



Western Washington University
Western CEDAR

WWU Graduate School Collection

WWU Graduate and Undergraduate Scholarship

Summer 2020

Engineering Class A Sortases: Activity and Selectivity of Hybrid and Ancestral Variants

Sarah Struyvenberg
sarahstruyvenberg@gmail.com

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

 Part of the [Chemistry Commons](#)

Recommended Citation

Struyvenberg, Sarah, "Engineering Class A Sortases: Activity and Selectivity of Hybrid and Ancestral Variants" (2020). *WWU Graduate School Collection*. 970.
<https://cedar.wwu.edu/wwuet/970>

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

**Engineering Class A Sortases: Activity and Selectivity of
Hybrid and Ancestral Variants**

By

Sarah A Struyvenberg

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

ADVISORY COMMITTEE

Dr. Jeanine Amacher, Chair

Dr. John Antos

Dr. Clint Spiegel

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.

Sarah A. Struyvenberg

8/4/20

**Engineering Class A Sortases: Activity and Selectivity of
Hybrid and Ancestral Variants**

A Thesis
Presented to
The Faculty of
Western Washington University

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

by
Sarah Struyvenberg
August 2020

Abstract

Bacterial sortases are cysteine transpeptidases that anchor virulence factors to the surface of bacterial cells. Sortases are a powerful tool utilized for protein engineering that allow researchers to modify proteins at the protein level, not the DNA level. However, important limitations to utilization of sortases for engineering purposes exist; namely, SrtA from *S. aureus* is a relatively modest enzyme compared to other SrtA enzymes and is very specific for the LPXTG motif. Previous work from our collaborators and others revealed that sortases from different species can recognize alternative sequences and that activities can vary widely. We were curious about how natural sequence variation in class A sortases affects activity and selectivity. To that end, a principle component analysis revealed that the structurally conserved β 7- β 8 substrate-interacting loop region may be a key component in substrate recognition and activity. We investigated this in two ways, by engineering eight *S. pneumoniae* β 7- β 8 loop variants with loop sequences from different bacterial species and by performing ancestral sequence reconstruction on extant class A sortase sequences. We then assayed all of our variants and found a SrtA construct, SPS_{faec} (*S. pneumoniae* core with a β 7- β 8 substrate-interacting loop from *E. faecalis*) which not only possessed an enhanced substrate promiscuity profile, recognizing seven 5th position substrates LPATGG, LPATSG, LPATAG, LPATVG, LPATTG, LPATNG, and LPATFG, but also displayed improved catalytic efficiency for all six of these substrates compared to the WT enzymes SrtA from *S. aureus* and SrtA from *S. pneumoniae*. Overall our engineered constructs provide further insight into the role of this β 7- β 8 substrate-interacting loop in class A sortases and provide additional framework for the design of sortases for future engineering purposes.

Acknowledgements

I express my deepest gratitude to my advisor Prof. Jeanine Amacher for her profound belief in me and my abilities as a researcher and her unwavering support and mentorship through both my Undergraduate and Graduate studies. My success and the completion of my thesis would not have been possible without her guidance and patience.

I am extremely grateful to my thesis committee member Prof. John Antos for the extensive sortase knowledge he has passed on to me in support of this project. This project would not have been possible without his guidance and contributions.

I extend my sincere thanks to my thesis committee member Prof. Clint Spiegel for his valuable advice and constructive criticism.

I cannot begin express my thanks to Izzi Piper who synthesized all the peptides for this project and helped develop the high throughput assay. I would not have been able to complete this project without her.

I express my sincere thanks to all members of my lab who contributed to this project. Special thanks to Jordan Valgardson who performed the PCA and reconstructed all of our ancestral sortase proteins. Alex Johnson, who helped me purify the enzymes for this project. Katherine Johnston for her assistance in both protein and peptide purification. I extend my thanks to current and past members of my lab for their support and helpful contributions. Melody Gao, Iain Mackley, Nick Pederson, and Haley Bamonte.

Special thanks to the members of the Antos Lab, past and present, who have contributed to this project, Nick Horvath, Nick McLaughlin, and Eddy Bartley.

I thank the members of the Spiegel Lab and Smirnov Lab for their advice, support, and company in the lab for all of these years.

I had the great opportunity to work with Helen Hobbs and Susan Marqusee at UC Berkeley for the HDX portion of this project. I thank them for the knowledge they have passed to me and their guidance.

I also gratefully acknowledge the help from members of the Chemistry Department for their assistance over the years I have attended Western Washington University. Special thanks to Spencer Anthony-Cahill, Jennifer Griffith, James Vyvyan, Sam Danforth, Gary Carlton, Steve Sible, and Alexi Guddal.

I thank the NSF and Western Washington University for funding my research fellowships and research assistantships for my Graduate Studies.

Finally, I thank my family and friends for their support and patience through the process of thesis writing and my graduate studies.

Table of Contents

Abstract	iv
Acknowledgements	v
List of Tables and Figures	vii
Introduction	1
Chapter 1: ‘Loop Swapped’ Engineered Sortase A.....	15
from <i>Streptococcus pneumoniae</i>	15
Introduction	16
1.1 Introduction to ‘Loop Swapped’ SrtA.....	16
Results and Discussion	18
1.2 Principle Component Analysis of Sortase A	18
1.3 Initial ‘Loop Swapped’ Constructs SAS _{Pneumoniae} and SPS _{Aureus}	19
1.4 High Throughput Fluorescence Assay Development.....	20
1.5 Selectivity and Activity of Initial ‘Loop Swapped’ Constructs.....	23
1.6 ‘Loop Swapped’ Complexes with New Sortase Homologues	25
1.7 Enzyme Inactivity.....	35
1.8 Preliminary HDX Experiments.....	40
1.9 Dermcidin Experiment.....	43
1.10 Expansion of Sortase Utilization and Concluding Remarks.....	44
Materials and Methods	47
Chapter 2: Ancestrally Reconstructed Sortase A.....	56
Introduction	57
2.1 Ancestral Sequence Reconstruction.....	57
Results and Discussion	59
2.2 Ancestral Constructs, Anc _{Staph} and Anc _{Strep}	59
2.3 Substrate Selectivity and Activity of Ancestral Constructs	62
2.4 Expansion into More Ancestral Relatives	67
2.5 β 7- β 8 and β 4- β 5 Loops and Enzymatic Behavior	72
2.6 Crystallization of Anc _{Staph}	73
2.7 Ancestral Sortase and Future Directions	75
Materials and Methods	76
References	77
Appendix A	81

List of Tables and Figures

Introduction:

Figure I-1. Sortase structure.....	1
Figure I-2. Illustration of Sortase A structure and sorting signal motif at the..... bacterial cell wall	3
Figure I-3. Sortase Mediated Ligation (SML) scheme.....	4
Figure I-4. NMR solution structure of Sortase A.....	7
Figure I-5. Catalytic active site of class A sortases.....	8
Figure I-6. Sortase A derived from <i>S. aureus</i> bound to the LPAT* sorting signal.....	10
Figure I-7. Overlay of NMR structures of apo-SrtA and bound SrtA.....	11

Chapter 1:

Figure 1-1. Heat map of substrate selectivity and catalytic activity of sortase..... enzymes isolated from differing bacterial species	17
Figure 1-2. Sequence alignment of SrtA _{staph} and SrtA _{strep}	17
Figure 1-3. Principle component analysis (PCA) of sortase superfamily and..... β 7- β 8 Loop Sequences	19
Figure 1-4. ‘Loop swapped’ SrtA constructs.....	20
Figure 1-5. Scheme illustrating the SrtA cleavage reaction.....	21
Figure 1-6. HPLC and calibration curve for high throughput assay.....	22
Figure 1-7. Heat map of initial ‘loop swapped’ SrtA enzymes.....	24
Figure 1-8. Heat map of substrate selectivity and catalytic activity of sortase enzymes isolated from differing bacterial species	26
Figure 1-9. Heat map of ‘loop swapped’ SrtA with new SrtA homologues.....	28
Figure 1-10. β 7- β 8 loop sequences from WT SrtA enzymes and SrtA homologues.....	29
Figure 1-11. Expanded graphical representation of ‘loop swapped’ SrtA with new..... SrtA homologues	30
Figure 1-12. SWISS modeled ‘loop swapped’ complexes with new SrtA homologues.....	32
Figure 1-13. SrtA _{strep} with measured angstrom distances between residues..... Glu-128 and Arg-104	33
Figure 1-14. Expanded graphical representation of mutated SPS _{faecalis}	34
Figure 1-15. Heat map of mutated SPS _{faecalis}	35
Figure 1-16. Isolation of monomeric species by SEC.....	36
Figure 1-17. Heat map of ‘loop swapped’ enzymes with a Trp mutation,..... truncated loop, or <i>S. suis</i> β 7- β 8 loop	39
Figure 1-18. Deuteration level of SrtA _{strep} and SPS _{faecalis} residues at different time..... points	42
Figure 1-19. Dermcidin modification experiment.....	43
Figure 1-20. FPLC chromatogram of SAS _{pneumoniae}	49
Figure 1-21. Gel image of SrtA enzymes.....	50
Figure 1-22. Gel image of SrtA enzymes.....	50
Figure 1-23. Mass spectrometry analysis of ‘loop swapped’ SAS _{pneumoniae}	51
Figure 1-24. SrtA peptide synthesis scheme.....	52

Figure 1-25. HPLC trace of Abz-LPATGG-K(Dnp) peptide.....	52
Figure 1-26. MS spectrum of Abz-LPATGG-K(Dnp) peptide.....	53

Chapter 2:

Figure 2-1. Steps for Ancestral Sequence Reconstruction (ASR).....	60
Figure 2-2. Phylogenetic tree displaying evolutionary branch points of sortase A.....	61
Figure 2-3. Heat map of initial ancestrally reconstructed SrtA enzymes.....	62
Figure 2-4. Expanded graphical representation of ancestrally reconstructed..... SrtA enzymes	63
Figure 2-5. BLAST sequence alignment of Anc _{staph} and Anc _{strep}	65
Figure 2-6. SWISS model of Anc _{staph} and WT SrtA from <i>S. aureus</i>	65
Figure 2-7. β 7- β 8 loop sequences of Anc _{strep} and WT SrtA homologues.....	66
Figure 2-8. SWISS model of Anc _{strep} and WT SrtA from <i>S. pneumoniae</i>	67
Figure 2-9. Phylogenetic tree of ancestral SrtA sequences.....	68
Figure 2-10. Heat map of more ancestral SrtA enzymes.....	69
Figure 2-11. Sequence comparison of ancestral SrtA enzymes and newly constructed..... nodes	70
Figure 2-12. β 7- β 8 loop and β 4- β 5 loop sequences of WT SrtA enzymes and sortase..... homologues	71
Figure 2-13. SWISS model of Anc ₄₀₈ , Anc ₅₀₃ , and Anc ₅₄₇ and WT <i>S. agalactiae</i>	72
Figure 2-14. Crystallization of AncStaph.....	74
Figure 2-15. Gel image of SrtA enzymes.....	76

Appendix 1:

Table 1. Standard deviation values from 5 th position kinetic enzyme assays.....	81
Table 2. Standard deviation values from 4 th position kinetic enzyme assays.....	81
Table 3. Molecular Weights and Extinction Coefficients of Sortase A Enzymes.....	82
Table 4. Sortase A Enzyme Sequences.....	82
List of Abbreviations	84
List of Enzyme Abbreviations	84

Introduction

Sortase Enzymes

Sortase enzymes, membrane associated cysteine transpeptidases, are a major contributor to the surface chemistry of live bacterial cells. Surface proteins play a number of key roles in bacterial virulence, including: promoting bacterial adhesion to host tissues, resistance to killing by phagocytic killing, essential nutrient uptake, and host cell invasion (1–5). For example, SrtA is required for virulence of *S. aureus* (MRSA). This bacterial infection is responsible for several difficult to treat infections in hospital settings which can lead to severe bloodstream infections and pneumoniae (6). Sortase enzymes can be organized into classes A-F, where each class plays a unique role on the cellular surface and can exhibit different substrate preferences. For example, class A sortase enzymes act as ‘housekeeping enzymes’ anchoring surface proteins to the cell wall recognizing a LPXTG motif while class B and C sortases assist with heme iron uptake and pilus polymerization recognizing the NPQTN and QVPTG motifs respectively (7). Class D and E sortases are involved in spore formation and aid with pilus attachment and aerial hyphae formation and in addition, have been shown to also act as ‘housekeeping enzymes’. The role of class F sortase enzymes is suggested to also be as a ‘housekeeping enzyme’ (8, 9). Class A sortases have been primarily studied due to their ability to act as drug targets, clinically relevant pathogenic bacteria such as *S. aureus* use class A SrtA enzymes to

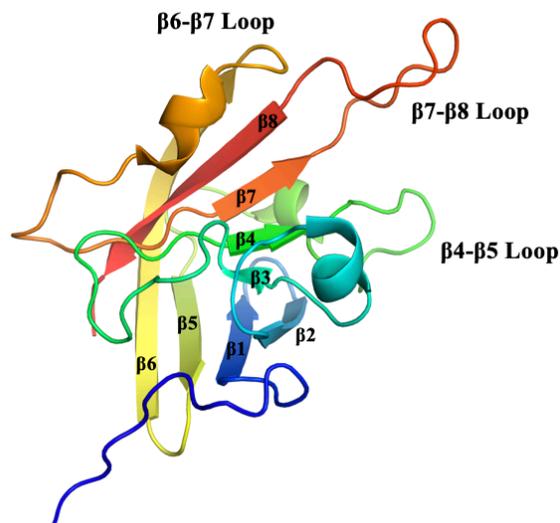


Figure I-1. Sortase structure. Solution structure of sortase A-substrate structure from *S. aureus* solved via solution NMR (PDB 2KID).

display virulence factors on the cell surface. Previous research has also indicated that knocking out the SrtA gene reduces the bacterial virulence (8). The NMR structure of SrtA reveals a β -barrel core structure, in which the conserved active site is made up of Cys, Arg, and His residues which uniquely position Cys towards the incoming canonical sortase A sorting signal, LPXTG, in order to facilitate a ligation mechanism (Figure I-1)(10).

Sortase, Inteins, and Protein Ligation Schemes

A variety of protein ligation methodologies exist currently, with prominent examples including sortase-mediated ligation and intein-based methods. More specifically sortase A (SrtA, see *List of Abbreviations in the Appendix*) is able to ligate a LPXTG tagged construct to any number of oligoglycine-containing structures in a process known as sortase-mediated ligation (SML). SML has a wide variety of uses such as *in vitro* site-specific modification of proteins and controlled attachment of proteins and peptides to live cells and solid supports as well as the ability to site specifically conjugate antibody drug conjugates with cytotoxic payloads (5, 11, 12). In addition, this technique opens up new routes for the creation of novel anti-infective agents, a necessity for contending against global spread of antibiotic resistant bacteria (7).

A second biochemical tool that is frequently used for protein ligation is inteins. Inteins are proteins that play a crucial role in protein splicing by removing themselves from a larger polypeptide chain by use of a ligation scheme. Inteins have been primarily used for protein expressed protein ligation, which in turn has a number of applications such as segmental isotopic labeling of proteins, or controlled expression of toxic proteins (13).

Class A sortase enzymes are able to ligate proteins containing the cell wall sorting signal to an amino group, displaying proteins on the cell wall. *In vitro*, SrtA recognizes and is able to

ligate a specific, C-terminal five amino acid ‘sorting motif’, LPXTG, to any number of oligoglycine-containing structures in a process known as sortase mediated ligation. SML has a wide variety of uses such as *in vitro* site-specific modification of proteins and controlled attachment of proteins and peptides to live cells and solid supports (5). On the surface of the cell, this sorting signal is bound to a segment of hydrophobic amino acids spanning the lipid membrane, and has a tail composed of positively charged residues, which initially localizes it to the cell membrane (Figure I-2)(3).

The SML ligation reaction scheme includes, first, recognition of the sorting motif by the membrane associated sortase. Then, cleavage occurs between the threonine and glycine residues in the LPXTG motif (positions 4 and 5 respectively) via an attack by the sulfhydryl group originating from the active site Cys residue in SrtA, in turn forming an labile thioester-linked acyl enzyme intermediate which is then resolved by a nucleophilic attack from the aminoglycine nucleophile, generating a site specifically ligated acyl donor and acceptor (Figure 1-3)(3).

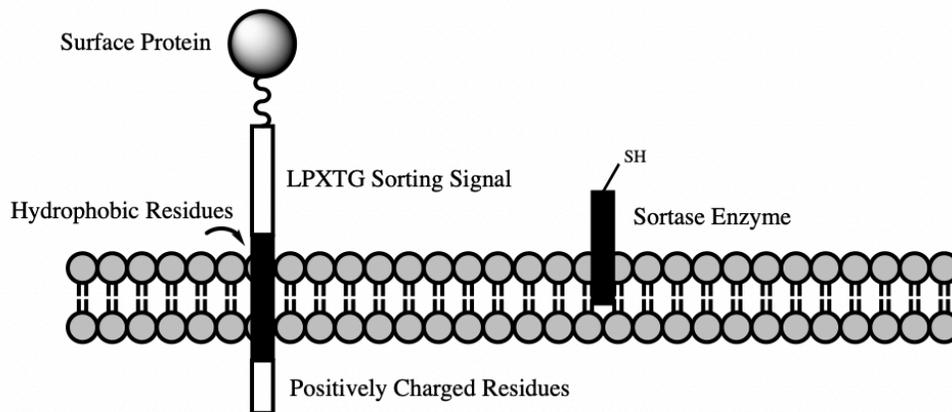


Figure I-2. Illustration of Sortase A structure and sorting signal motif at the bacterial cell wall. The cell wall sorting signal is adjacent to a stretch of hydrophobic residues and a group of positively charged amino acids anchoring the sorting signal to the cell membrane. This complex will interact with the sortase enzyme illustrated to the right as part of the sortase ligation scheme.

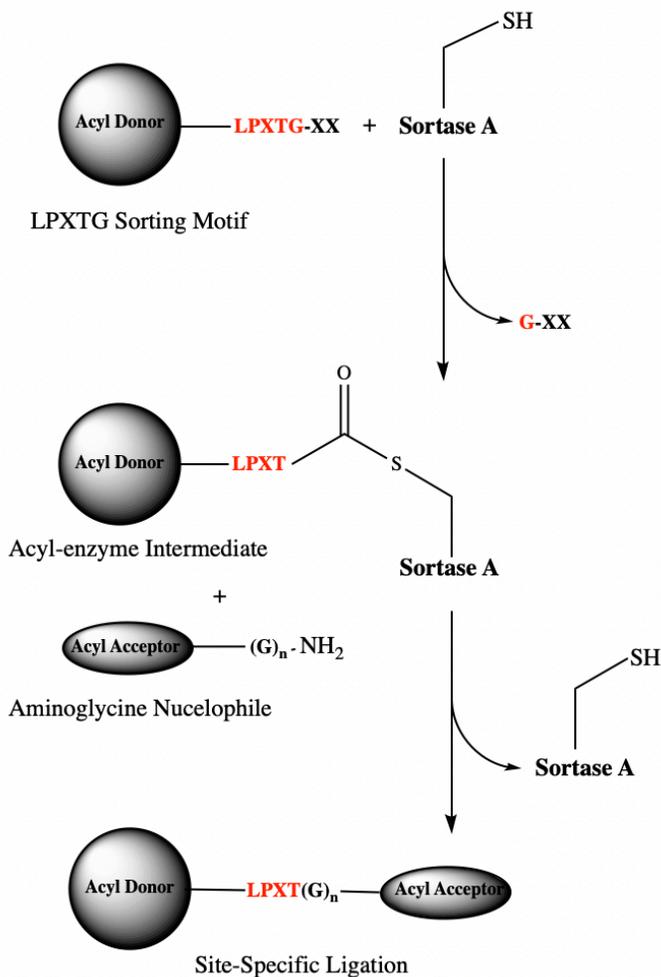


Figure I-3. Sortase Mediated Ligation (SML) scheme. The sorting motif, LPXTG (acyl acceptor)(X denotes any amino acid) is cleaved between the Thr and Gly residues by the catalytic cysteine in SrtA, this in turn forms a thioester intermediate which is immediately attacked by the aminoglycine nucleophile (acyl acceptor), forming a new amide linkage, ligating the sorting signal to acyl acceptor.

