTDTTVTAGDQIVTLSTCDYALDPEAGRLVVHAKLVKRQ

>**Self-Cleaved_C233A_baSrtB_NPTKG_C233** **(N-terminal** **extension;** **DNPKTGDEARI**
cleaved by baSrtB)

GDEARIYYENRKVMAEAQNIYEKSPMEEQSQDGEVRKQFKALQQINQEIVGWITMDDTQINYPI

VQAKDNDYYLFRNYKGEDMRAGSIFMDYRNDVKSQNRNTILYGHRMKDGSMFGSLKKMLDEE

FFMSHRKLYYDTLFEGYDLEVFSVYTTTTDFYYIETDFSSDTEYTSFLEKIQEKSLYKTDTTVTA

GDQIVTLSTCDYALDPEAGRLVVHAKLVKRQ

Sequences Submitted to AlphaFold2 (Galaxy)

Full-length *B. anthracis* IsdC joined to the N-terminus of full-length baSrtB through a glycine-serine linker. This sequence was submitted to AlphaFold2 (Galaxy) to generate the substrate-bound baSrtB model referenced in chapters 3-5. NPKTG pentapeptide motif highlighted orange; Gly-Ser linker in red text; baSrtB sequence highlighted blue.

MKLADGTYDINYVIQKAENDSASMANDYFEKPAKLIVKNGEMRVQVPMNHSAWITEFKAPEN
GNFVDAKVVSKDESADKRTVEFKVDDLSKPEAVKIHVVVPNANYDHHYTIRFAFDANVKAVG
GDNGVAATTKNNDQAKTDTQVKEEKTKVESKETAKEVNKETKNENGKAEKTD<mark>NPKTG</mark>DEA
RIGLFAALILISGVFLIGGGGSGGGGS<mark>IFFQRILTVVFLGTFFYSVYELGDIFMDYYENRKVMAE
AQNIYEKSPMEEQSQDGEVRKQFKALQQINQEIVGWITMDDTQINYPIVQAKDNDYYLFRNYK
GEDMRAGSIFMDYRNDVKSQNRNTILYGHRMKDGSMFGSLKKMLDEEFFMSHRKLYYDTLF
EGYDLEVFSVYTTTTDFYYIETDFSSDTEYTSFLEKIQEKSLYKTDTTVTAGDQIVTLSTCDYAL
DPEAGRLVVHAKLVKRQ</mark>

Full-length *L. monocytogenes* Lmo2185 joined to the N-terminus of full-length lmSrtB through a glycine-serine linker. This sequence was submitted to AlphaFold2 (Galaxy) to generate the substrate-bound lmSrtB model referenced in chapter 3. NAKTN pentapeptide motif highlighted orange; Gly-Ser linker in red text; lmSrtB sequence highlighted blue.

MKKLWKKGLVAFLALTLIFQLIPGFASAADSRLKDGGEYQVQVNFYKDNTGKTTKESSEADK
YIDHTATIKVENGQPYMYLTITNSTWWQTMAVSKNGTRPEKPAQADVYQDRYEDVQTVSTDA
AKDTRVEKFKLSSLDDVIFSYMHIKVDAISYDHWYQVDLTIDPSTFKVISEPAVTTPVTLSDGIY
TIPFVAKKANDDSNSSMQNYFNNPAWLKVKNGKKMVAMTVNDNKTVTALKTTLAGTLQDVK
VVSEDKDANTRIVEFEVEDLNQPLAAHVNYEAPFNGSVYKGQADFRYVFDTAKATAASSYPG
SDETPPVVNPGETNPPVTKPDPGTTNPPVTTPPTTPSKPAVVDPKNLLNNHTYSIDFDVFKDG
TTETSMMESYVMKPALIKVENNQPYVYLTLTNSSWIKTFQYKVNGVWKDMEVVSGDINKNTR
TVKYPVKDGTANTDVKTHVLIEDMPGFSYDHEYTVQVKLNAATIKDITGKDVTLKEPVKKDILN

TGNVASNNNAGPKLAKPDFDDTNSVQKTASKTEK<mark>NAKTN</mark>DSSSMVWYITLFGASFLYLAYRL
GGGGSGGGGS<mark>LTLVVLGVFLFSGWKIGMELYENKHNQTILDDAKAVYTKDAATTNVNGEVR
DELRDLQKLNKDMVGWLTIIDTEIDYPILQSKDNDYYLHHNYKNEKARAGSIFKDYRNTNEFLD
KNTIIYGHNMKDGSMFADLRKYLDKDFLVAHPTFSYESGLTNYEVEIFAVYETTTDFYYIETEFP
ETTDFEDYLQKVKQQSVYTSNVKVSGKDRIITLSTCDTEKDYEKGRMVIQGKLVTK</mark>

Full-length *L. monocytogenes* Lmo2186 joined to the N-terminus of full-length lmSrtB through a glycine-serine linker. This sequence was submitted to AlphaFold2 (Galaxy) to generate the substrate-bound lmSrtB model referenced in chapter 3. NPKSS pentapeptide motif highlighted orange; Gly-Ser linker in red text; lmSrtB sequence highlighted blue.

MKKVLVFAAFIVLFSFSFLSTGLTAQAALKDGTYSVDYTVIQGDSDSASMANDYFDKPATVTV
NGGKSTVSLQVNHSKWITGLWVEGNAVSVTSKNASSDTRKVSFPVSTLSNPVNAKIKVDIDD
DDLNYHHEYQIKLRFDEGSAKALAGAVKSSDNNTTTPATKSDSSNKVT<mark>NPKSS</mark>DSSQMFLYG
IIFVATGAGLILLGGGGSGGGGS<mark>LTLVVLGVFLFSGWKIGMELYENKHNQTILDDAKAVYTKD
AATTNVNGEVRDELRDLQKLNKDMVGWLTIIDTEIDYPILQSKDNDYYLHHNYKNEKARAGSI
FKDYRNTNEFLDKNTIIYGHNMKDGSMFADLRKYLDKDFLVAHPTFSYESGLTNYEVEIFAVY
ETTTDFYYIETEFPETTDFEDYLQKVKQQSVYTSNVKVSGKDRIITLSTCDTEKDYEKGRMVIQ
GKLVTK</mark>

Full-length *S. aureus* IsdC joined to the N-terminus of full-length saSrtB through a glycine-serine linker. This sequence was submitted to AlphaFold2 (Galaxy) to generate the substrate-bound saSrtB model referenced in chapter 3. NPQTN pentapeptide motif highlighted orange; Gly-Ser linker in red text; saSrtB sequence highlighted blue.

MKNILKVFNTTILALIIIIATFSNSANAADSGTLNYEVYKYNTNDTSIANDYFNKPAKYIKKNGKL
YVQITVNHSHWITGMSIEGHKENIISKNTAKDERTSEFEVSKLNGKIDGKIDVYIDEKVNGKPFK
YDHHYNITYKFNGPTDVAGANAPGKDDKNSASGSDKGSDGTTTGQSESNSSNKDKVE<mark>NPQT</mark>

NAGTPAYIYAIPVASLALLIAITLFV GGGGSGGGGS FLTIVQILLVVIIIIFGYKIVQTYIEDKQERA
NYEKLQQKFQMLMSKHQEHVRPQFESLEKINKDIVGWIKLSGTSLNYPVLQGKTNHDYLNLD
FEREHRRKGSIFMDFRNELKNLNHNTILYGHHVGDNTMFDVLEDYLKQSFYEKHKIIEFDNKY
GKYQLQVFSAYKTTTKDNYIRTDFENDQDYQQFLDETKRKSVINSDVNVTVKDRIMTLSTCED
AYSETTKRIVVVAKIIKVS

Full-length *C. difficile* CD2831 joined to the N-terminus of full-length cdSrtB through a glycine-serine linker. This sequence was submitted to AlphaFold2 (Galaxy) to generate the substrate-bound cdSrtB model referenced in chapter 3. PPKTG pentapeptide motif highlighted orange; Gly-Ser linker in red text; cdSrtB sequence highlighted blue.