Sortase enzymes can be used for the *in vitro* modification of live cells, solid supports, proteins, or synthetic peptides (5). Previous engineering studies have primarily focused on altering the substrate specificity of SrtA_{staph}, improving the modest kinetics of SrtA_{staph}, measured by k_{cat}/K_m and reducing the need for a Ca^{2+} cofactor, leading to the development of a number of variants, including the so-called “pentamutant” and “heptamutant” SrtA enzyme (11, 14–16). Chen *et al.*, by use of yeast display, were able to evolve a SrtA enzyme with improved catalytic efficiency. This pentamutant has five mutations (P94R/D160N/D165A/K190E/K196T) which yielded a 120-fold improvement in k_{cat}/K_m (aka catalytic efficiency) in comparison to the original WT

SrtA_{staph} construct (14). Though a rate increase was observed for this pentamutant, its reliance on a Ca²⁺ cofactor makes it difficult to use in environments with a low Ca²⁺ concentration or in the presence of Ca²⁺ binding constructs, leading to the development of a heptamutant SrtA enzyme. This heptamutant added two additional mutations (E105K/E108A) which eliminated the need for Ca²⁺ as a cofactor for the engineered SrtA_{staph} construct (14). This exclusion of the Ca²⁺ cofactor is useful for *in vivo* studies where Ca²⁺ concentrations are usually lower than observed in *in vitro* studies (15). These engineered constructs prove useful for many studies which utilize SrtA_{staph} and require rate enhancement or the elimination of the Ca²⁺ cofactor.

Engineering changes in the overall specificity profile of SrtA_{staph} can be accomplished in many ways. Dorr, *et al.*, were able to utilize a bond-forming enzyme screening system which allowed for the evolution of a SrtA variant with a mutated β 6- β 7 loop which possessed an altered substrate specificity profile, recognizing LPXSG or LAXTG substrates with around a 51,000 fold change in overall substrate specificity and minimal reduction in catalytic efficiency (11). In addition, the Schwarzer group reported a second generation sortase library utilizing a randomized β 6- β 7 loop. They screened this library for sortase mutants that accepted the LPXTG and the FPXTG motifs. These screens yielded multiple mutants that displayed the desired substrate specificity, the F-21 mutant was the most promising out of their study, accepting the LPXTG and the FPXTG motifs and displaying improved catalytic activity (17).

Protein Engineering

Proteins, such as sortases, are a desired drug target due to their ability to catalyze highly specific reactions as well as taking regio- and stereoselectivity into account (18). To understand

the structure-function relationships of proteins and develop specialized pharmaceuticals, researchers may utilize a technique commonly known as protein engineering.

Protein engineering involves the design of new polypeptides, not found in nature, by either mutation of existing native proteins or the *de novo* production of new structures. By engineering these proteins, researchers are able to produce functional changes or which shape the overall usage of these proteins (19). Protein engineering can be accomplished by many different strategies; some examples of these are knowledge-based mutagenesis (KBM), computational protein design (CPD), directed evolution (DE), and sortase based modification (19). For the purposes of this project we will only be focusing on the use of KBM. KBM involves the utilization of biochemical knowledge to identify key components of a protein structure that when mutated, can impact the functional profile of the protein, such as mutating peptide-agonist binding sites to determine potential pharmaceutical targets (20). Researchers are able to utilize KBM to identify and further modulate protein tools that may be utilized for protein ligation schemes, specifically those from bacterial sortases.

Structural Components of Class A Sortases

As described, the canonic catalytic domain structure of sortases is composed of an eight-stranded β -barrel fold, the ‘sortase fold’ (21). The archetypal sortase, sortase A, derived from *S. aureus* (SrtA_{staph}) was the first solved sortase structure, determined by the Clubb and Schneewind groups with nuclear magnetic resonance (NMR) spectroscopy (Figure I-4)(21). This structure provided researchers the ability to investigate fundamental structural components of the class A sortase

family. Notable conserved structural components of class A sortases include the presence of multiple substrate interacting loops, these being the β 7- β 8, β 6- β 7, and β 4- β 5 loops are near or adjacent to the substrate binding groove. Though these loops vary widely in length and identity there are key conserved residues in each loop and modulation of these residues can result in a decrease in catalysis and

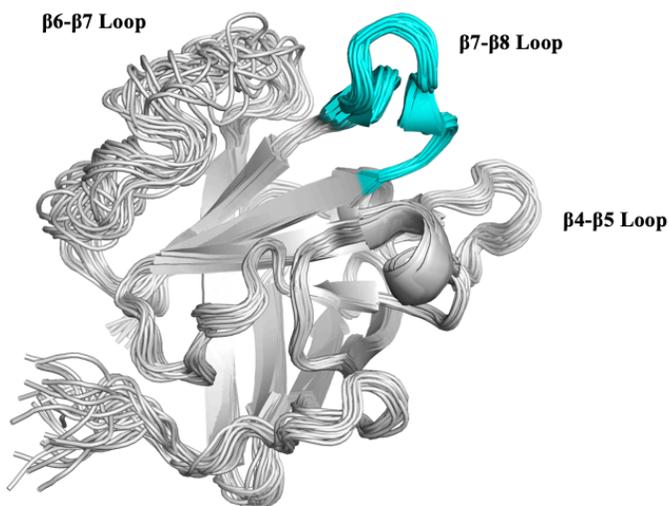


Figure I-4. NMR solution structure of Sortase A. Derived from *Staphylococcus aureus*. The protein is shown in cartoon representation and colored grey. The active site β 7- β 8 loop is adjacent to the substrate binding groove and colored cyan (PDB 1IJ1).

substrate promiscuity (11, 22). Another conserved structural component is the ‘catalytic triad,’ containing a catalytically active Cys, His, and Arg residues (Figure I-5)(4, 23). Cys acts as the catalytic cysteine, required for the first step of the sortase-mediated ligation reaction, cleaving between the Thr and Gly residues in the LPXTG sorting motif. His acts as a general acid/base (24, 25), while Arg may help create a stabilizing oxyanion hole in correlation with the amide from the backbone of the β 7- β 8 loop (Figure I-5)(7, 10, 26). The RMSD values for this alignment of the main chain atoms of *S. pyogenes* and other SrtA enzymes was between .506 and 1.691 Angstroms over roughly 400 main chain atoms (Figure I-5). In addition, in WT SrtA_{staph} a residue in the β 7- β 8 loop, Trp-194 partially shields active site residues from the solvent in an apo state, possibly playing a role in catalysis (4).

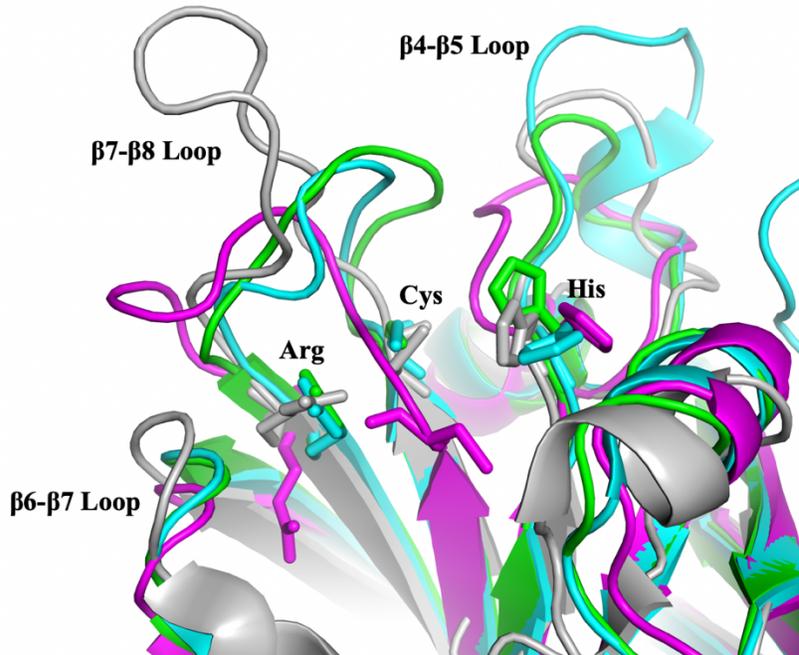


Figure I-5. Catalytic active site of class A sortases. Catalytic residue side chains are shown in stick representation and labeled. Grey= *S. aureus*, Cyan= *S. pyogenes*, Magenta= *B. anthracis*, Green= *L. monocytogenes*. Cys is the catalytic cysteine cleaving in between the Thr and Gly residues of the archetypal LPXTG motif. His acts as a general acid/base. Arg may help stabilize the oxyanions generated during acyl enzyme intermediate formation and the subsequent attack by amine nucleophiles (PBD 2KID, 3FN5, 2KW8, and 5HU4).

Calcium binding in WT SrtA_{staph} is also indicated by structural NMR studies. Ca²⁺ binding occurs in an ordered pocket formed by the β 3- β 4 and the β 7- β 8 loops and is required for catalytic activity of SrtA_{staph} (26). Calcium is thought to promote substrate binding due to bacteria commonly encountering Ca²⁺ ions at sites of infection due to the high concentrations of Ca²⁺ in the extracellular fluid (21). Ca²⁺ ion binding allosterically controls enzymatic activity in SrtA_{staph} by influencing the β 6- β 7 loop dynamics, allowing for adaptive recognition of the LPXTG substrate by modulation of the β 6- β 7 loop (27). Ca²⁺ dependence is specific to SrtA from *S. aureus* and in some cases, Ca²⁺ actually inhibits the activity of sortases such as SrtA from *S. pyogenes* (7).

Sorting Signal Binding

In class A sortases the LPXTG substrate binds in a ‘binding pocket’ formed by a matrix of β sheets, surface loops, and α helices. The base of the pocket is comprised of residues from the $\beta 4$ and $\beta 7$ loops, directly interacting with the proline (Pro) residue of the LPXTG substrate (Figure I-6). This proline residue is said to play an ‘architectural’ role by producing a kink in the middle of the substrate, so that, when bound, the kinked L-shape substrate orients the C-terminus of the sorting motif towards the catalytically active cysteine (26). The walls lining the binding pocket are comprised of residues that form the surface loops. These loops originate from the $\beta 6$ - $\beta 7$ strands, $\beta 3$ - $\beta 4$ strands, and the $\beta 2$ strands- $\alpha 2$ helix (Figure I-6)(26). Binding of the LPXTG motif will cause the active site to reorganize. The flexible, and highly mobile $\beta 6$ - $\beta 7$ loop will undergo a disorder-to-order transition, forming an 3_{10} helix (26, 28).

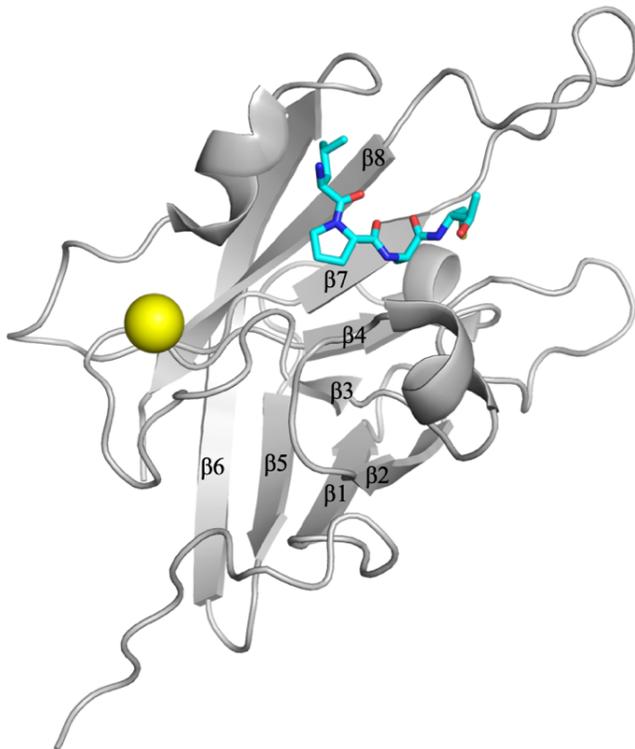


Figure I-6. Sortase A derived from *S. aureus* bound to the LPAT* sorting signal. Loop and helix structures are shown interacting with the bound LPAT* substrate illustrated in ball-and-stick form colored cyan. Ca^{2+} bound to the distal $\beta 3$ - $\beta 4$ pocket is also shown colored yellow. Solved via solution NMR (PDB 2KID).

This 3_{10} helix is able to interact with the bound substrate. In addition, binding of the LPXTG motif will also cause a displacement of the $\beta 7$ - $\beta 8$ loop. This displacement could play a role in exposing the catalytically active His-120 and assisting in the next step of SML, integration of the lipid II complex (27). The binding of the sorting signal is commonly described as an “induced fit” because when the sorting signal is bound the $\beta 7$ - $\beta 8$ loop will transition to a more “open” conformation, allowing for improved contact with the covalently bound sorting signal in the binding groove (29).

Active Site Loop Structure and Dynamic Movement

As discussed above, there are many functional components that make up the substrate binding groove of SrtA_{staph}. Out of these components, both the $\beta 7$ - $\beta 8$ and the $\beta 6$ - $\beta 7$ active site loops are crucial for effective binding of the LPXTG sorting motif in class A sortases (Figure I-7). Before substrate binding occurs, the “closed position”, the apo-SrtA_{staph} $\beta 7$ - $\beta 8$ loop is highly mobile. Previous studies indicate that this loop is also unstable and requires a calcium cofactor to be bound in order to modulate hinge motions in the mobile $\beta 7$ - $\beta 8$ loop, allowing for reordering into proper orientation of the loop before substrate binding may occur (1, 21). This reordering is achieved by signal transmission from the Ca²⁺ binding pocket to the $\beta 7$ - $\beta 8$ loop, via repetitive folding and unfolding of short helical stretches in the $\beta 6$ - $\beta 7$ loop (30).

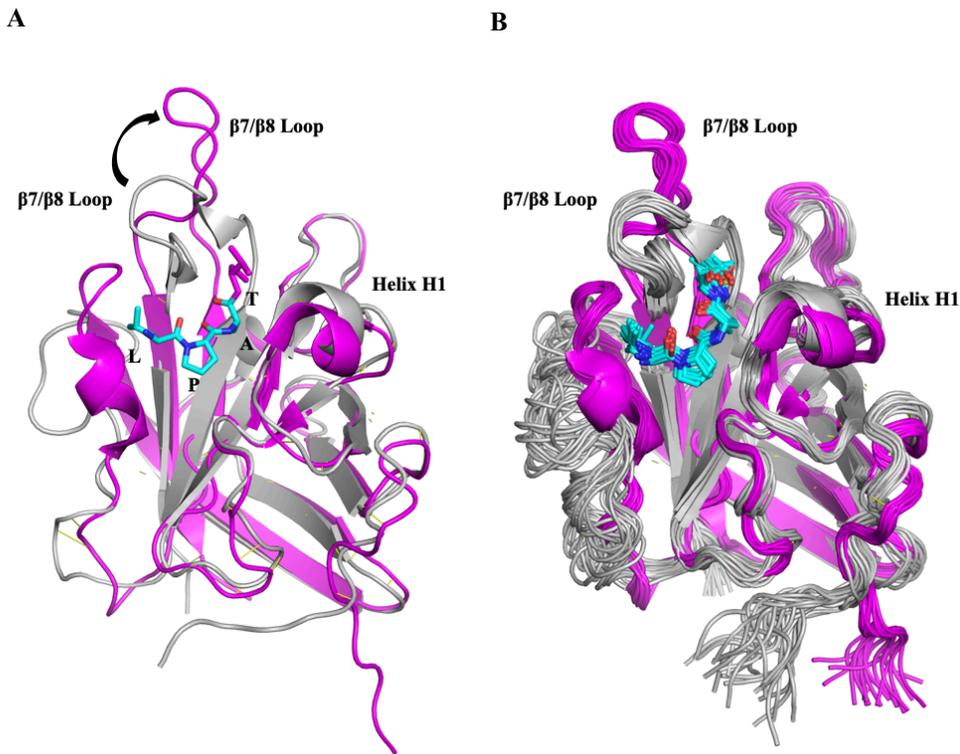


Figure I-7. Overlay of NMR structures of apo-SrtA and bound SrtA. (A) Apo-SrtA from *S. aureus* (PDB 1IJA) shown in grey and SrtA with bound LPAT* sorting signal shown in magenta, LPAT* sorting signal shown in ball and stick format colored cyan. Arrows indicate the unbound to bound state (PDB 2KID). **(B)** NMR ensemble structures of apo and bound SrtA.

The $\beta 6$ - $\beta 7$ loop will then adopt a conformation in which the side chains of Val-168 and Leu-169 are rotated away from the body of the protein (27). Structurally the Glu-171 residue that coordinates Ca^{2+} originates from the $\beta 6$ - $\beta 7$ loop. When the sorting signal binds, the complex orients into the final “open conformation” where the $\beta 7$ - $\beta 8$ loop is partially displaced, leaving room for proper binding of the sorting signal where the active site catalytic cysteine, Cys-184, is positioned to cleave between the Thr and Gly residues of the LPXTG motif, the integral first step of sortase mediated ligation.

Substrate Specificity of SrtA and ‘Loop Swapped’ SrtA

Previous *in vitro* studies on SrtA_{staph} have revealed optimal catalytic activity using an expanded sorting motif with the LPXTGG substrate but no catalytic activity for any additional 5th position substrates (31). Though this substrate specificity is advantageous for researchers looking to perform site-specific modifications where cross reactions would be unfavorable this requirement for substrates containing the LPXTGG motif can be considered a limitation of SrtA_{staph}, and the ability to target variants of the LPXTGG motif would increase the versatility of SML (11, 31, 32). SrtA_{strep} on the other hand displays modest efficiency (31). But this low catalytic efficiency is countered with a broader substrate profile. SrtA_{strep} is quite nonselective at the 5th position in the LPATXG motif, and the ability to harness this selectivity profile and engineer an enzyme that maintains the high catalytic activity, like that from SrtA_{staph} but also possesses an broader substrate specificity profile, would extend the capabilities of established SML schemes.

Previous research that has explored ‘loop swapped’ constructs has primarily focused on swapping both the β 6- β 7 and β 7- β 8 loops. A ‘loop swapped’ SrtA enzyme study was implemented by Bentley, *et al.*, where the β 6- β 7 loop sequence was swapped from *S. aureus* SrtB into *S. aureus* SrtA, altering the substrate specificity profile of SrtA to accommodate recognition of NPQTN substrates and modulating the overall catalytic specificity profile; the ability of an enzyme to process NPTQN the given reaction was 700,000-fold higher compared to WT SrtA. Though they only observed substrate cleavage for this loop swapped construct but could not complete the ligation reaction (22). A recently published study out of the University of Groningen investigated a “loop grafted” β 7- β 8 loop to engineer the specificity of *Streptococcus pyogenes* SrtA. By grafting in β 7- β 8 loops from *S. aureus* and *B. anthracis* researchers found that the engineered *S. pyogenes* SrtA with the *S. aureus* β 7- β 8 loop showed improved activity

toward the LPETG substrate, the established sorting motif that is recognized by *S. aureus*. Their results indicated that the β 7- β 8 loop may be modulating substrate access to the active site groove (33). Similar to this study we have also explored the impact of swapping in this β 7- β 8 loop between homologous SrtA enzymes in order to modulate the overall specificity and activity of these constructs, as described below.

A Multi Direction Approach

Utilizing these two constructs, SrtA_{staph} and SrtA_{strep}, we designed a compatible mutant enzyme where the desirable aspects of both enzymes, high catalytic activity from SrtA_{staph} and a broader substrate profile from SrtA_{strep} are displayed. By means of a principle component analysis (PCA) we were able to globally analyze the sortase network and identified a region of variability in the β 7- β 8 loop, this loop region near the catalytic domain was swapped between SrtA_{staph} and SrtA_{strep}. The β 7- β 8 loop is recognized has also previously been recognized as a potential component to substrate binding (33).