MKKGNRKALLISLIMILSMVVSTIYPTVSYASELGENSQIQSGSTNSSTGEEKESDNKKPEQTP
EKDKATDNKKPEQTPEEDKSTDNKKSEQALEDEKPLDNKNTEKTPEEDNLLEDENLLKVLEE
ELNEENEDYGFVVKINNNTIETESMKKISFNLTYTPTSKGIQAGDSITFKVPDVFNKVNLDYTSE
CFDKTESNGEYTLTFRELPNGQSVMQGKIGLEAYVKKVDEDTNAKIHIETTGKIESGSGDIDVE
IKPGDKTDVPDAKGTLKKLVEGKKSTTVFMPVKNKDINYSIQVNEKQEELKDIILYDELPEGLT
LINGSVSVVTSDGKEVSDFNIEQSKNSISVNFGNIDKSYTVKYKARISDKNAKHGNKYKNVARI
ESDGKKIQEDDATVSIFDRGDDYLLTKGHSGATNITQVGQVINYQISINDDKSPISNVVITDNIPE
GMRLTTSGEAGHDFRVVEIPMNGSWTPWSKEKIANNISYKVEEKRNESGQVDKVITGFTINLS
KEEVESKFFIAYTLKVISIEDSYINRAVLDANNSEIDKNDEINFKKNSGLISAKKEVDKKVLNSSD
NQIVKYKINMSTYGVYDAGQVNLLDEVNSVLEISNIKYSDNLELKKEAGDGKNTIRLVNKYEFK
QIKEGEPVQSWVTFDANFTNVKVGETIRNVAQINGSSPPGVETTKQGYAFEAKKVDALDKNV
LSGAKFNLEDAFGNIVVKDLVSDEDGIIQSSVKNPGTYYLVEIMAPRGYEKLKDKVKVEIGNED
IGKIVDIGNIENLKQENPPVNPPIPPDTDEPIVNPPVPPSTDKPRKPSSSSDTENTIVINPPVPPS
EDIINPPIPEVLNPPVPPSEEMIETPVKQIIPIPEVVKPSVSEEKNNKVKDDTLVNPPV PPKTG DS
TTIIGEILLVIGAIVGLIVLRRNKNTN GGGGSGGGG MKKLYRIVINIILVLVILYSGFNIYSKLTKYN
HDTKISSELQKKEYKKEDLSKINSDFKFWLSVENTNINYPVVQSKDNSYYLDKDFYKKDSISGT

LFMDYRNKSIDDKNIIIYGHNMKNKTMFNNLNKFKDADFFKKNNKIKITLNGKEFLYDVFSAYIV
ESDYDYLKTNFNNESDYQNYINDITSKSLYKSPIKVNSNDKIVTLSTCTYEFDDARMVIHGRLI

>baSrtBswapAureus_IsdC_galaxy_SJ_A

MKLADGTYDINYVIQKAENDSASMANDYFEKPAKLIVKNGEMRVQVPMNHSAWITEFKAPEN
GNFVDAKVVSKDESADKRTVEFKVDDLSKPEAVKIHVVVPNANYDHHYTIRFAFDANVKAVG
GDNGVAATTKNNDQAKTDTQVKEEKTKVESKETAKEVNKETKNENGKAEKTDNPKTGDEA
RIGLFAALILISGVFLIGGGGSGGGGSIFFQRILTVVFLGTFFYSVYELGDIFMDYYENRKVMAE
AQNIYEKSPMEEQSQDGEVRKQFKALQQINQEIVGWITMDDTQINYPIVQAKDNDYYLFRNYK
GEDMRAGSIFMDYRNDVKSQNRNTILYGHRMKDGSMFGSLKKMLDEEFFMSHRKLYYDTLF
EGYDLEVFSVYTTTTDFYYIETDFSSDTEYTSFLEKIQEKSLYKTDTTVTAGDQIVTLSTCEDAY
SETTKRLVVHAKLVKRQ

>AlphaFold_SJ_baSrtBswapDifficile_IsdC_A

MKLADGTYDINYVIQKAENDSASMANDYFEKPAKLIVKNGEMRVQVPMNHSAWITEFKAPEN
GNFVDAKVVSKDESADKRTVEFKVDDLSKPEAVKIHVVVPNANYDHHYTIRFAFDANVKAVG
GDNGVAATTKNNDQAKTDTQVKEEKTKVESKETAKEVNKETKNENGKAEKTDNPKTGDEA
RIGLFAALILISGVFLIGGGGSGGGGSIFFQRILTVVFLGTFFYSVYELGDIFMDYYENRKVMAE
AQNIYEKSPMEEQSQDGEVRKQFKALQQINQEIVGWITMDDTQINYPIVQAKDNDYYLFRNYK
GEDMRAGSIFMDYRNDVKSQNRNTILYGHRMKDGSMFGSLKKMLDEEFFMSHRKLYYDTLF
EGYDLEVFSVYTTTTDFYYIETDFSSDTEYTSFLEKIQEKSLYKTDTTVTAGDQIVTLSTCTYEF
DDARLVVHAKLVKRQ

>baSrtBswapMono_IsdC_A

MKLADGTYDINYVIQKAENDSASMANDYFEKPAKLIVKNGEMRVQVPMNHSAWITEFKAPEN
GNFVDAKVVSKDESADKRTVEFKVDDLSKPEAVKIHVVVPNANYDHHYTIRFAFDANVKAVG

GDNGVAATTKNNDQAKTDTQVKEEKTKVESKETAKEVNKETKNENGKAEKTDNPKTGDEA

RIGLFAALILISGVFLIGGGGSGGGGSIFFQRILTVVFLGTFFYSVYELGDIFMDYYENRKVMAE

AQNIYEKSPMEEQSQDGEVRKQFKALQQINQEIVGWITMDDTQINYPIVQAKDNDYYLFRNYK

GEDMRAGSIFMDYRNDVKSQNRNTILYGHRMKDGSMFGSLKKMLDEEFFMSHRKLYYDTLF

EGYDLEVFSVYTTTTDFYYIETDFSSDTEYTSFLEKIQEKSLYKTDTTVTAGDQIVTLSTCDTEK

DYEKGRLVVHAKLVKRQ

The Effect of DMSO on baSrtB Activity. Peptide stock solutions typically contain DMSO to aid

solubility. An activity assay was performed using baSrtB and a DMSO concentration of 1%, 2%, and 5% (n = 2). Activity appears to be slightly inhibited at 5%, but is comparable in 1% and 2%. Sortase reactions presented in this work had residual DMSO <2%.
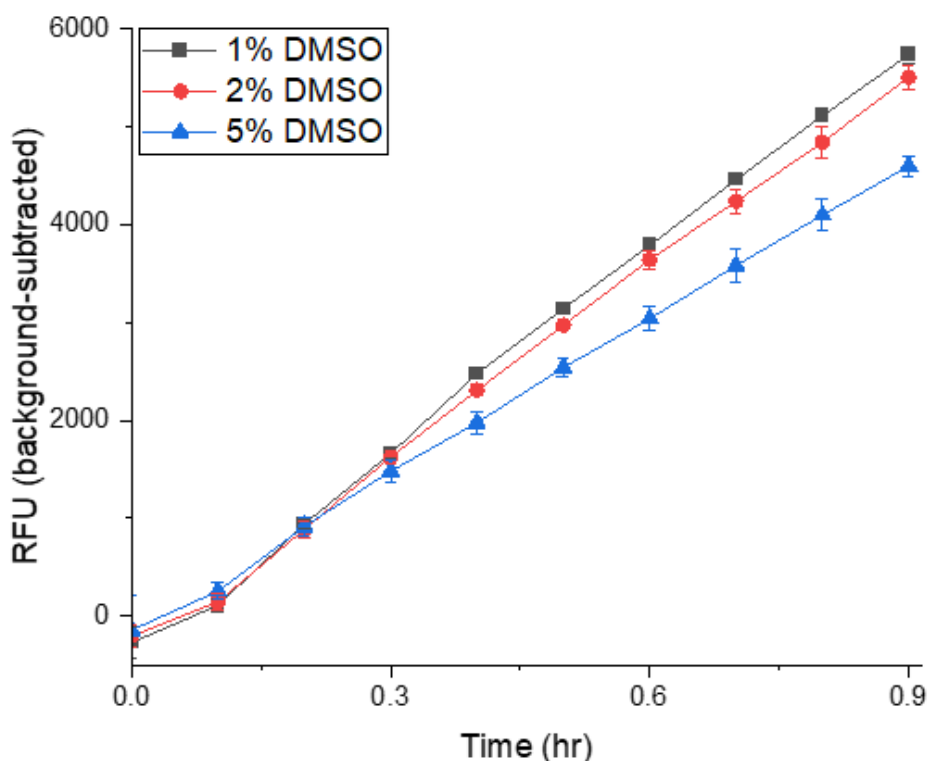
Figure A13. BaSrtB activity appears reduced when reaction conditions contain 5% DMSO. Peptide substrate stocks contain DMSO to aid solubility and residual DMSO remains in assays. Here, we test the effect of DMSO on baSrtB activity. Activity appears comparable at 1% and 2% DMSO but is reduced when DMSO levels rise to 5%.