This ‘loop swap’ concept is not only limited to these two enzymes, SrtA_{staph} and SrtA_{strep}, but any number of loops may be swapped in, with a SrtA enzyme core and any β 7- β 8 loop that is adjacent to the active site engineered on. The β 7- β 8 loop boundaries are the N-terminal Cys and the C-terminal Arg of the catalytic triad. By utilizing these species with increased promiscuity in a hybrid enzyme schematic as described previously, it may be possible to alter the substrate specificity to include amino acids not recognized by SrtA_{staph} or SrtA_{strep} enzymes or improve catalytic activity.

Another tool for investigating sequence variation in protein families is ancestral sequence reconstruction (ASR) where ancestral protein sequences are reconstructed using an alignment of

extant protein sequences. These ancestral sequences provide insight into the natural sequence variations around the extant sequences and may reveal novel links between sequence variation and biochemical behaviors such as substrate promiscuity and catalytic activity (34–36).

We hoped to identify an ancestral SrtA sequence that would display an improved substrate specificity profile and/or improved catalytic efficiency compared to the WT SrtA_{staph} and SrtA_{strep}. Though we were not limited to just these two ancestral sequences, we were able to reconstruct additional sequences further back on the phylogenetic tree. Though, these enzymes are catalytically dead we were still able to explore the natural sequence variation and the investigation into these constructs is ongoing.

By engineering over eight *S. pneumoniae* β 7- β 8 loop variants with loop sequences from different bacterial species and by performing ancestral sequence reconstruction on extant class A sortase sequences we were able to broadly explore the natural sequence variation of class A sortase enzymes and deepen our understanding of sortase biology, especially of the role of the β 7- β 8 loop. Specifically, the loop's recognition of ligands in SrtA enzymes, in particular, SrtA from *S. pneumoniae*.

**Chapter 1: ‘Loop Swapped’ Engineered Sortase A
from *Streptococcus pneumoniae***

Introduction

1.1 Introduction to ‘Loop Swapped’ SrtA

Previous studies on SrtA_{staph} revealed that the preferred substrate was LPXTGG, indicating a highly stringent substrate profile as compared to that seen in SrtA from *Streptococcus pneumoniae* (SrtA_{strep}) or other SrtA homologues (31). This specificity motif does offer benefits to researchers who are looking to perform modifications in complex settings (31). But, this rigid motif specificity can also be considered a significant drawback to usage of SrtA_{staph} for techniques such as simultaneous conjugation of multiple peptide substrates to a target (11). As mentioned previously, SrtA_{strep}, displays a broader substrate specificity profile, a more ‘promiscuous’ enzyme, but exhibits poor catalytic efficiency.

The activity and selectivity of class A sortases is primarily based on the substrate interacting loops that border the active site of sortase. Previous research has indicated both the β 6- β 7 and the β 7- β 8 loops play a role in substrate recognition and catalysis (7, 22, 28, 32, 33). We performed a PCA which reasserted the β 7- β 8 loop as a region of high variability. We hypothesized that this loop may play a role in the biochemical differences observed between SrtA species, and selected it as a ‘loop swap’ target for the purposes of this study. The β 7- β 8 loop of the *S. aureus* was swapped onto the core of the *S. pneumoniae* core (SPS_{aureus}) or the β 7- β 8 loop from *S. pneumoniae* was swapped onto the core of the *S. aureus* (SAS_{pneumoniae}).

This ‘loop swap’ concept is not only limited to these two WT enzymes, SrtA_{staph} and SrtA_{strep}. Any number of loops may be swapped in, with a WT SrtA_{staph} or SrtA_{strep} ‘core’ and a new β 7- β 8 loop swapped on. Research conducted by the Antos group revealed that SrtA enzymes from a number of bacterial species exhibited differing substrate selectivities and

catalytic activity while still maintaining similar identities and catalytic activity (Figure 1-1, 1-2). Catalytic activity was observed for both 4th and 5th position substitutions but 5th position substitutions were of primary interest due to the observed enhanced substrate promiscuity (31). By utilizing these species with enhanced promiscuity in a hybrid enzyme schematic as described previously, we hypothesized that it may be possible to alter the substrate specificity to include amino acids not recognized by the WT SrtA_{staph} or SrtA_{strep} enzymes, and in turn and provide further insight into the role of this β 7- β 8 substrate-interacting loop in class A sortases.

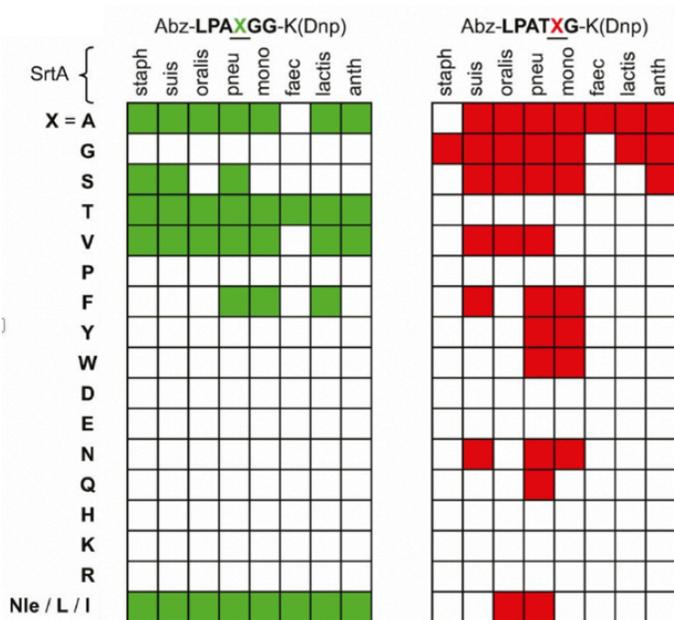


Figure 1-1. Heat map of substrate selectivity and catalytic activity of sortase enzymes isolated from differing bacterial species. “Hits” on the heat map (colored green or red) indicate that cleavage had occurred and was measured via MS after 24 hrs when the enzyme and substrate were incubated together. Colored letters indicate the substituted amino acid in either the 4th or 5th position (left to right) in the canonical LPXTG motif (Adapted from Nikghalb KD, *et al.*, 2018).

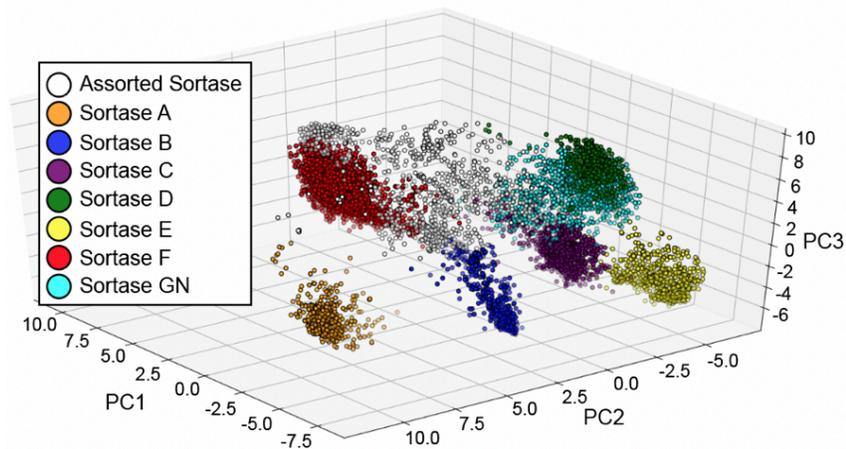


Figure 1-2. Sequence alignment of SrtA_{staph} and SrtA_{strep}. Colored red (small and hydrophobic residues), green (hydroxyl, sulfhydryl, amine, and glycine residues), magenta (basic residues), and blue (acidic residues). Aligned using Clustal Omega. Sequence identity between SrtA_{staph} and SrtA_{strep}, 77.56%.

Results and Discussion

1.2 Principle Component Analysis of Sortase A

A Principle Component Analysis is a statistical method which allows data in a higher dimensional space to be projected into a lower dimensional space (e.g. 2-D or 3-D), this is achieved by maximizing the variance of the data set so that even though the dimensionality is reduced, the variability remains relatively high (37). This PCA allowed us to globally analyze the sortase family tree, examining each sequence as a whole all at one time, differing from a network analysis in which only portions of the sequence are analyzed. Every published sortase enzyme sequence was sourced from the UniProt data base, and a multi sequence alignment (MSA) informed us as to how the sequences were related. From there, individual residues and their chemical properties were introduced for every single amino acid and sequence gap, producing a Protein Similarity Matrix (PSM). In this PSM each sequence holds a position in some higher dimensional space wherein each protein sequence correlates to a data point in this protein sequence space. PCA allowed us to simplify these data points down into a 3-D or 2-D space for further analysis of the data shape (Figure 1-3). Classes of sortases clustered, and we selected and analyzed the Class A sortase cluster. Within this cluster of class A sortase enzymes, we were able to identify the β 7- β 8 loop and the β 6- β 7 loop as regions of variability within class A sortase enzymes. We hypothesize that these regions may play a role in the observed biochemical differences in class A sortase enzymes such as selectivity and activity.



	β6-β7 Loop	β7-β8 Loop	
SrtA_staph	VGNETRKYKMTSIRDVKPTDVGVLDEQKGGKDKQLTLIT	CDDYNEKTGVWEKR	KIFVATEV 166
SrtA_faec	DLSTVYTYKITSVEKIEPTRVELIDDPGQ-NMITLIT	CGDLQ-ATTRIAVQGTLAATTP	160
SrtA_lactis	DKDNIYQYKIDKIDVVTPEHVEVINDSPGK-KEITLVT	CADAE-ATHRIIVHGTFVDKTS	166
SrtA_suis	DKTNVYTYVIDSVEIVSPESVYVIDDVEGR-TEVTLVT	CTDYY-ATQRIIVKGVLESTTP	166
SrtA_strep	DKNKVYTYEIREVKRVTDPDRVDEVDDRDGV-NEITLVT	CEDLA-ATERIIVKGDVKETKD	157
SrtA_oralis	DKDKVYTYEITEVKRVTDPDRVDEIEDRDGV-KEITLVT	CVDYN-ATERIIVKGIKESKA	166
SrtA_mono	DLENLYEYTVTETKTIDETEVSVIDDIT-KD-ARITLIT	CDKPTETTKRFVAVGELEKTEK	149
SrtA_anth	DNENEYEVAVTGVSEVTPDKWEVVEDH-GK-DEITLIT	CVSVKDNSKRYVAVAGDLVGTKA	172
	* : : :: : ** : *	: : .	

Figure 1-3. Principle component analysis (PCA) of sortase superfamily and β7-β8 loop alignment. Sortase sequences sourced from UniProt were subjected to a multisequence alignment in which they were filtered by the presences of large gaps (from 16164 to 9427). Data was simplified to 3D space (PCA1-PCA3 axis) and class A sortase was selected. A BLAST sequence comparison displays variation in the β7-β8 region, boxed in black. β6-β7 loop boxed in red.

1.3 Initial ‘Loop Swapped’ Constructs SAS_{Pneumoniae} and SPS_{Aureus}

As discussed, the β7-β8 loop is recognized as a component to substrate binding, laying adjacent to the catalytic residues, Cys-184, Arg-197, and His-120. Between SrtA_{staph} and SrtA_{strep} the β7-β8 loops differ significantly in length, with the SrtA_{staph} β7-β8 loop containing 14 residues (CDDYNEKTGVWEKR), and the SrtA_{strep} β7-β8 loop containing 9 residues (CEDLAATER). Additionally, there is only one residue conserved between the loops, an Asp located two residues C-terminal to Cys-184. The residue numbering for this study is based on the WT SrtA_{staph} unless otherwise specified. Taking these β7-β8 loop residues and swapping them between the β-barrel

core of the two enzymes resulted in two hybrid, 'loop swapped' chimeric constructs, SPS_{aureus} and SAS_{pneumoniae} (Figure 1-4).

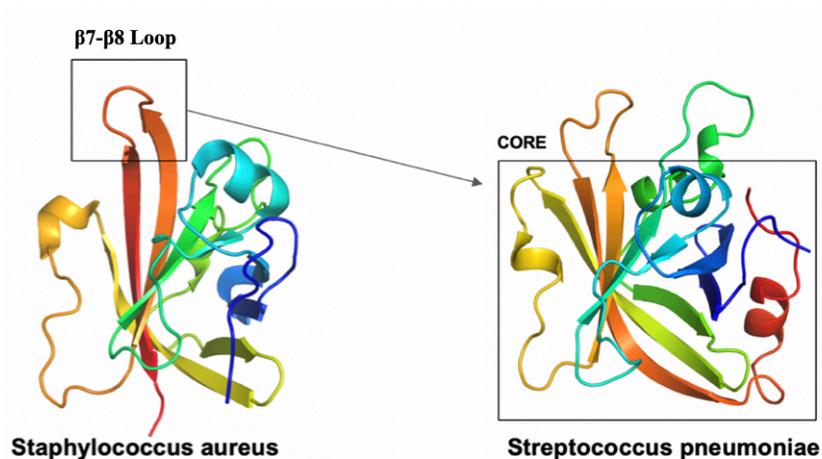


Figure 1-4 'Loop swapped' SrtA constructs. The β7-β8 loop from the SrtA_{staph} construct is boxed, the core from the SrtA_{strep} construct is boxed. These two species are merged together to form one of the new constructs, SPS_{aureus}. (PDB 2KID and homology model made using SWISS Model with 3RCC used as the template structure).

These engineered constructs were hypothesized to follow the substrate specificity and catalytic activity of the enzyme from which the loop originated. Consequently, we hypothesized that the substrate specificity and catalytic activity of the SPS_{aureus} should be most similar to the behavior of SrtA_{staph} with a possibility of increased substrate specificity due to the fact that this loop has been spliced onto the core of the more promiscuous SrtA_{strep}. The opposite was anticipated hold true for SAS_{pneumoniae}.

1.4 High Throughput Fluorescence Assay Development

In order to measure the activity and promiscuity of our engineered SrtA enzymes, the SrtA reactions were monitored using model substrates containing an attached fluorophore and quencher (Abz and Dnp) to estimate overall conversion from starting material to product (Figure 1-5). Using these substrates, we were able to develop an efficient assay in which multiple enzyme-substrate pairs could be monitored in parallel via the increase in the observed Abz fluorescence.

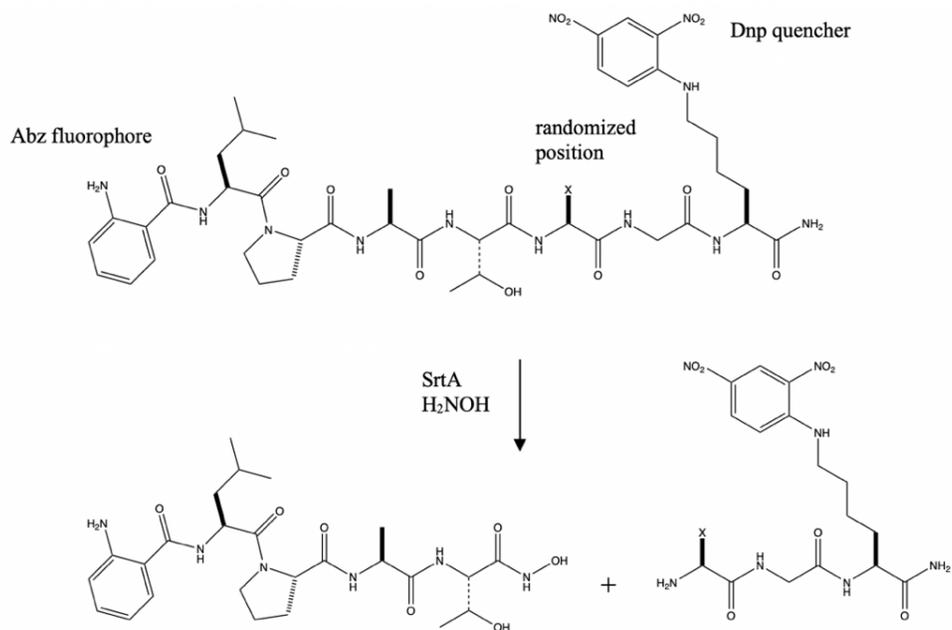
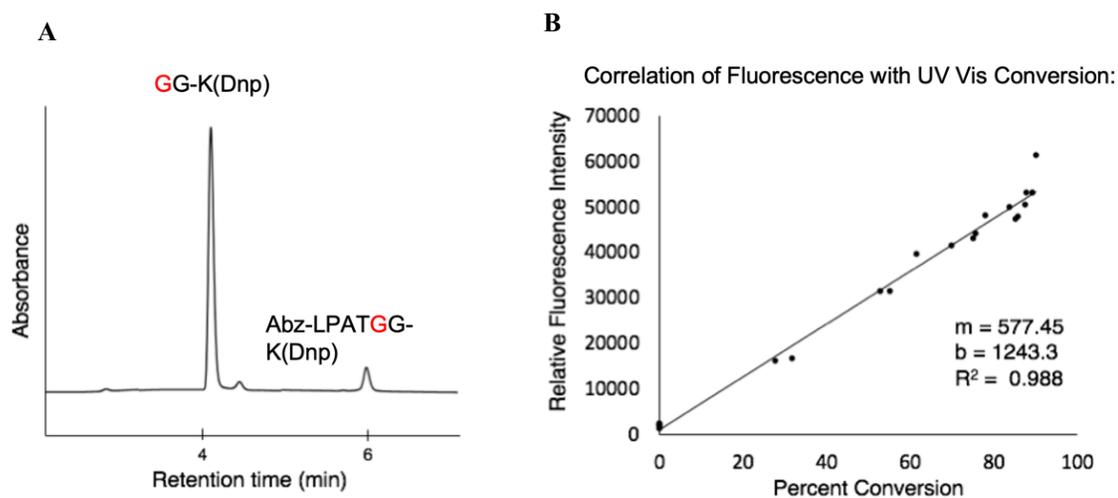


Figure 1-5. Scheme illustrating the SrtA cleavage reaction. The starting material, Abz-LPATXG-K(Dnp) is cleaved to form the products Abz-LPAT and XG-K(Dnp) after the specified peptide is reacted with SrtA and the H₂NOH nucleophile.

This high throughput kinetic fluorescence assay was executed on a microplate reader, where each individual reaction was performed in a single well of a 96 well plate for a maximum of 96 reactions per 2 hr time period, improving our testing speed significantly compared to the previous HPLC method (31). Though this assay offered an improved testing rate, it only provided a fluorescence intensity reading, a unitless number that, without a correlated peak area, such as that observed when performing the reaction on a HPLC, could not be correlated to overall conversion of substrate starting material to product.

The overall conversion rate of these reactions was originally calculated by comparing the starting material peak, the Abz-LPATXG-K(Dnp) species, and the product peak, the XG-K(Dnp) species. Dnp has a strong UV absorption at 360 nm, which allowed us to observe both the starting material peak and the product peak on a HPLC. In addition once the Dnp group is

cleaved it will no longer quench the fluorescent Abz group, resulting in a measurable fluorescent signal. This fluorescent behavior can be harnessed when performing the reaction in a plate reader assay where measuring starting material and product peaks by HPLC is more time consuming. By measuring a fluorescent signal of the fluorophore Abz over a 2 hr time period and correlating it to the more precise UV vis traces obtained on the HPLC we were able to create a standardized calibration curve that can be applied to any sortase mutant to estimate overall percent conversion from the starting material, Abz-LPATXG-K(Dnp) to the product, XG-K(Dnp) without needing to perform the reaction on a HPLC in tandem (Figure 1-6). This novel high throughput screen permitted a broader subset of our sortase enzyme mutants to be screened for selectivity behaviors and resulting catalytic activity.



$$(1) y=577.45x+1243.3$$

Figure 1-6. HPLC and calibration curve for high throughput assay. (A) Representative HPLC reaction peaks, starting material (Abz-LPATGG-K(Dnp)) and product (GG-K(Dnp)), absorbance at 360 nm. (B) Calibration curve correlating fluorescence from plate reader assay to percent conversion calculated via peak ratios obtained from HPLC traces such as that shown in (A).