Figure A14. Multiple sequence alignment of srtB enzymes. A multiple sequence alignment of wild-type sortase B homologs studied in this work. The conserved leucine (*) and serine (**) residues and the β7 - β8 loop (***) are shown boxed in red.
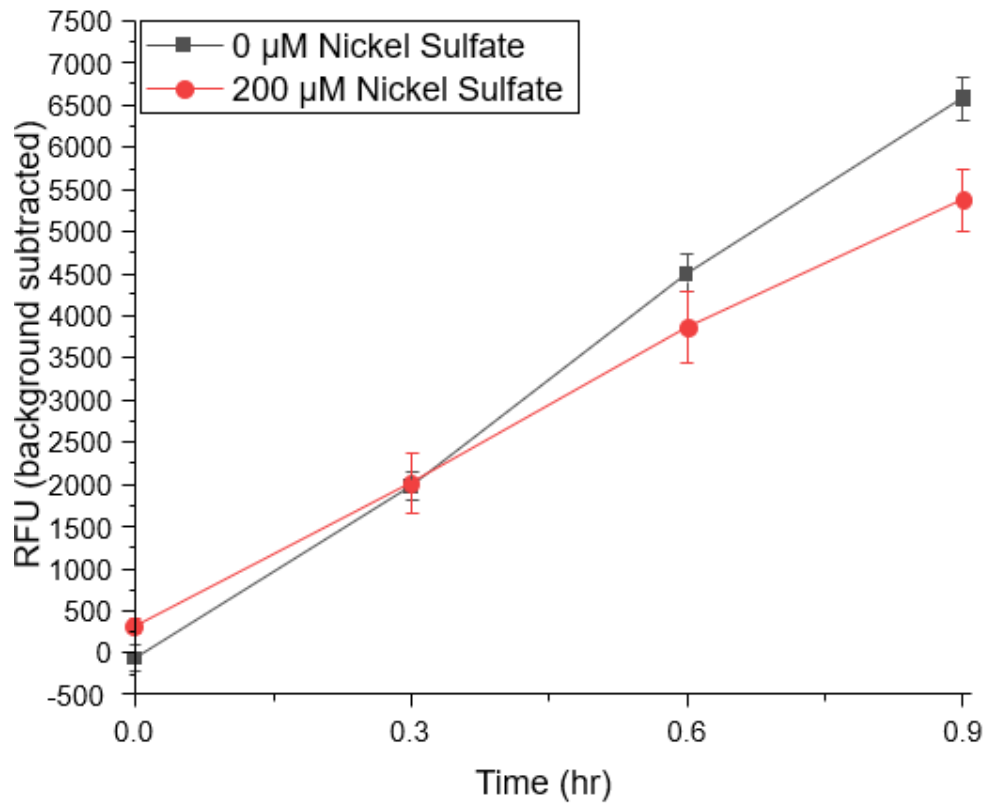
145