1.5 Selectivity and Activity of Initial ‘Loop Swapped’ Constructs

When substituting residues in for the 5th position in the LPATXG motif we selected LPATGG, LPATSG, and LPATAG as the representative substrates based on a previous study out of the Antos lab which indicated that if a sortase enzyme is not catalytically active with one or more of these substrates, catalytic activity will not be observed in any other tested substrates (31). A similar pattern is observed for 4th position motif LPAXGG in which LPAAGG, LPAEGG, and LPAIGG were selected as the representative substrates. For the purposes of this study we determined that a 20% conversion from our starting material, Abz-LPATXG-K(Dnp) to our desired cleavage product XG-K(Dnp) was sufficient to claim that a WT or engineered construct was catalytically active with the specified substrate. This cut off percentage is based on the magnitude of experimental error which was consistently 15-17%. In addition, for this study we acknowledge that the presence of a His-tag used for recombinant protein expression can affect enzyme activity. We chose to keep the His-tag on all of our *S. pneumoniae* SrtA variants in order to compare with our WT enzyme, a construct that does not contain a protease cleavage site for His-tag removal, as well as previously published data (31)

Consistent with previous literature, SrtA_{staph} displayed catalytic activity with only the LPATGG and LPAAGG substrates. SrtA_{strep} displayed lowered and roughly similar catalytic activity for the 5th position LPATGG, LPATSG, and LPATAG substrates while displaying no catalytic activity for the 4th position LPAAGG, LPAEGG, and LPAIGG substrates (Figure 1-7). As described earlier, a value of 35, for example, means that at a time point of 2 hrs, there was a 35% total conversion from starting material, Abz-LPATXG-K(Dnp) to the product XG-K(Dnp) measured by fluorescence intensity.

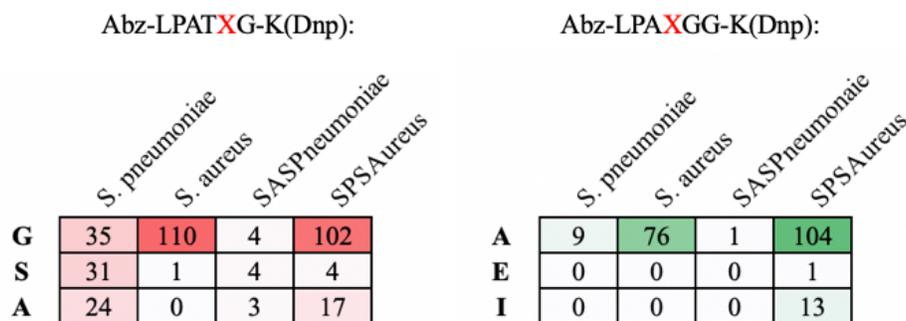


Figure 1-7. Heat map of initial ‘loop swapped’ SrtA enzymes. Displays measured catalytic activity of WT and initial ‘loop swapped’ SrtA enzymes with a 5th and 4th position substitutions (LPATXG and LPAXGG). Each “hit” corresponds to final percent conversion from starting material to product measured via florescent plate reader assay after 2 hrs. Darker shades of red/green indicate an enhanced overall percent conversion.

Subjecting our engineered constructs SPS_{aureus} and SAS_{pneumoniae} to the same 4th and 5th position panels as the WT constructs revealed that swapping the more catalytically active SrtA_{staph} loop onto the more promiscuous SrtA_{strep} core (SPS_{aureus}) produced roughly a 3-fold improvement in catalytic activity for the 5th position LPATGG substrate compared to that of SrtA_{strep}, while almost completely knocking out activity for the LPATSG and LPATAG substrates. A similar behavior was observed for the 4th position substitutions where the previously inactive SrtA_{strep} enzyme had activity completely restored by adding on the SrtA_{staph} β7-β8 loop (Figure 1-7). This observed catalytic activity was consistent with our initial hypothesis that the β7-β8 loop may play a role in target selectivity, and found that indeed, the sequence of these residues can modulate both activity and selectivity as sortase substrate interacting loops and have been indicated a playing a role in substrate recognition and processing (1, 22, 26, 28, 33, 38). The SAS_{pneumoniae} construct produced unexpected results where no catalytic activity was observed for any 4th or 5th position substitutions wherein we expected to possibly observe an increase in substrate promiscuity due to the addition of the β7-β8 loop from the promiscuous SrtA_{strep} onto the catalytically active SrtA_{staph} core, especially due to our results obtained for the SPS_{aureus}

construct. We are uncertain as to why this inactivity is occurring and we investigate potential solutions to address this issue by identifying key residues and structural components that may be causing this inactivity in Section 1.7 *Enzyme Inactivity*. The results we obtained for the SPS_{aureus} construct led us to explore additional loop swapped constructs utilizing the same *S. pneumoniae* core and different substrate interacting loops from sortase homologues.

1.6 ‘Loop Swapped’ Complexes with New Sortase Homologues

By addition of the β 7- β 8 substrate interacting loop from the SrtA_{staph} onto the core of the SrtA_{strep} core (SPS_{aureus}) we were able to narrow the substrate specificity profile such that only the LPATGG substrate was recognized. Though we did improve catalytic efficiency with this SPS_{aureus} construct and LPATGG, the substrate scope was still limited. Expanding this panel to include all 5th position amino acid substitutions allowed us to determine if the loop swapped constructs, SPS_{aureus} and SAS_{pneumoniae} had enhanced substrate profiles outside of the initial substrates tested. Based on initial results for our ‘loop swapped’ constructs SPS_{aureus} and SAS_{pneumoniae}, we were interested in exploring the usage of new β 7- β 8 substrate interacting loops from multiple distinctive sortase homologues. Previous results showed SPS_{aureus} to be a promising engineered construct, displaying a slightly increased substrate promiscuity compared to the SrtA_{staph} construct in which the β 7- β 8 loop originated from. Although this result was encouraging, we considered that the overall substrate and catalytic activity may be modulated by the addition of substrate interacting loops from new sortase homologues onto the SrtA_{strep} core.

The sortase family encapsulates thousands of sortase genes that may be studied and utilized for the purposes of SML but we wanted to identify a SrtA enzyme that has a differing selectivity and activity from those that are currently available to researchers. In a previously published study out of the Antos Lab at Western Washington University, our collaborators

identified and tested 6 new WT sortase homologues, *Streptococcus suis*, *Streptococcus oralis*, *Listeria monocytogenes*, *Enterococcus faecalis*, *Lactococcus lactis*, *Bacillus anthracis*. Overall these homologues, specifically *S. suis*, *S. pneumoniae*, *S. oralis*, and *L. monocytogenes*, displayed higher substrate promiscuity compared to that of SrtA_{staph} in the previously published study (31). This substrate selectivity was

measured by “hits” seen via mass spectrometry (MS) and informed an HPLC based assay in which total percent conversion was measured at a time point of 24 hrs for selected peptides (Figure 1-1, 1-8).

We engineered 6 new constructs utilizing these new sortase homologues $\beta 7$ - $\beta 8$ loops and the SrtA_{strep} core based on our previous loop swapped results which indicated improved catalytic efficiency for the SPS_{aureus} construct. By swapping in these new $\beta 7$ - $\beta 8$ loops we hoped to engineer a construct that would display improved catalytic activity and a more promiscuous substrate specificity profile. We utilized our high throughput kinetic fluorescence assay to assess of the impact these new sortase homologue $\beta 7$ - $\beta 8$ loops on the overall behavior of an engineered construct.

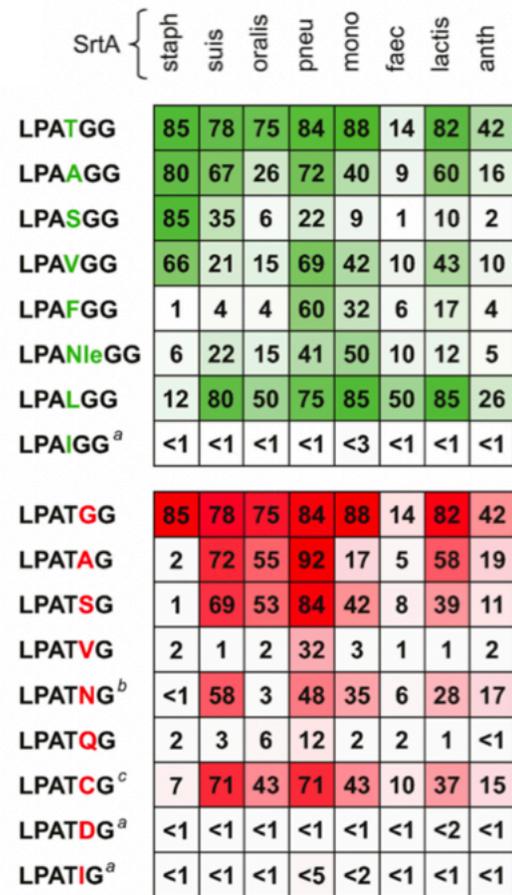


Figure 1-8. Heat map of substrate selectivity and catalytic activity of sortase enzymes isolated from differing bacterial species. Darker colors of green or red on the heat map indicate that cleavage had occurred and was measured via an HPLC based assay after 24 hrs when the enzyme and substrate were incubated together. Substrates with substituted residue (red) in either the 4th, or 5th position in the canonical LPATXG motif. Adapted from Nikghalb KD, *et al.*, 2018.

Initial results obtained via our fluorescence assay exhibited improved catalytic activity for the SPS_{lactis}, SPS_{faec}, SPS_{oralis}, and SPS_{suis} constructs with all of the representative 5th position substrates (LPATGG, LPATSG, and LPATAG) compared to the initial SrtA_{staph} and SPS_{aureus} constructs. For 4th position representative substrates (LPAAGG, LPAEGG, LPAIGG) catalytic activity was only observed for the LPAAGG substrate. For the constructs, SPS_{anth} and SPS_{mono} no catalytic activity was observed for any 4th or 5th position substitutions (Figure 1-9). The sequence similarity between catalytically inactive enzymes versus the catalytically active enzymes displayed a trend in β 7- β 8 loop length and loop composition related to catalytic behavior. The active SrtA enzymes have the same loop length (7 residues) and a 43% loop sequence identity to the SrtA_{strep} β 7- β 8 loop while the inactive enzymes have a longer loop length (8 residues) and have no loop sequence similarity to that from SrtA_{strep} (Figure 1-10). It seems that loop sequence and loop length may influence if the selected β 7- β 8 loop on a SrtA_{strep} core will exhibit catalytic activity with the representative panel of substrates for both 4th and 5th position substitutions. Notably, the mutant, SPS_{faec} showed promising catalytic activity for all of the 5th position representative substrates, leading us to expand our substrate panel to include all 20 amino acids for a 5th position substitution.

Abz-LPATXG-K(Dnp):

	<i>S. pneumoniae</i>	<i>S. aureus</i>	SASPneumoniae	SPSAureus	SPSLactis	SPSAnthracis	SPSFaecalis	SPSOralis	SPSSuis	SPSMonocytogenes
G	35	110	4	102	93	3	98	48	95	12
S	31	1	4	4	49	1	88	28	77	1
A	24	0	3	17	77	1	94	53	94	1

Abz-LPAXGG-K(Dnp):

	<i>S. pneumoniae</i>	<i>S. aureus</i>	SASPneumoniae	SPSAureus	SPSLactis	SPSAnthracis	SPSFaecalis	SPSOralis	SPSSuis	SPSMonocytogenes
A	9	76	1	104	48	1	65	13	56	5
E	0	0	0	1	0	0	0	0	0	0
I	0	0	0	13	0	0	0	0	0	0

Figure 1-9. Heat map of ‘loop swapped’ SrtA with new SrtA homologues.

Displays measured catalytic activity of SrtA enzymes with a 5th and 4th position substitutions (LPATXG and LPAXGG). Each “hit” corresponds to final percent conversion from starting material to product measured via florescent plate reader assay after 2 hrs. Darker shades of red/green indicate an enhanced overall percent conversion.

To better understand how the substrate specificity of our engineered constructs may be modulated by β 7- β 8 loop mutations, all of our constructs including the original loop swapped variants were tested against a 19 amino acid panel for 5th position substitutions (excluding Trp due to issues with peptide purification). We observed an overall trend matching that seen with the initial substrate panel (LPATGG, LPATSG, and LPATAG), wherein catalytic activity was only observed for SPS_{lactis}, SPS_{faec}, SPS_{oralis}, and SPS_{suis} and no catalytic activity was observed for SPS_{anth} and SPS_{mono} (Figure 1-11). Using a 20% conversion cut off, in the expanded panel the SPS_{lactis}, SPS_{faec}, and SPS_{suis} displayed improved catalytic activity for the LPATFG and

LPATNG substrates, while only SPS_{lactis} and SPS_{faec} displayed catalytic activity for all LPATVG, LPATLG, LPATFG, and LPATNG substrates.

SrtA Enzyme	β7-β8 Loop Sequence
<i>Staphylococcus aureus</i>	DDYEKTGVWEK
<i>Streptococcus pneumoniae</i>	EDLAATE
<i>Enterococcus faecalis</i>	GDLQATT
<i>Streptococcus suis</i>	TDYYATQ
<i>Streptococcus oralis</i>	VDYNATE
<i>Lactococcus lactis</i>	ADAEATH
<i>Bacillus anthracis</i>	VSVKDNSK
<i>Listeria monocytogenes</i>	DKPTETTK

Figure 1-10. β 7- β 8 loop sequences from WT SrtA enzymes and SrtA homologues.

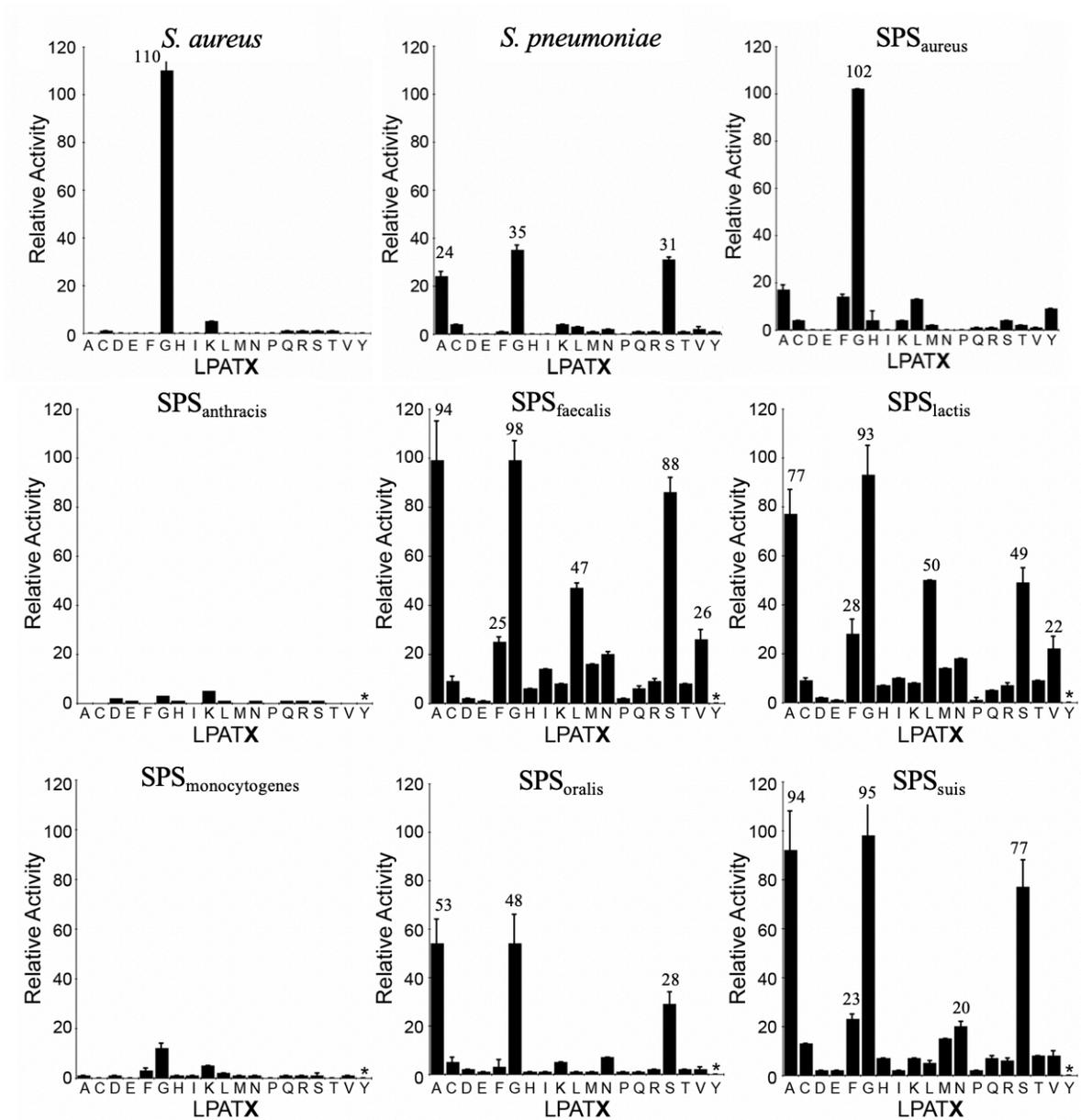


Figure 1-11. Expanded graphical representation of 'loop swapped' SrtA with new SrtA homologues. Displays measured catalytic activity of SrtA enzymes with a 5th position substitutions (LPATXG). Final percent conversion from starting material to product measured via florescent plate reader assay after 2 hrs. Percent conversions over 20% are labeled above bar. * indicates residue substitution was not determined.

Comparing the sequences between SPS_{faec} and SPS_{lactis} we see that both loop sequences contain the three residues Asp, Ala, and Thr, similar to that of SrtA_{strep} but also contain small, uncharged residues next to the catalytic cysteine in the active site (Figure 1-10, 1-11, 1-12).

On the other hand the construct SPS_{oralis} contains 4 out of 7 similar residues to SrtA_{strep} but its catalytic behavior across the substrate panel is lowered, implying that there may be specific residues in the β 7- β 8 loop that are influencing catalytic behavior to a large degree. The two residues in the loop sequence Asn and Val show no similarity between either the SrtA_{strep} or any other catalytically active loops. We also see that the SPS_{anth} β 7- β 8 loop contains a Val residue in the same position. When comparing the SPS_{mono} to the SPS_{anth} construct we actually see that SPS_{mono} is marginally active and does not contain a Val residue, though this could be coincidental. Future studies could explore the impact of swapping in a Gly residue for the Val residue, similar to that seen in the SPS_{faec} mutant.

The complete inactivity observed for the SPS_{anth} and SPS_{mono} constructs could potentially be due to the fact that both of these constructs are longer than the original SrtA_{strep} β 7- β 8 loop, possibly impacting the ability of these constructs from forming key residue interactions necessary for catalysis (Figure 1-10, 1-11). In addition, these sequences also have no sequence similarity compared to that of the SrtA_{strep} β 7- β 8 loop. Future experiments could shorten the β 7- β 8 loops to match the loop length of SrtA_{strep} or even substitute in residues closest to the active site in order to determine if site specific mutations could restore activity.

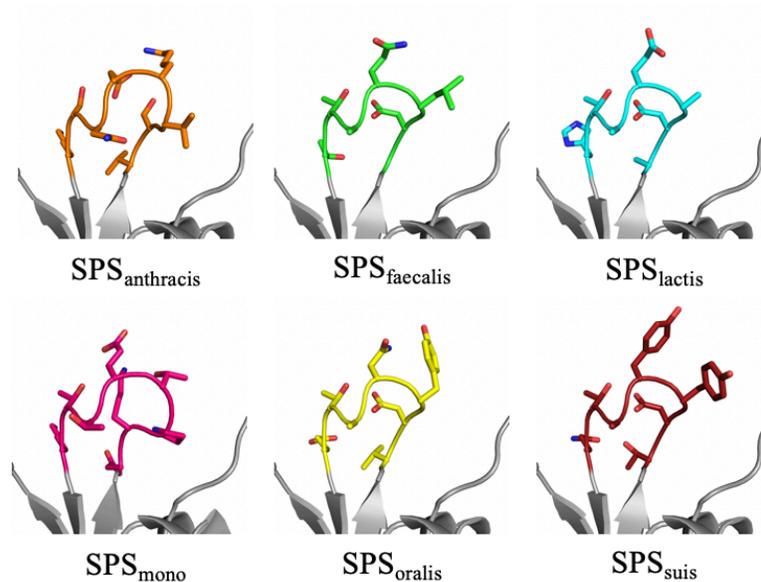


Figure 1-12. Swiss modeled ‘loop swapped’ complexes with new SrtA homologues. SrtA_{strep} core colored grey, β 7- β 8 loop colored by species (A) SPS_{anth}, (B) SPS_{faec}, (C) SPS_{lactis}, (D) SPS_{mono}, (E) SPS_{oralis}, (F) SPS_{suis}. Adapted from homology model made using SwissModel with PDB 3RCC used as the template structure

Based on the results obtained from the expanded sortase homologue panel, we identified our most promising construct, SPS_{faec}. SPS_{faec} had the most expansive substrate promiscuity profile with 7/20 amino acids meeting or exceeding the 20% conversion rate cut off for determining improved catalytic activity. In addition SPS_{faec} showed improved substrate promiscuity with a 10% conversion rate cut off, expanding the recognized substrate profile by an additional two residues. Though these additional two residues would not be as useful for protein engineering purposes due to their slower catalytic rate over the course of 2 hrs, they could potentially still be utilized for SML if run over the course of 24 hrs. The promiscuity of this construct was surprising as the WT *E. faecalis* is highly selective, only recognizing the LPATAG substrate when tested in the study out of the Antos lab via HPLC/MS (Figure 1-1, 1-8). By generating this ‘loop swapped’ construct with a *S. pneumoniae* core we produced a construct with the highest substrate promiscuity out of any of our SrtA enzymes.

With the generation of this SPS_{facc} construct we show that we were able to successfully engineer a unique SrtA enzyme which displayed an broader substrate specificity profile and overall improved catalytic efficiency (Figure 1-11). The vastly improved catalytic activity and promiscuity of this enzyme was achieved by changing only three residues from the initial SrtA_{strep} sequence, these being the Gly, Gln, and Thr in the loop sequence (GDLQATT). The origins of the effect of this loop swap are not entirely clear but we speculate that specific residue interactions on the *S. pneumoniae* scaffold may be the key to understanding the behavior of this construct, as detailed below.

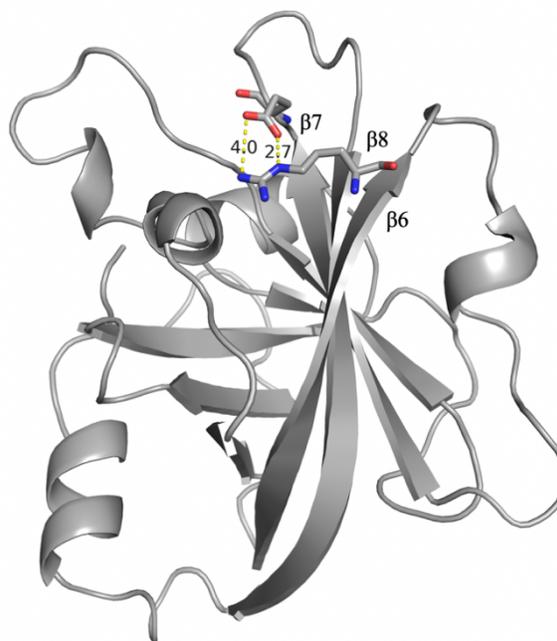


Figure 1-13. SrtA_{strep} with measured angstrom distances between residues Glu-128 and Arg-104. β 6, β 7, and β 8 loops are highlighted. Adapted from homology model made using SWISS Model with PDB 3RCC used as the template structure.

Identifying key residues that may play a role in maintaining catalytic activity is important for this study as the identification of specific residue positions

can possibly be employed by researchers to further mutate SrtA enzymes which currently struggle from nominal catalytic activity, such as our SPS constructs, SPS_{anth} and SPS_{mono}.

We identified one potential residue target for investigation, the Gly residue in the (GDLQATT) sequence displayed in SPS_{facc}. We hypothesized that this Glu-128 residue in SrtA_{strep} may be interacting with the Arg-104 residue near the N-terminus of the β 6- β 7 loop, reducing the flexibility of the loop, which may be deleteriously impacting substrate promiscuity

and overall catalytic activity (Figure 1-13). The numbering of these residues is based on the *S. pneumoniae* sequence, up until now we have only used *S. aureus* numbering. This Gly was mutated back to a Glu, similar to that displayed in the SrtA_{strep} sequence. Since SPS_{faec} displays high catalytic efficiency this Glu mutation near the active site may demonstrate if this interaction is impacting catalytic activity and promiscuity and identify Glu as a key residue which may limiting the overall catalytic activity and substrate specificity of SrtA_{strep} as well as demonstrating Gly near the catalytic Cys may be a necessity for a highly active and promiscuous enzyme.

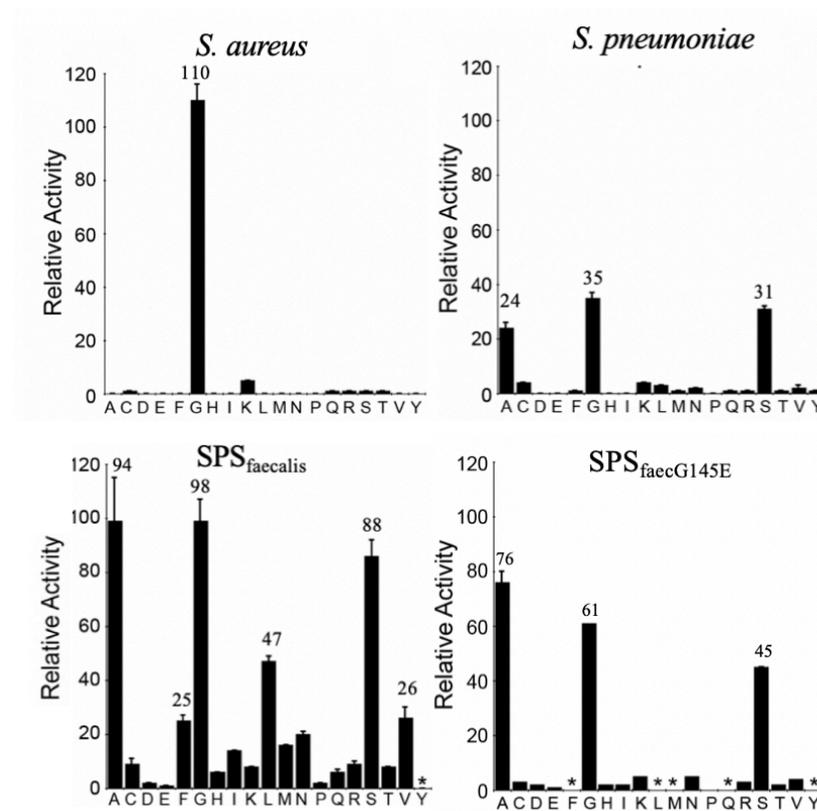


Figure 1-14. Expanded graphical representation of mutated SPS_{faecalis}. Displays measured catalytic activity of WT and the ‘loop swapped’ SPS_{faec} mutant SrtA enzyme with a 5th substitutions (LPATXG). Final percent conversion from starting material to product measured via florescent plate reader assay after 2 hrs. Percent conversion over 20% labeled over bar. * indicates peptide was not measured.

This mutated construct, SPS_{faecG145E}, resulted in an overall decrease in catalytic activity for the 5th position LPATGG, LPATSG, and LPATAG substrates (Figure 1-14). This decrease in catalytic activity was not as sizable as we expected but the overall reduction in catalytic

activity indicated that the negatively charged Glu residue near the catalytic cysteine is either limiting the activity by interaction with this Arg or that a small residue, such as Gly is necessary for activity with our SPS_X constructs. The impacts on substrate promiscuity were unknown due to limitations of enzyme and peptide availability.

The 4th position LPAAGG, LPAEGG, and LPAIGG substrates were also tested via the fluorescence assay and we observed a 3-fold drop in catalytic activity compared to SPS_{faec.}. The panel in general displayed lowered catalytic activity for the LPAAGG substrate and no catalytic activity was exhibited for either the LPAEGG or LPAIGG substrates

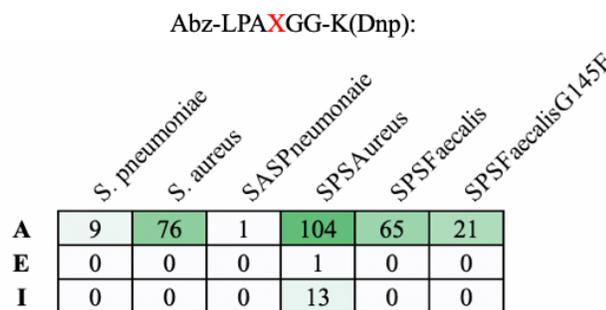


Figure 1-15. Heat map of mutated SPS_{faecalis}. Displays measured catalytic activity of WT, ‘loop swapped’, and SPS_{faec.} mutant SrtA enzymes with a 4th substitutions (LPAXGG). Each “hit” corresponds to final percent conversion from starting material to product measured via florescent plate reader assay after 2 hrs. Darker shades of green indicate an enhanced overall percent conversion.

(Figure 1-15). Future studies could address why these residues are not catalytically active as well as expanding the 4th position LPAXGG motif to include all 20 amino acid substitutions in order to identify any unexpected amino acid substitutions that may produce catalytic activity.

1.7 Enzyme Inactivity

The results of our enzymatic assay revealed multiple inactive SrtA enzymes. Though this inactivity may be truly due to the enzymes inability to recognize the substrates, we wanted to ensure that the inactivity was not due to outlying factors such as enzyme contamination with dimer or oligomeric fractions or protein expression/purification issues. We primarily focus on the inactive SAS_{pneumoniae} enzyme in this section as its counterpart, the SPS_{aureus} construct behaved so well under similar conditions.

Based on the initial results obtained via the kinetic fluorescence assay we observed a completely inactive SAS_{pneumoniae} enzyme (Figure 1-7). This result was intriguing due to the fact that we expected to observe an enzyme that may have displayed improved substrate promiscuity or catalytic activity, but instead catalytic activity was completely knocked out of this engineered construct. Originally thought to be due to a protein expression or purification issue, SAS_{pneumoniae} was re-expressed and re-purified under standard, non-denaturing conditions with similar results, indicating that there may be additional properties of the enzyme at play in determining catalytic activity.

Previous studies have indicated that the dimer and oligomeric states of sortase A are catalytically inactive (31). Due to potential perturbations of the monomeric identity of our enzymes by influence of freeze/thaw cycles we were uncertain if inactivity issues in our SAS_{pneumoniae} construct or in other constructs could be due to a transformation of the SrtA enzymes from a monomeric to dimeric/oligomeric state. Therefore, a control experiment was performed where previously purified protein samples were re-run over a S75 SEC column to gauge the level of monomer versus dimer or oligomer (Figure 1-16). This revealed that the majority of our

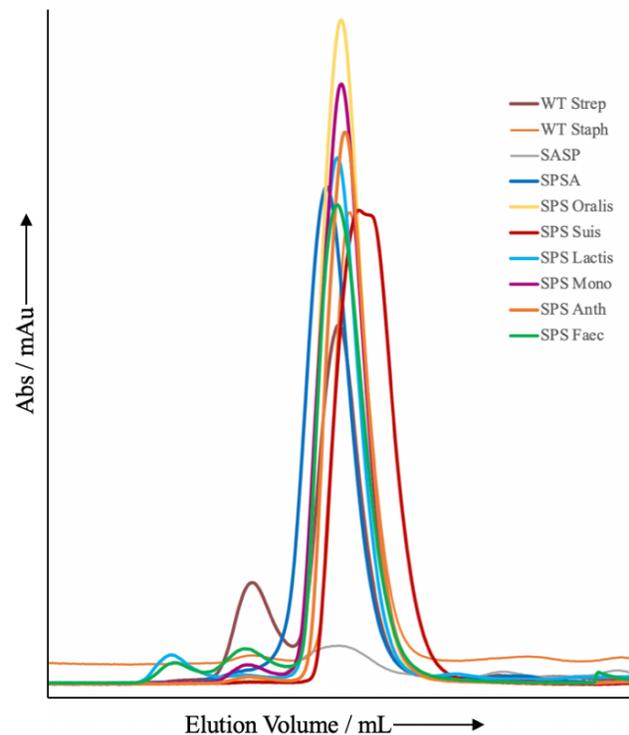


Figure 1-16. Isolation of monomeric species by SEC. SEC chromatogram illustrating isolation of monomeric complexes for WT SrtA enzymes and new SrtA homologues.

sortase enzymes did possess a monomeric identity, with only SrtA_{strep} displaying a potentially dimer or oligomer contaminated sample. When tested with our fluorescence assay SrtA_{strep} displayed no difference in overall activity compared to previous fluorescence data collected for the monomeric enzyme stock solution that was contaminated with dimer using the LPATXG substrate with either a Gly, Ser, or Ala substitution. Inconclusive results from this control experiment for SAS_{pneumoniae} led us to consider additional methodologies directly involving manipulation of the β 7- β 8 loop sequence when addressing causes of inactivity.

In the introduction, the significance of the tryptophan residue on catalytic activity in SrtA_{staph} was presented. Tryptophan shields the active site residues in the sortases' active site from the surrounding solvent, thus this Trp residue was a promising target for site specific mutation. Using our two previously tested constructs, SPS_{aureus} and SAS_{pneumoniae}, we engineered two new constructs with a T194W mutation in the construct SAS_{pneumoniae}T194W or a W194T mutation in the construct SPS_{aureus}W194T. When the Trp was added back into the SAS_{pneumoniae}T194W construct we expected to observe partially restored catalytic activity. But for the SPS_{aureus}W194T construct, we expected that swapping the Trp residue out of the SPS_{aureus} would greatly reduce catalytic activity for this enzyme. A result for either of these scenarios could indicate a potential correlation between the presence of this residue in our engineered constructs and overall catalytic activity.

The SPS_{aureus}W194T construct displayed a decrease in overall catalytic activity for the 5th position LPATGG substrate and the 4th position LPAAGG substrate but overall catalytic activity was not knocked out completely. The catalytic behavior for SPS_{aureus}W194T construct was in line with the proposed result of this mutation and these results support the findings from previous literature (Figure 1-17)(27). Results for the SAS_{pneumoniae}T194W construct displayed unresolved

inactivity with substitutions for both the LPATXG motif and the LPAXGG motif, wherein we expected to observe an increase in catalytic activity due to the restoration of this Trp residue (Figure 1-17). Sustained inactivity even with this Trp mutation and the reduction in activity observed for the SPS_{aureusW194T} indicates that this Trp residue may play a role in determining overall catalytic activity on the *S. aureus* scaffold.

To address loop length differences and β 7- β 8 loop composition as a potential cause of catalytic inactivity in our SAS_{pneumoniae} construct we engineered SrtA_{staph} in which the β 7- β 8 loop (DDYEKTGVWEK) was truncated by removing the (EKTG) residues resulting in a loop length which is similar to that of SrtA_{strep}. The (EKTG) portion of the β 7- β 8 loop was selected due to its distance from the active site of the enzyme and was assumed to not play a substantial role in substrate recognition as it was not near the active site of SrtA_{staph}.

The new SA Δ EKTG construct was tested for both 4th and 5th position substitutions using the representative substrates, LPATGG, LPATSG, and LPATAG (5th position) and LPAAGG, LPAEGG, and LPAIGG (4th position) to determine if this mutation could provide insight as to why the SAS_{pneumoniae} construct was catalytically inactive. The truncation of the (EKTG) residues completely knocked out catalytic activity for the 5th position LPATGG substrate and the 4th position LPAAGG substrate, compared to the SrtA_{staph} control which displayed 100% total conversion over 2 hrs for both of these substrates (Figure 1-17). This reduction in catalytic activity with a truncation of the β 7- β 8 loop could be due to a reduction in loop flexibility that may be necessary for substrate recognition in SrtA_{staph} by means of key residue interactions. In addition, the (EKTG) residues may play an unknown role in substrate recognition.

Rather than site specifically engineering a new $\beta 7$ - $\beta 8$ loop to potentially restore activity of the SAS_{pneumoniae} construct, we instead opted to utilize an existing $\beta 7$ - $\beta 8$ loop from a sortase homologue, SrtA from *Streptococcus suis*, due to the $\beta 7$ - $\beta 8$ loop similarities to that of SrtA_{strep}. The sequence from *S. suis* (TDYYATQ) has the same number of residues and a roughly 43% loop sequence identity (3 out of 7 residues) to that of SrtA_{strep}, (EDLAATE). By use of this sortase homologue $\beta 7$ - $\beta 8$ substrate interacting loop we hoped to observe a partial or full restoration of catalytic activity in a construct with the SrtA_{staph} core and identify key residues that may influence catalytic activity, especially those closest to the catalytic active site.

The mutated construct, SAS_{suis}, with the *S. suis* $\beta 7$ - $\beta 8$ loop exhibited no catalytic activity when tested for both 4th and 5th position substitutions using the representative substrates LPATGG, LPATSG, LPATAG (5th position) and LPAAGG, LPAEGG, LPAIGG (4th position), similar to that of the SA Δ EKTG construct (Figure 1-17).

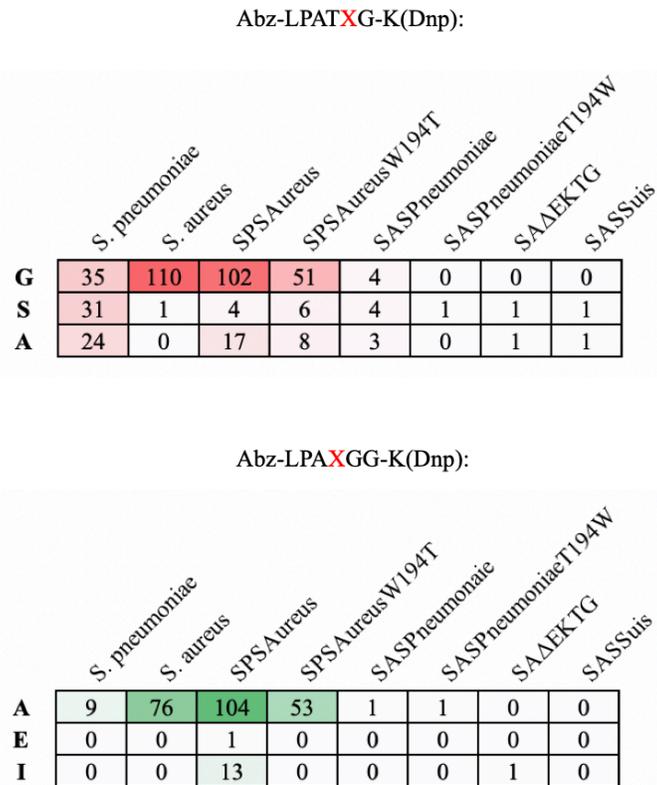


Figure 1-17. Heat map of ‘loop swapped’ enzymes with a Trp mutation, truncated loop, or *S. suis* $\beta 7$ - $\beta 8$ loop. Displays measured catalytic activity of SrtA enzymes with a 5th and 4th position substitutions (LPATXG and LPAXGG). Each ‘hit’ corresponds to final percent conversion from starting material to product measured via florescent plate reader assay after 2 hrs. Darker shades of red/green indicate an enhanced overall percent conversion.

We were unable to restore catalytic activity in our SAS_{pneumoniae} construct and we are still unsure as to what specific issue is resulting in this catalytically dead enzyme but we hypothesize that both loop composition and substrate interacting loop length, especially on the *S. aureus* scaffold, may be responsible for substrate recognition and catalytic activity.

1.8 Preliminary HDX Experiments

Promising results for the SPS_{faec} construct and the identification of key residues in the β 7- β 8 loop that may be modulating activity assists our understanding of the localized effects of this ‘loop swap’. To comprehend the bigger picture of how this chimeric construct behaves, and the dynamic movement experienced by this β 7- β 8 loop and the surrounding substrate interacting loops we employed Hydrogen Deuterium Exchange (HDX) to investigate dynamic substrate loop movement and conformational changes that relate to SrtA function.