Figure A15. BaSrtB activity is inhibited in the presence of Ni²⁺. BaSrtB activity is slightly reduced in the presence of 200 µM Ni²⁺ (n = 2).
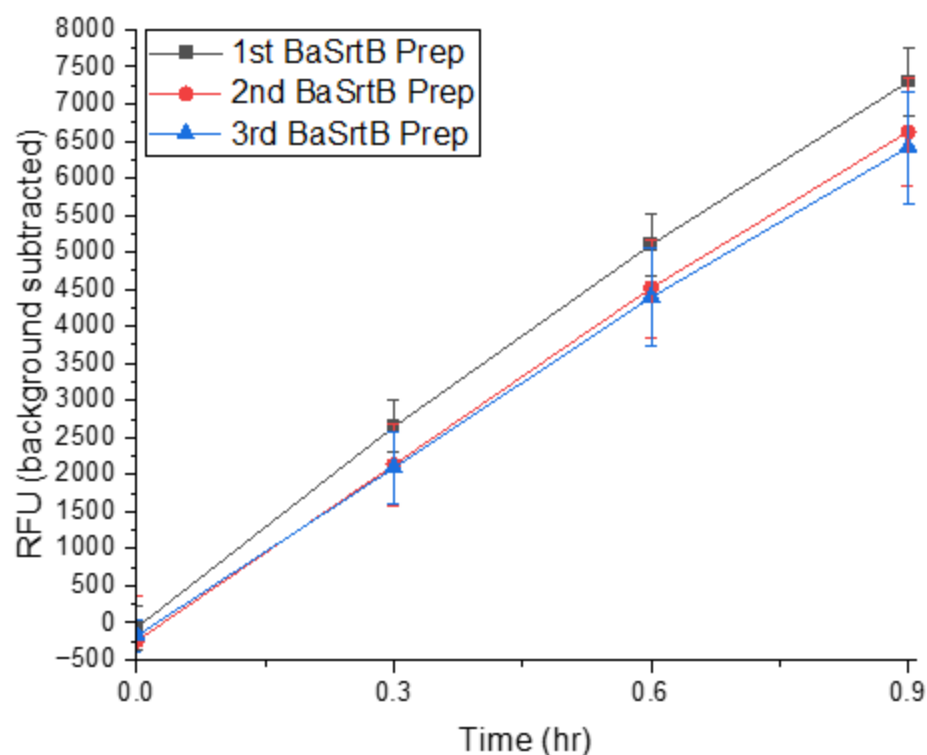
Figure A16. Activity of 3 baSrtB preps. The work presented in this thesis was completed using three separate baSrtB (wild-type) protein preparations. Briefly, the first prep (black squares) was used for preliminary assays and a portion of the wild-type controls for the loop chimera experiments. The 2nd prep (red circles) was used for substrate length experiments and for the wild-type controls for point mutants and some loop chimera experiments. The 3rd prep (blue triangles) was used for the nucleophile experiments.
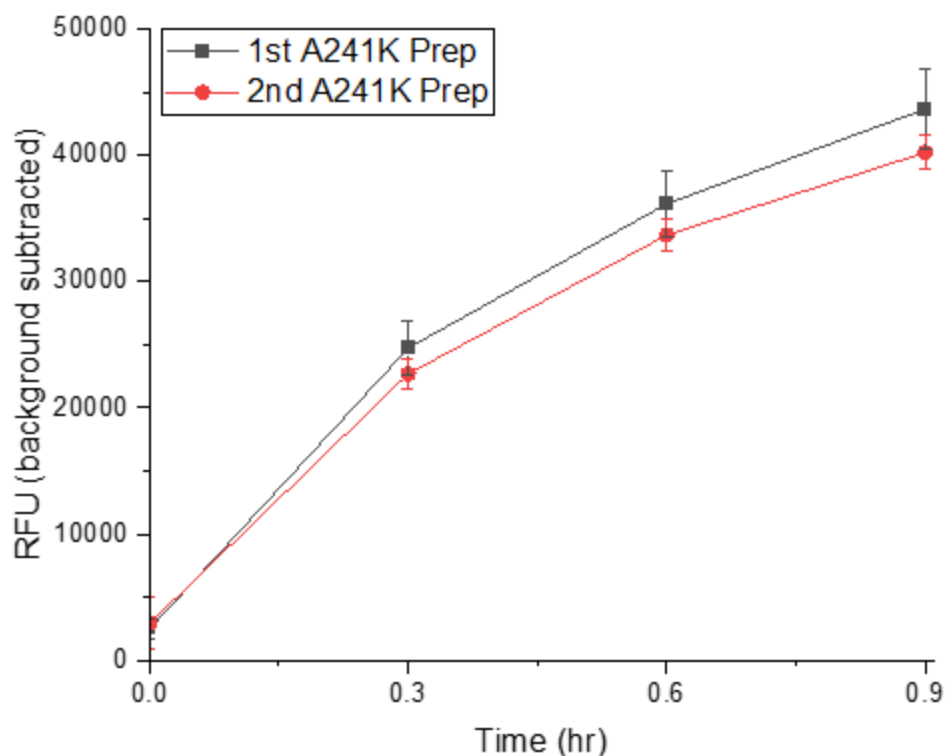
Figure A17. Activity of 2 A241K baSrtB preps. The work presented in this thesis was completed using two separate A241K baSrtB protein preparations. The first prep (black squares) was used for all experiments except the nucleophile assay, which was performed using the second prep (red circles).