As described previously, before substrate binding occurs, the apo-SrtA β 7- β 8 and β 6- β 7 loops are unstructured and flexible. When the substrate binds, the β 6- β 7 loop will then adopt a conformation in which the side chains of Val-168 and Leu-169 are rotated away from the body of the protein and the complex orients into the conformation wherein the β 7- β 8 loop is partially displaced, leaving room for binding of the sorting signal where the active site catalytic Cys, Cys-184, is positioned to cleave between the Thr and Gly residues of the LPXTG motif (27).

Previous results in our study indicated that β 7- β 8 loop flexibility and length may be a key component mediating catalytic activity, exemplified by the SA Δ EKTG mutant in which catalytic activity was knocked out by truncating the SrtA_{staph} loop by 4 residues, (EKTG). Investigating the flexibility and dynamic movement of these β 7- β 8 and β 6- β 7 loops should help us evaluate

why some of our constructs are catalytically inactive and also why our highly successful mutant, SPS_{faec}, performs so well with many 5th position substitutions in the LPATXG motif.

One way one we can examine this loop flexibility and dynamic movement is through hydrogen deuterium exchange (HDX). HDX is a biophysical technique that allows researchers to investigate the dynamic loop movements and structural characteristics of proteins. HDX can be used to examine protein conformations, identify substrate binding sites, and investigate the dynamics of protein domains (39–41).

In a protic solution covalently bonded amide hydrogens of the protein backbone with exchangeable protons will exchange with the deuterated solvent, incorporating deuterium in at these positions, causing a mass change. The rate of exchange the detectable by MS. This “exchange” allows for detection of dynamic movement by means of measuring increased or decreased hydrogen-exchange and the mass change is dependent on the folded state of the protein and exposed loop surface area, where more shielded areas will experience less exchange from protons to deuterium. The stability of the hydrogen bonding networks and the chemical properties sequence also play a role in the rate of exchange (42, 43).

Partnering with a PhD candidate, Helen Hobbs in Susan Marqusee’s Lab at UC Berkeley we tested numerous constructs including our WT SrtA enzymes and our promising enzyme candidate, SPS_{faec}, to determine how the improved activity of the engineered construct would correlate to flexibility and dynamic movement of its loops.

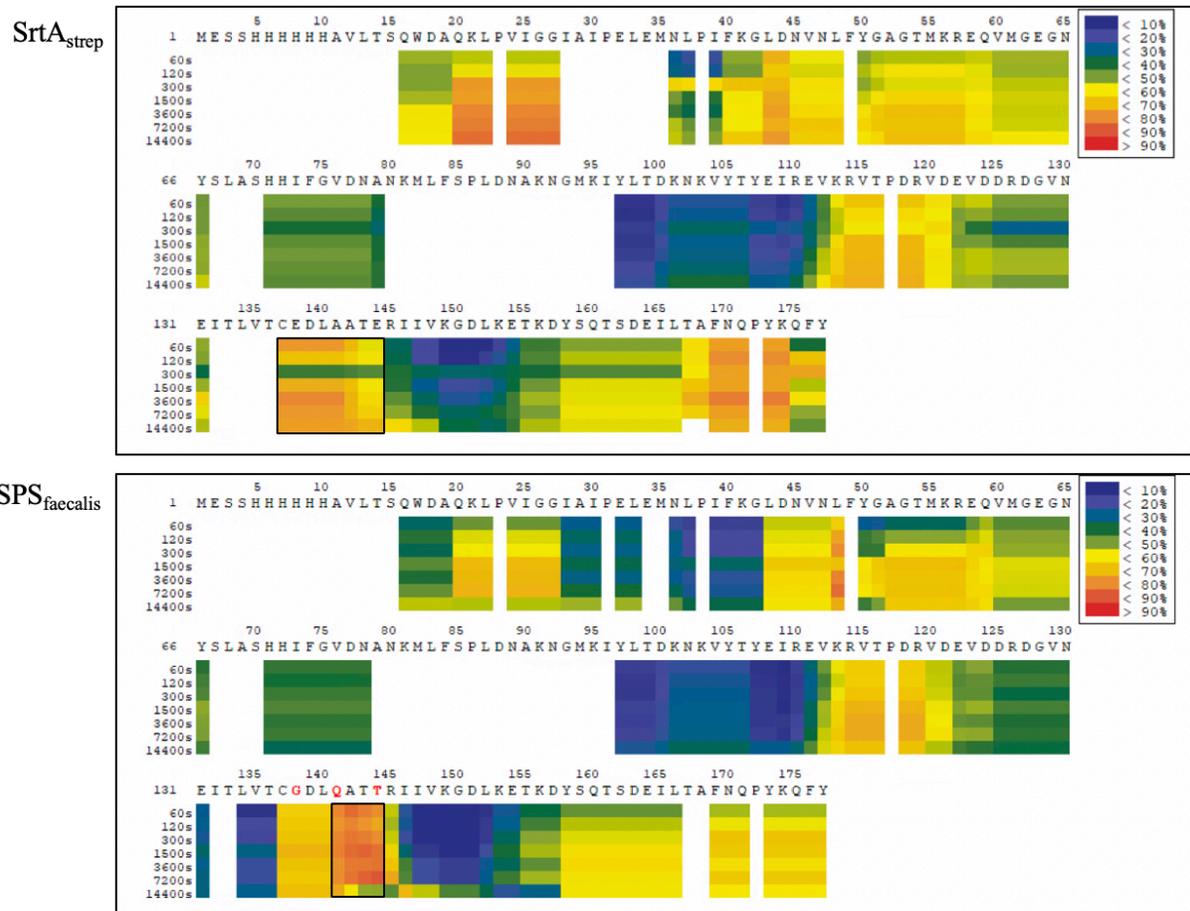


Figure 1-18. Deuterium level of SrtA_{strep} and SPS_{faecalis} residues at different time points. Regions in the β 7- β 8 loop with enhanced rates of exchange are boxed in black. Scale of % deuteriation listed from 10%-100% in increments of 10%.

Results of the HDX experiments showed that in our SPS_{faec} mutant there was an enhancement in the percent deuteriation indicating that this β 7- β 8 loop region may be more flexible and dynamic (Figure 1-18). The improvement in the flexibility of this loop may help the SPS_{faec} accommodate a broader scope of substrates compared to the SrtA_{strep}. Within this loop, most of the enhancements in the flexibility seem to be focused around the C-terminus of this loop sequence, specifically near the Gln and Thr residues. Though, the residue where we expected to potentially see a change in the flexibility was the Gly residue in SPS_{faec} based on the results from the SPS_{faec}G145E mutant but we did not see this in the deuteriation heat maps.

Conceivably this residue change may “loosen up the C-terminus of the loop. This increase in flexibility near the C-terminus may be resulting in a loop-loop interaction or a change in structure that accommodates additional substrates. Though this experiment was limited as we were only able to test the apo state of our enzymes. To really get a deeper understanding of the differences in loop dynamics and draw conclusions about the behavior of these enzymes we should test both bound and unbound states of the SPS_{face} and SrtA_{strep}.

1.9 Dermcidin Experiment

Our project has predominantly focused on determining if these newly engineered constructs were able to form the cleavage product by either observing a product peak for (XG-K(Dnp)) via

HPLC or producing a

fluorescent signal (Abz) via our

plate reader assay, and

correlating these results to

overall enzymatic behaviors.

Though useful for making

general observations about

these new enzymes, we were

limited by the scope of this

assay. Therefore, we were

interested to explore how these

new constructs, specifically the SPS_{face}, would behave when used in a ligation method

application which expands beyond the scope of simple model substrates.

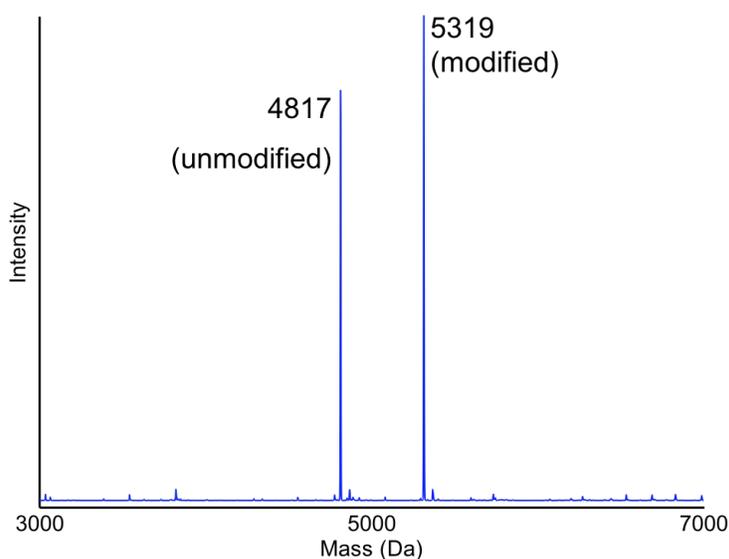


Figure 1-19. Dermcidin modification experiment. MS of dermcidin experiment utilizing a five-fold molar excess of F* tagged peptide, FITC-Axh-LPATSG in combination with the SPS_{face} enzyme after 2 hrs. The modified peak indicates the presence of a modified N-terminus.

This experiment utilized dermcidin-1L (DCD-1L), a small 48 residue antimicrobial peptide (roughly 4.8 kD), which has a naturally occurring N-terminal Ser residue. The utilization of this peptide is indicated by results obtained with our SPS_{faec} enzyme and the LPATSG substrate where improved catalytic activity was observed via the fluorescence plate assay. In previous experiments performed in the Antos lab, DCD-1L has shown the ability to be ligated to the LPATSG substrate using SrtA and Ser also displays as a naturally occurring N-terminal nucleophile in DCD-1L.

This DCD-1L was incubated with a five-fold molar excess of fluorescently tagged peptide, FITC-Axh-LPATSG, and our SPS_{faec} enzyme. We observed a conversion from our unmodified DCD-1L to a modified N-terminal DCD-1L product. This conversion was monitored by LC-ESI-MS (Figure 1-19).

The ability our SPS_{faec} enzyme to not only form a cleavage product with a variety of 5th position substitutions (LPATGG, LPATSG, LPATAG, LPATVG, LPATLG, and LPATEFG) but also successfully progress through the transacylation reaction and form the ligation product when utilized in the DCD-1L modification scheme, similar to that used by the Antos lab, is indicative that this enzyme can be used for the purposes of sortase-mediated ligation in a research setting, potentially out-performing previously engineered constructs. Future experiments could apply this same dermcidin experiment to our other constructs which displayed improved catalytic activity, SPS_{lactis} and SPS_{suis}.

1.10 Expansion of Sortase Utilization and Concluding Remarks

By means of engineering existing SrtA enzymes we produced a novel sortase by loop swapping a β 7- β 8 loop from *E. faecalis* onto the core of the *S. pneumoniae*, producing an enzyme, SPS_{faec}, that not only displayed an improved catalytic profile but in addition had vastly

expanded substrate promiscuity compared to WT SrtA_{strep} and SrtA_{staph}. Though we had additional engineered enzymes, specifically SPS_{lactis} and SPS_{suis}, which showed enhanced catalytic activity and a relatively similar substrate promiscuity, SPS_{faec} was the construct which we considered to be the most promising for future protein engineering endeavors. This consideration was supported by the behavior of SPS_{faec} when tested against the antimicrobial peptide dermcidin, in which SPS_{faec} was able to successfully perform the transacylation reaction, producing a modified N-terminal DCD-1L product. The ability of a SrtA construct to ligate is necessary for the labeling of antibodies with small molecule labels or the formation of antibody drug conjugates, such as used with sortase-mediated antibody drug conjugation technology (SMAC-technology)(12).

We have also identified key structural components and residues that may be modulating the activity of our SrtA constructs. The HDX experiment reasserted this β 7- β 8 loop as a region of flexibility in our SPS_{faec} enzyme. In addition, loop length seems to possibly play a role in determining overall catalytic activity and substrate promiscuity of our SPS_X constructs. The inactive constructs, SPS_{mono} and SPS_{anth}, both exhibit longer β 7- β 8 loops compared to their active counterparts, SPS_{faec}, SPS_{suis}, and SPS_{lactis}. In addition, our ‘inactive’ SAS_{pneumoniae} construct followed this same trend where the shorter *S. pneumoniae* loop which was engineered on to the SrtA_{staph} core resulted in no catalytic activity observed. When testing this with our SA Δ EK_{TK} construct, a shortened SrtA_{staph} loop completely knocked out catalytic activity, reinforcing the idea that overall loop structure and length may be modulating substrate interaction and resulting catalytic activity. The key residues that seem to possibly be regulating the behavior of our constructs are those nearest to the active site of the SrtA enzyme. When we tested the mutated enzyme, SPS_{faecG145E}, the slight reduction in catalytic efficiency illustrated that the Glu in the

SrtA_{strep} β 7- β 8 could be interacting with the Arg residue, limiting catalytic activity. This Glu residue seems to be necessary for effective catalysis.

Future studies should first, explore the impacts of additional site-specific mutations in both our active SPS_{faec} mutant and our ‘inactive’ enzymes to determine if loop composition and/or loop length is the determining factor of catalytic activity and substrate promiscuity. In addition these site-specific mutations may reveal if we may restore the activity of the SAS_{pneumoniae} enzyme. Additionally, the other constructs SPS_{suis} and SPS_{lactis} which displayed catalytic activity and promiscuity close to that of SPS_{faec} should be tested to determine if they are also able to form a modified N-terminal DCD-1L product.

The engineering of the *S. pneumoniae* β 7- β 8 loop by means of a ‘loop swap’ to produce a more promiscuous and catalytically active enzyme as well as identifying potential residue interactions that may be limiting WT *S. pneumoniae* catalysis illustrates a compelling opportunity to further explore sortase biology and the role of the β 7- β 8 loop in the biochemical characteristics of this class of enzymes. Furthermore, sortases with altered substrate specificity are of interest to protein engineers as they expand the applications of SML and our findings may assist in the design of sortase for future engineering purposes.

Materials and Methods

Instrumentation

Protein purification by immobilized metal affinity chromatography (IMAC) and size exclusion chromatography (SEC) was conducted on a GE AktaPrime Plus FPLC system with a GE Healthcare HisTrap HP column (5 x 5 mL) for IMAC in either a Ni²⁺-NTA wash buffer (50 mM Tris pH 7.5, 150 mM NaCl, 20mM Imidazole pH 7.5, 1mM TCEP) or a Ni²⁺-NTA elution buffer (50mM Tris pH 7.5, 150mM NaCl, 300mM Imidazole pH 7.5, 1mM TCEP). For SEC we used a HiLoad 16/600 Superdex 75 pg column in SEC running buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM TCEP).

RP-HPLC purifications and analyses were performed on a Dionex Ultimate 3000 HPLC system. Phenomenex Kinetex® 2.6 µm C18 100 Å column (100 x 2.1 mm), aqueous (95% H₂O, 5% MeCN, 0.1% formic acid) / MeCN (0.1% formic acid) mobile phase, flow rate = 0.3 mL/min, hold 10% MeCN (0.0-0.5 min), linear gradient 10-90 (0.5-7.0 min), hold 90% MeCN (7.0-8.0 min), re-equilibrate at 10% MeCN (8.0-13.5 min).

For LC-ESI-MS analyses, the Dionex Ultimate 3000 HPLC was interfaced with an Advion CMS expression mass spectrometer. LC-ESI-MS data were analyzed using Advion Data Express software.

Kinetic fluorescence assays were conducted on a BioTek Synergy H1 Microplate Reader. Excitation: 320, Emission: 420, Gain: 75, Light source: Xenon flash lamp.

HDX experiments were performed using a LEAP H/D-X PAL robotic automatic sampling system. Tandem MS and HPLC was performed on a Thermo Scientific LTQ Orbitrap Discovery system. Data analyzed using HDExaminer software from Sierra Analytics.

SrtA Mutant Cloning

All SrtA mutant clones were purchased from Genscript. Wild type SrtA from *Staphylococcus aureus* and *Streptococcus pneumoniae* were from the Antos Lab at Western Washington University. Cloning errors did occur with a number of our SrtA sequences, experimental data did not seem to be affected by an additional His-tag added on the N-terminus. Sequences for all constructs are listed in the Appendix.

Expression and Purification of SrtA Wild Types and Mutants

E. coli BL21(DE3) competent cells were transformed with pET-28a(+) vector plasmids for all sortase constructs, both the WT and mutated constructs (all uncleaved molecular weights and extinction coefficients are listed in Table 3).

Transformed cells were plated on an agar plate (200 μ L) with kanamycin (KAN) (50 μ g/mL) resistance and grown at 37° C overnight. Single colonies were then selected and cultured overnight in 10mL of Luria Broth (LB) supplemented by 50 μ g/mL KAN. These overnight cultures were then used to inoculate 1000mL of LB media, supplemented by 50 μ g/mL KAN, and grown to an optical density (OD₆₀₀) between .6-.8 at 37°C. When the desired OD was reached, temperature was changed to 18°C and growths were inoculated with 150 μ L of 1 M Isopropyl β -D-1-thiogalactopyranoside (IPTG) and induced overnight. Cell pellets were obtained by centrifugation (4,000rpm, 10 min, 4°C), the cell pellet was then resuspended in lysis buffer (50mM Tris pH 7.5, 150mM NaCl, 0.5mM ethylenediaminetetraacetic acid (EDTA)). The resuspended cell pellets were then lysed by sonication and the lysate was then centrifuged (17,500 rpm, 30 min, 4°C). The resulting supernatant was run over a nickel nitrilotriacetic acid (Ni²⁺-NTA) resin column equilibrated with Ni²⁺-NTA wash buffer (50 mM Tris pH 7.5, 150 mM NaCl, 20mM Imidazole pH 7.5, 1mM TCEP). Using fast phase liquid chromatography (FPLC),

the column was rinsed with wash buffer for 50 mL to eliminate non-specifically bound proteins and then, running a gradient elution using wash buffer and elution buffer (50mM Tris pH 7.5, 150mM NaCl, 300mM Imidazole pH 7.5, 1mM TCEP) from 0-100% buffer over a 50mL elution a peak was obtained, indicating the presence of our desired protein. The peak was collected and concentrated down to desired volume utilizing a 10kD molecular weight cut off (MWCO) ultrafiltration device (Millipore).

Concentrated and partially purified protein was then loaded onto a size exclusion column (SEC) equilibrated with running buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM TCEP) to obtain three peaks corresponding to oligomer, dimer, and monomeric fractions, the monomeric fractions were only collected for the purpose of this study (Figure 1-20). Protein fraction purity was analyzed by tricine gel (Figure 1-21,1-22). We noticed slight degradation over time from proteins that were stored in 4 °C for longer than a week. Using a 10 kD MWCO ultrafiltration device, the purified monomeric fraction concentration was calculated at A280 absorbance on a NanoDrop Lite Spectrometer utilizing extinction coefficient information. MS analysis was performed on initial WT and 'loop swapped' constructs to confirm identity (Figure 1-23).

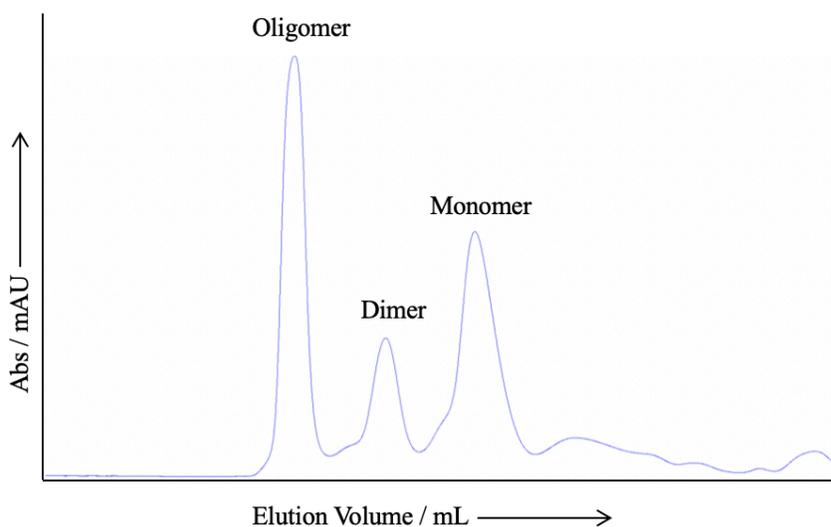


Figure 1-20. FPLC chromatogram of SAS_{pneumoniae}. Elution off SEC column delivers three peaks (oligomer, monomer, and dimer). Monomeric fractions were collected for kinetic fluorescence assays.

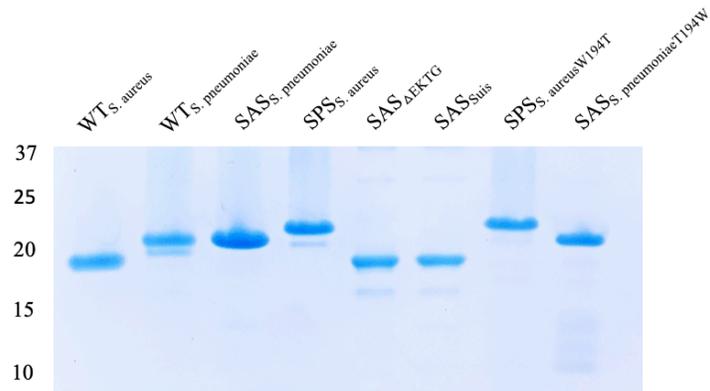


Figure 1-21. Gel image of SrtA enzymes. Tricine gel showing purified SrtA protein samples of WT, initial ‘loop swapped’ constructs, and *S. aureus* tryptophan mutants.

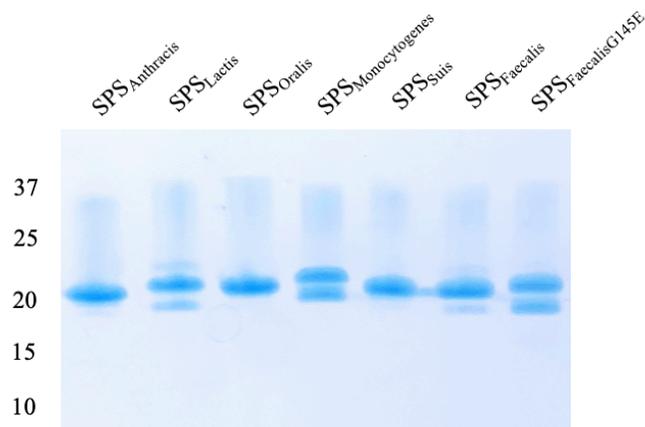


Figure 1-22. Gel image of SrtA enzymes. Tricine gel showing purified SrtA protein samples of six sortase homologues and the SPS_{faecG145E} mutant.

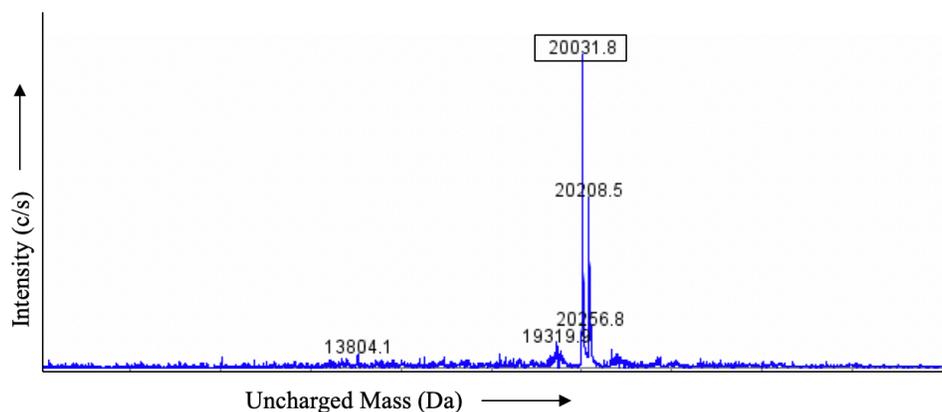


Figure 1-23. Mass spectrometry analysis of 'loop swapped' SAS_{pneumoniae}. Expected mass: 20,032.39 Da.

Peptide Synthesis

We used a widely accepted scheme for solid phase synthesis similar to that utilized in the 2017 paper out of the Antos Lab (31). Briefly, model peptides (Abz-LPATXG-K(Dnp) and Abz-LPAXXGG-K(Dnp)) were synthesized using solid phase synthesis. The utilization of Fmoc protecting groups allowed for a stepwise addition of amino acids. The synthetic scheme began with Fmoc-protected rink amide solid support. We have used both resin and synphase lanterns as solid supports for the purposes of this project, but the advantage of synphase lanterns is the ability to create multiple distinct peptides in tandem. The base-labile Fmoc was removed using a 20% piperidine/NMP mixture, followed by additional NMP washes to ensure that excess reagents/amino acids are washed out from the previous step, and then an additional Fmoc protected amino acid was coupled to the deprotected amine using a mixture of Fmoc-K(Dnp)-OH, HBTU, DIPEA, NMP. These deprotection and addition steps were repeated until all the desired amino acids had been added. In addition, the chromophores were added to the N and C terminus of the peptide to allow for reaction monitoring when performing the sortase catalyzed transpeptidation reaction (Scheme 2). When the desired sequence had been synthesized, the

peptide was cleaved off of the solid support using a mixture of TFA:TIPS:H₂O/95:2.5:2.5. Model peptides were purified vis RP-HPLC and their identities were confirmed using mass spectrometry (Figure 1-24, 1-25, 1-26).

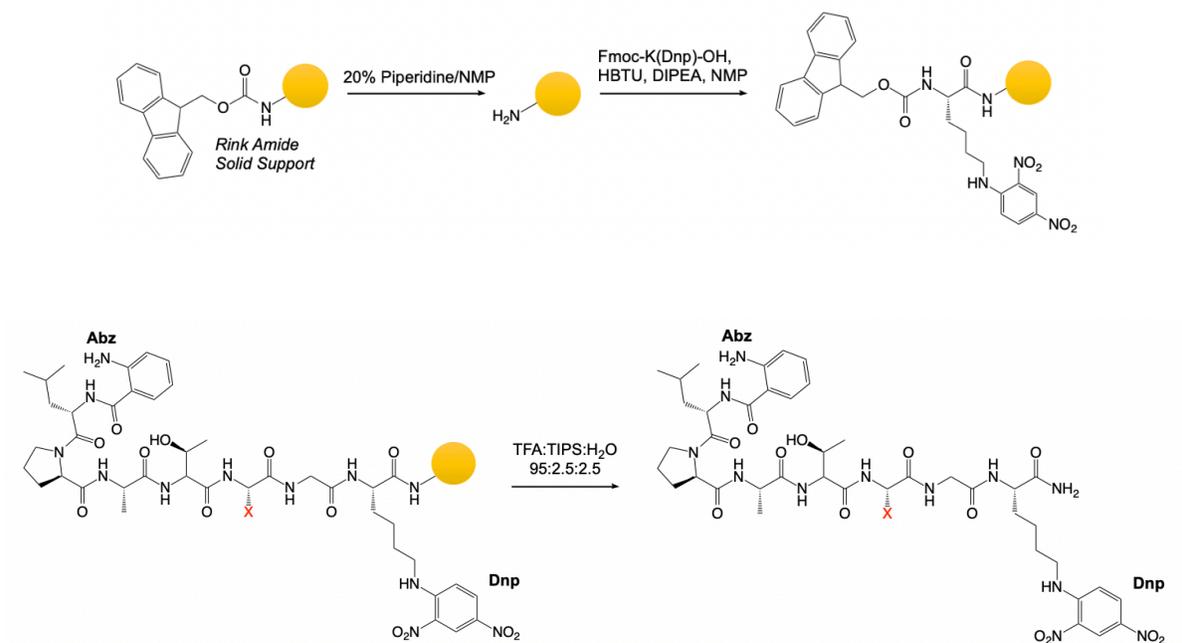


Figure 1-24. SrtA peptide synthesis scheme. Synthesis of LPATXG and LPA_XGG SrtA peptides utilized with HPLC/MS and F* plate reader assay.

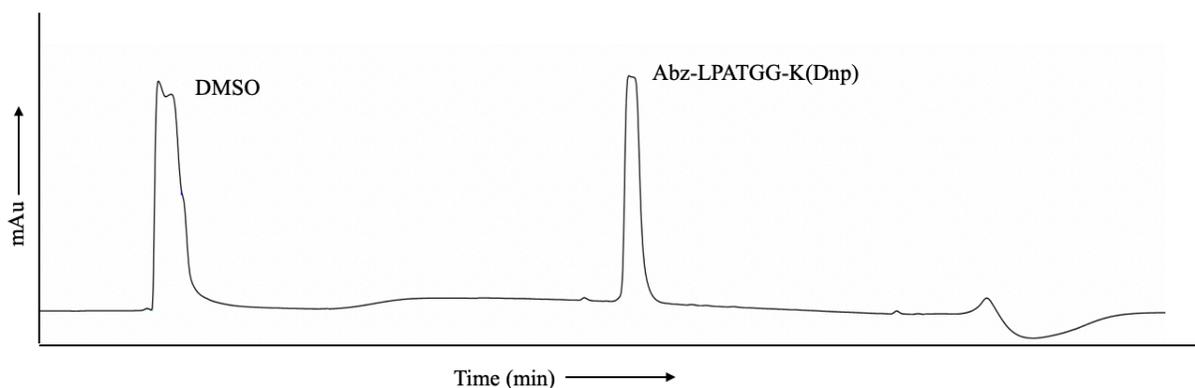


Figure 1-25. HPLC trace of Abz-LPATGG-K(Dnp) peptide.

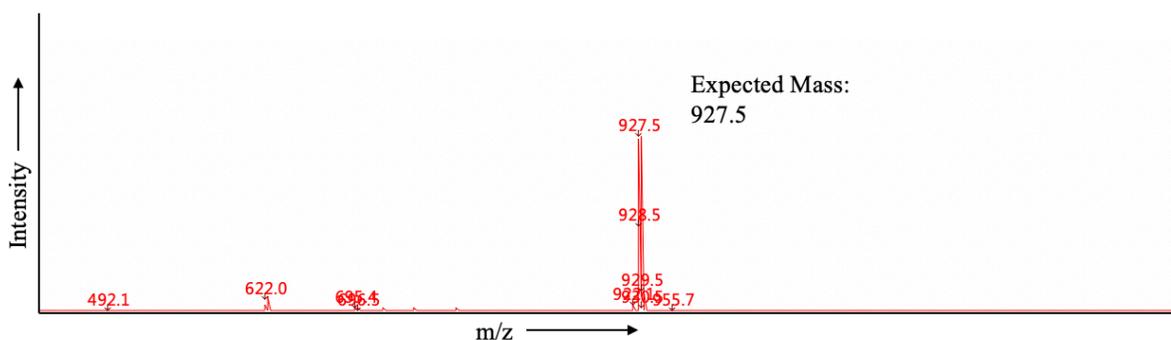


Figure 1-26. MS spectrum of Abz-LPATGG-K(Dnp) peptide. Expected m/z=927.5.

Kinetic Enzyme Assays

Protocol 1: HPLC

Individual reaction pools of 100 μ L containing, 50 μ M peptide (LPATXG or LPA \underline{X} GG, X denotes any amino acid substitution), 5 μ M sortase, 5 mM NH₂OH, 10% (v/v) 10x sortase reaction buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl₂) as well as residual glycerol (<6%, v/v) and DMSO (\leq 5%, v/v) from sortase and peptide substrate stock solutions were mixed. Reactions involving cysteine containing peptide substrates were supplemented with 1mM TCEP to prevent undesired disulfide bond formation. Using the established protocol from the Antos Lab for reaction monitoring we used RP-HPLC (Phenomenex Kinetex 2.6 μ M 100 A C₁₈ column, 3.0 x 100 mm) with a H₂O (0.1% Formic Acid)/MeCN (5% 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 0.5-6.0 min, hold 90% MeCN 6.0-7.0 min) and by LC-ESI-MS. To determine overall percent conversion, peak areas for the starting material and product, measured at 365 nm on the RP-HPLC chromatogram, were compared.

Protocol 2: Plate Reader, Kinetic Enzyme Assay

Individual reaction pools of 100 μL containing, 50 μM peptide (LPATXG or LPAXGG, X denotes any amino acid substitution), 5 μM sortase, 5 mM NH_2OH , 10% (v/v) 10x sortase reaction buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl_2) as well as residual glycerol (<6%, v/v) and DMSO (5%, v/v) from sortase and peptide substrate stock solutions were mixed. Reactions involving Cys containing peptide substrates were supplemented with 100mM TCEP to prevent undesired disulfide bond formation. The reaction mixture was combined without sortase into a 96 well plate. Immediately before the start of the 2 hr time run 10 μL of a 10X stock sortase enzyme (5 μM) was added, allowing for precise monitoring of the cleavage reaction start point and progress. Each set of reactions was measured over a 2 hr time period via fluorescent output readings obtained at 2 min intervals on a Synergy H1 Microplate Reader. These unitless fluorescent values were then compared against the calibration curve equation ($y=577.45x+1243.3$) calculated via standardized UV-vis data obtained from the HPLC to obtain an overall percent conversion for each SrtA reaction. The standardized UV-vis values are readings obtained the WT type SrtA enzymes, SrtA_{staph} and SrtA_{strep}, measured over a 2 hr time period in 30 min time intervals when in combination with the substrates LPATGG, LPATSG, or LPATAG by HPLC providing a percent conversion from starting material (Abz-LPATXG-K(Dnp)) to the cleavage product (XG-K(Dnp)).

Hydrogen Deuterium Exchange (HDX-MS)

HDX experiments and digestions were completed utilizing the robotic LEAP H/D-X PAL automated sample preparation system. Peptic peptide mass fingerprinting from purified SrtA samples (SrtA_{staph}, SrtA_{strep}, and SPS_{faec}) was performed using an online pepsin and fungal protease digestion. This was immediately followed by RP-HPLC and tandem MS (Thermo

Scientific LTQ Orbitrap Discovery) to identify a list of common peptide peptides and retention times. For the HDX experiments the SrtA samples were diluted 1:10 in D2O buffer (50 mM Tris, 150 mM NaCl, pD 7.5). At time points 60 s, 120 s, 300 s, 1500 s, 3600 s, 7200 s, and 14,400 s deuterated aliquots were quenched with quench buffer (3.5 M GdnHCl, 1.5 M Glycine, pH 2.5). These samples were then digested on column as described previously and analyzed by LC-MS. Data was analyzed using HDExaminer (Sierra Analytics).

Chapter 2: Ancestrally Reconstructed Sortase A

Introduction

2.1 Ancestral Sequence Reconstruction

Ancestral Sequence Reconstruction (ASR) is a technique utilized by researchers to investigate the evolution of structure-function relationships of protein families. ASR has allowed protein scientists to bridge the gap “between mechanistic biochemistry and evolutionary biology” (44). ASR involves identifying key evolutionary relationships via a statistical analysis of amino acid substitutions in which the probability of replacing any amino acid with another amino acid is calculated. At the ancestral nodes on a phylogenetic tree a maximum likelihood (ML) sequence is calculated along with a confidence score for that residue substitution. This sequence is the most likely to have generated the following sequences that are observed in more current proteins (45, 46). The probabilistic ML method has been more commonly used in recent ASR studies due to the more statistically reliable information it provides. This is in contrast to previous methodologies which utilize the maximum parsimony (MP) method, in which the phylogenetic tree with the least amount of amino acid substitutions was selected for sometimes leading to inaccurate conclusions regarding homoplasy of these enzymes (34). This amino acid sequence can then be encoded in a DNA plasmid and expressed and purified recombinantly. ASR allows researchers to analyze the activity of ancestral sequences and the functional changes that result from evolved mutations as well as exploring the sequence space between enzymes (47). Though there is an expected uncertainty to the validity of reconstructed sequence as there is no way to be fully confident that an ancestrally reconstructed sequence would match that of the actual protein which existed so many years ago but the general biochemical properties of these enzymes can still allow researchers to rationalize the data and form conclusions about potential enzymatic behavior (46). ASR is a powerful tool that aids researchers in tracing the ancient

mutations that led to current functional sites, structural features such as ligand binding pockets and loop dynamics, as well as biochemical characteristics seen in their extant relatives. By filling in the natural sequence space between extant enzymes and adding these sequences to a MSA researchers are able to predict functional sites and detect homologues in database searches (48, 49).

Ancestral proteins found in bacteria such as Sortase A from *Streptococcus pneumoniae* or *Staphylococcus aureus* along with other eubacteria, archaea, yeast, and vertebrates have been hypothesized to exist roughly between several million to around 3 billion years ago (34). The first studied examples of ancestral enzymes were translation elongation factors from organisms that lived roughly 3.5 billions years ago (35). Ancestral sequences reconstructed via ASR exhibit a pattern of expanding substrate specificity at older branch nodes on a phylogenetic tree, indicating that more ancestral sequences may have a more promiscuous substrate profile (50). These ancestral sequences also tend to possess enhanced stability, possibly due to the high-temperature environment of ancient times, especially in sequences reconstructed from the Pre-Cambrian era (36).

We reconstructed ancient SrtA proteins and tested them against our established assay which allowed us to explore how natural sequence variation of SrtA enzymes related to the overall promiscuity and activity of this important class of enzymes. We predicted that the ancestral SrtA enzymes will display improved substrate promiscuity for the target motif as they may retain ancestral generalist traits allowing for recognition of these target motifs (51).

Results and Discussion

2.2 Ancestral Constructs, Anc_{Staph} and Anc_{Strep}

By use of ASR we initially obtained two ancestral sortase sequences, Anc_{strep} and Anc_{staph}. These sequences correspond to the ancestral forms of the WT SrtA_{strep} and SrtA_{staph} enzymes, selected based on the high statistical support for these nodes on the ancestral SrtA phylogenetic tree. When expanding the scope of this study via utilization of ancestrally reconstructed constructs, the identification and selection of constructs with high statistical support is vital, especially when attempting to discover viable, catalytically active constructs. In addition, the quality of the phylogenetic tree which informs these reconstructed sequences is of the utmost importance (48). Errors in alignment, reconstructing longer sequences than the true ancestors, tree topology, or errors in insertions/deletions of residues can dramatically alter the reliability of the ASR results (52). The techniques for reconstruction of these ancestral sequences are described.

The potential behavior of these initial constructs, Anc_{staph} and Anc_{strep} was uncertain due to limited supporting information regarding how SrtA may perform when subjected to ASR as currently there are no studies in which SrtA has been reconstructed by use of ASR to elucidate the characteristics of an improved SrtA enzyme. The ancestral Anc_{staph} was of particular interest as its extant relative, SrtA_{staph}, has a highly limited substrate scope where only the LPXTG motif is recognized. These ancestral prokaryotic enzymes, Anc_{staph} and Anc_{strep}, have experienced many mutations altering their overall functionality but have retained conserved key residues, and maintained roughly a 44% sequence identity, due to the need for these catalytically active residues to maintain enzymatic function. These conserved residues seem to be generally focused

around the SrtA active site and previous research has indicated that promiscuous enzymes tend to share the same catalytic active site (53).

By restoring these ancestrally reconstructed sequences we hope to at a minimum, improve the substrate promiscuity of SrtA with these new constructs, Anc_{staph} or Anc_{strep}, and potentially improve the resulting overall catalytic activity in conjunction with exploring the natural sequence variation of

these ancestral enzymes and their extant relatives, especially considering the low sequence identity between the WT SrtA_{staph} and SrtA_{strep}.

To determine the sequences of these ancestral constructs, a member of our lab, Jordan Valgardson, applied multiple statistical modeling systems to ancestrally reconstruct SrtA sequences. First, non-redundant sortase sequences were sourced from NCBI protein database. Cluster Database at High Intensity with Tolerance

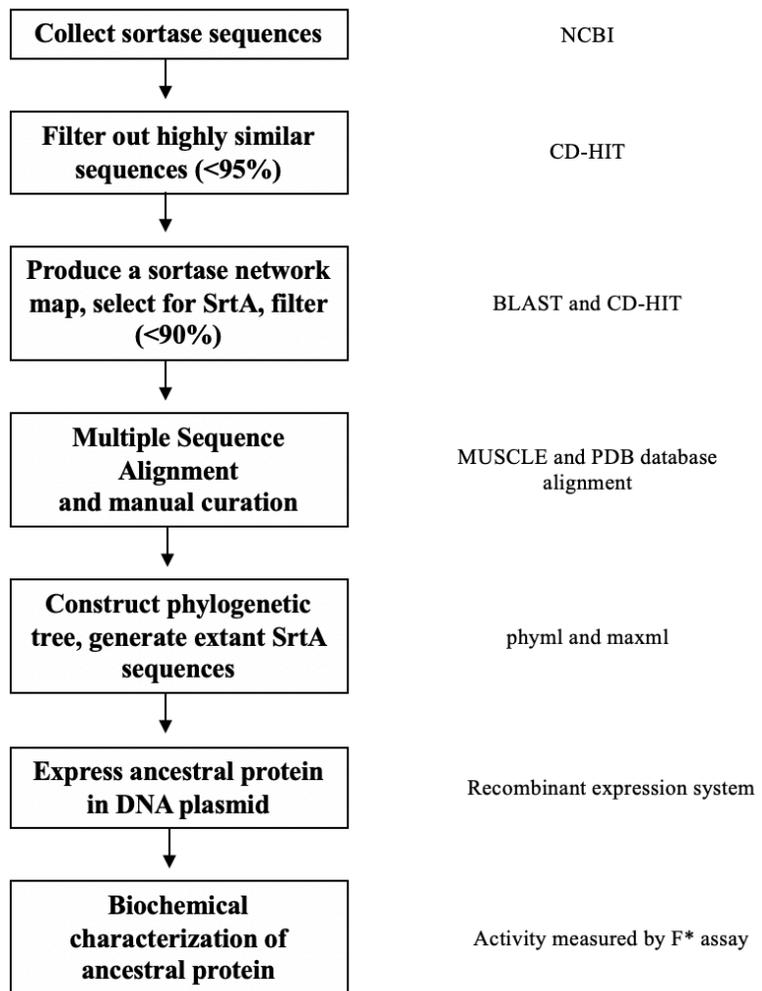


Figure 2-1. Steps for Ancestral Sequence Reconstruction (ASR). Allowed for the production of the ancestral protein sequences tested via our F* assay.

program (CD-HIT) was used to filter out highly similar (>95%) identical sequences sourced

from NCBI. An All-vs-all basic local alignment search tool (BLAST) was used on the remaining sortase sequences, producing a sortase network which informed the assignment of sortase class groups (A-F) by using labeled sortase sequences to assign a class to each grouping. Proteins surrounding the class A group were selected. An additional round of filtering was performed, and all highly similar proteins (>90%) were filtered out via CD-HIT. The remaining pool of sortase sequences was then subjected to alignment by Multiple Sequence Comparison by Log-Expectation (MUSCLE), and then manually curated to remove any outlying sequences. SrtA structures sourced from the PDB database were structurally aligned and sequence similarity between structural sequences (via PDB) and sortase sequences from the multi sequence alignment (MSA) (via ASR) then informed the true alignment of the MSA. A phylogenetic tree was constructed from the MSA via phylml and ancestral sequences were then generated at each node via multi-channel access XML (maxml) (Figure 2-1). The nodes preceding the SrtA_{strep} and SrtA_{staph} branches with high statistical support, designated Anc_{staph} and Anc_{strep}, were selected, and the sequences were cloned into DNA plasmids for further study (Figure 2-2). This same process informed the selection of other ancestral SrtA sequences used for this study.

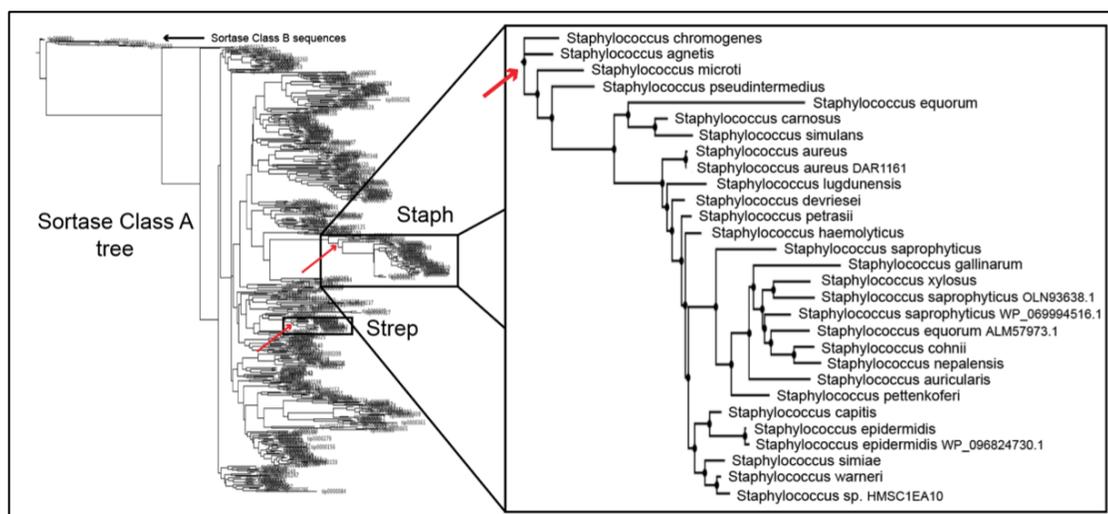


Figure 2-2. Phylogenetic tree displaying evolutionary branch points of sortase A. Red arrows indicate ancestrally reconstructed nodes for SrtA_{staph} and SrtA_{strep}. (Unpublished work, Jordan Valgardson)

2.3 Substrate Selectivity and Activity of Ancestral Constructs

Initial results for the ancestral constructs Anc_{staph} and Anc_{strep} showed that only the Anc_{strep} construct displayed improved catalytic activity when tested with the representative 5th position LPATGG, LPATSG, and LPATAG substrates. The most noticeable improvement in catalytic activity was observed for the LPATAG substrate wherein we observed a 4-fold improvement in activity compared to $SrtA_{staph}$ (Figure 2-3). Anc_{staph}

exhibited a sharp decrease in catalytic activity of roughly 50% compared to that of the WT $SrtA_{strep}$.

Regarding the substrate promiscuity of these two new constructs, neither the Anc_{staph} or the Anc_{strep} displayed a more promiscuous substrate selectivity profile when tested with the LPATGG, LPATSG, or LPATAG substrates. This is contrary to the hypothesis that ancestral enzymes tend to possess broader specificity, recognizing not only the canonic substrates but also additional substrate binding motifs, and their

extant relatives tend to be specialists, catalyzing specific reactions (53). But these substrate promiscuity results were limited to this representative substrate panel so to better understand how the substrate specificity of our ancestral constructs may be expanded by the utilization of ancestral constructs, both the Anc_{staph} and Anc_{strep} were tested against a 19 amino acid panel for a 5th position substitution (excluding Trp due to issues with peptide purification) to determine if they possessed improved substrate profiles outside of the initial substrates tested.

Abz-LPATXG-K(Dnp):

	<i>S. pneumoniae</i>	<i>S. aureus</i>	Anc_{strep}	Anc_{staph}
G	35	110	77	41
S	31	1	57	1
A	24	0	73	1

Figure 2-3. Heat map of initial ancestrally reconstructed $SrtA$ enzymes. Displays measured catalytic activity of WT and ancestral $SrtA$ enzymes with a 5th substitutions (LPATXG). Each “hit” corresponds to final percent conversion from starting material to product measured via florescent plate reader assay after 2 hrs. Darker shades of red indicate an enhanced overall percent conversion.

Similar to the substrate panel developed for the engineered ‘loop swapped’ constructs in Chapter 1, our ancestral constructs were tested against 19 amino acid substitutions in the 5th position LPATXG motif. Results indicated the Anc_{staph} exhibited a substrate promiscuity profile similar to that of WT

SrtA_{staph} except reduced catalytic activity was observed for the LPATGG substrate. Anc_{strep} displayed a slightly enhanced substrate selectivity profile in which catalytic activity was observed for the LPATCG and LPATNG substrates, in addition to the LPATGG, LPATSG, and LPATAG

substrates (Figure 2-4). This catalytic activity was measured with a 20%

conversion cut off. This improvement in promiscuity is consistent with literature that has explored ASR and enzymatic function wherein ancestral proteins were capable of recognizing a multitude of substrates compared to their extant relatives. The improved substrate promiscuity can possibly be linked to conformational changes of the substrate interacting loops as the active sites of enzymes such as SrtA tend to be highly conserved (35, 46). Our construct, Anc_{strep} not

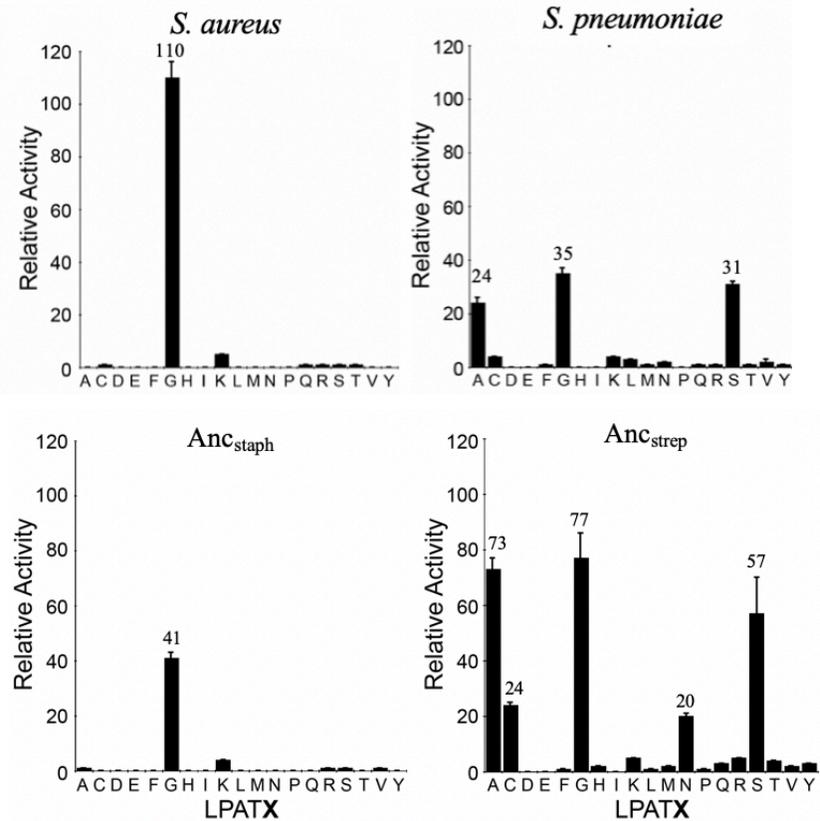


Figure 2-4. Expanded graphical representation of ancestrally reconstructed SrtA enzymes. Displays measured catalytic activity of WT and ancestral SrtA enzymes with a 5th substitutions (LPATXG). Final percent conversion from starting material to product measured via florescent plate reader assay after 2 hrs. Percent conversions over 20% labeled above bar. * indicates residue substitution was not determined.

only followed this trend, but in addition displayed higher catalytic activity for these newly recognized substrates.

A BLAST sequence comparison indicated twenty mutations between Anc_{staph} and SrtA_{staph} and fifty-three mutations between Anc_{strep} and SrtA_{strep} (Figure 2-5). This result is intriguing as the Anc_{strep} displayed an improved substrate specificity though it has over two times the number of mutations of the Anc_{staph}, which has fewer mutations but its catalytic activity has been halved. This difference in activity and promiscuity between the Anc_{staph} construct and its extant relative, WT SrtA_{staph}, is most likely due to the majority of mutations occurring in the β 6- β 7 loop region (Figure 2-6, 2-7). The β 6- β 7 loop has been indicated as playing a role in substrate motif recognition and promiscuity as well as making up a part of the binding groove (38). These mutations in WT SrtA_{staph}, K162N, T165D, G167E, K175E, D176K, and K177N could be impacting the dynamic movement of the β 6- β 7 loop required for substrate binding or eliminating necessary contact points required for substrate recognition and processing. One of these contact points, the Gly residue in the WT *S. aureus* seems to be interacting with the Pro residue of the LPXTG substrate. When this Gly is mutated to a Glu, a necessary interaction for catalysis may not be able to occur (Figure 2-6). Regarding the movement of the β 6- β 7 loop, during substrate binding the β 6- β 7 loop experiences repetitive folding and unfolding of the short helical stretches, and once the substrate is bound β 6- β 7 loop will then adopt a final conformation which accommodates the bound substrate (27, 30). These mutations could be altering the conformation of this region so that processing of the substrates cannot occur.

	Score	Expect	Method	Identities	Positives	Gaps
	260 bits(665)	5e-96	Compositional matrix adjust.	127/147(86%)	135/147(91%)	1/147(0%)
SrtA_{ancStaph}:	Query 3	QKPPEIPKDKSKMAGYISVPDADIKEPVYPGPATPEQLNRGVSF AEEDSLDDQNISIAG				62
	Sbjct 1	Q P+IPKDKSK+AGYI +PDADIKEPVYPGPATPEQLNRGVSF AE+ESLDDQNISIAG				60
20 mutations versus	Query 63	HTFTDRPHYQFTNLKAAKKGSKVYFKVGNETRKYKMTSIRDVNPDDVEVLDEQ-GEKNQL				121
<i>S. aureus</i>	Sbjct 61	HTF DRP+YQFTNLKAAKKGSKVYFKVGNETRKYKMTSIRDV P DV VLDEQ G+ QL				120
	Query 122	TLITCDDYNEQTGVWEKRKIFVAEQVK	148			
	Sbjct 121	TLITCDDYNE+TGVWEKRKIFVA +VK	147			
	Sbjct 121	TLITCDDYNEKTGVWEKRKIFVATEVK	147			

	Score	Expect	Method	Identities	Positives	Gaps
	214 bits(544)	1e-76	Compositional matrix adjust.	111/164(68%)	132/164(80%)	1/164(0%)
SrtA_{ancStrep}:	Query 20	LVAQAQSNLFPVIGGIAIPELGINLPIFKVGNNTSLLYGAGTMKEDQVMGEGNYALASHH				79
	Sbjct 3	L +Q + LPVIGGIAIPEL +NLPFKG+ N +L YGAGTMK +QVMGEGNY+LASHH				62
53 mutations versus	Query 80	IFGV-TASDMLFSPLERAKNGMKIYLTDKDNVYTYTITSVEVTPPERVDIDDTEGKKEI				138
<i>S. pneumoniae</i>	Sbjct 63	IFGV A+ MLFSP+ AKNGMKIYLTDK+ VYTY I V+ VTP+RVD +DD +G EI				122
	Query 139	TLVTCDDYEAQR IIVKGELEETTPYNEASEDILNAPNQSYNQF	182			
	Sbjct 123	TLVTC D AT+RIIVKG+L+ET Y++ S++IL AFNQ Y QF	166			
	Sbjct 123	TLVTCEDLAATER IIVKGD LKTKDYSQTSDEILTAFNQPKQF	166			

Figure 2-5. BLAST sequence alignment of Anc_{staph} and Anc_{strep}. The number of mutations between the ancestrally reconstructed SrtA enzymes and their WT SrtA mates are identified.

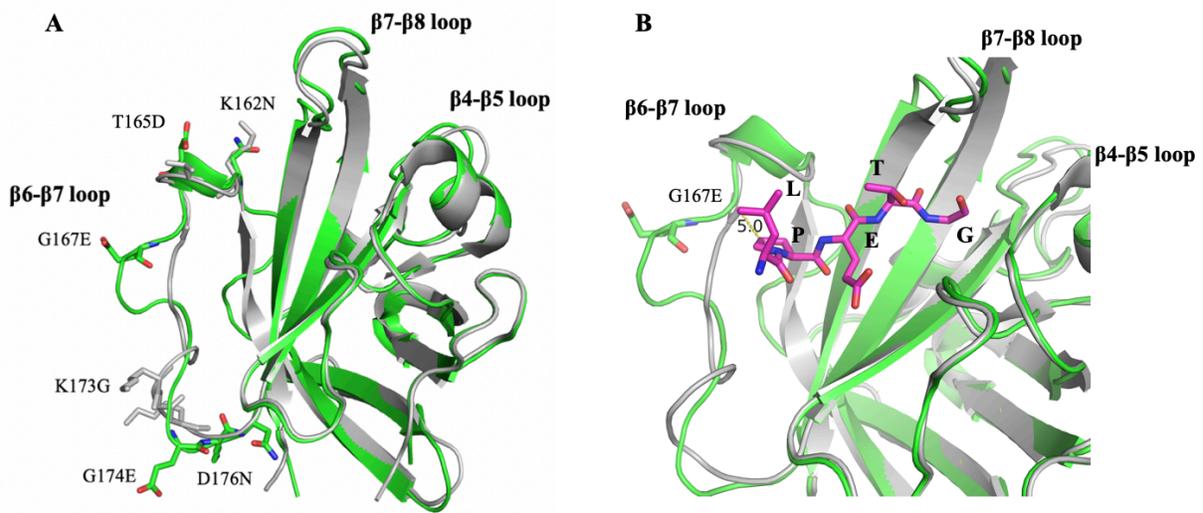


Figure 2-6. SWISS model of Anc_{staph} and WT SrtA from *S. aureus*. Modeled using 1T2W as template. WT SrtA from *S. aureus* colored grey, Anc_{staph} colored green. (PDB: 1T2W). **(A)** Mutations along the β 6- β 7 loop. **(B)** Bound LPETG peptide and G167E mutation interacting with Pro residue of peptide.

The Anc_{strep} enzyme has 53 mutations as compared to wildtype SrtA_{strep}, and these mutations resulted in a roughly 2-fold increase in catalytic activity and a slightly enhanced substrate promiscuity. By modeling this construct via SWISS-MODEL we can elucidate the key structural differences that may be producing this improvement in catalytic activity (Figure 2-8) (54). Similar to the Anc_{staph}, most of the mutations are the β 6- β 7 loop (Figure 2-7). But, we also observe mutations in the β 7- β 8 and the β 4- β 5 regions. As with the ‘loop swapped’ constructs we saw that mutations in the β 7- β 8 loop can modulate substrate promiscuity and catalytic activity.

We see a similar mutation in the Anc_{strep} as we saw with the SPS_{suis}, where a Thr residue directly follows the catalytic Cys which may have

SrtA Enzyme	β7-β8 Loop Sequence
<i>Streptococcus pneumoniae</i>	EDLAATE
<i>Streptococcus suis</i>	TDYYATQ
<i>Anc_{strep}</i>	TDYEATQ

Figure 2-7. β 7- β 8 loop sequences of Ancstrep and WT SrtA homologues.

resulted in an improvement in catalytic activity. Comparing the WT *S. pneumoniae* to the Anc_{strep}, we know that a Glu residue next to the catalytic Cys resulted in a decrease in catalytic activity as we saw with the SPS_{faecG145E} mutant. Perhaps this mutation from a Glu to Thr is causing the extra boost in activity and promiscuity for this ancestral construct. Another mutation, the E138T mutation in the Anc_{strep} β 7- β 8 loop seems to result in an interaction between the β 7- β 8 and β 4- β 5 loops. We are uncertain if this mutation is resulting in the modulation of activity and promiscuity of the Anc_{strep} but this possible interaction led us to consider that interactions between the β 7- β 8 and the β 4- β 5 loop may be a cause of previously unseen catalytic results, supported by PCA which illustrated that not only did the β 7- β 8 region show variability but the β 4- β 5 region showed variability as well.

Based on our demonstrated ability to reconstruct these ancestral SrtA sequences we speculated that we may reconstruct even more ancestral sequences in the phylogenetic tree to

expand our initial investigation into how the sequence variation and the space between these constructs may be affecting target motif recognition and also address the possible β 7- β 8 and β 4- β 5 loop interactions and how they could modulate activity in our SrtA enzymes.

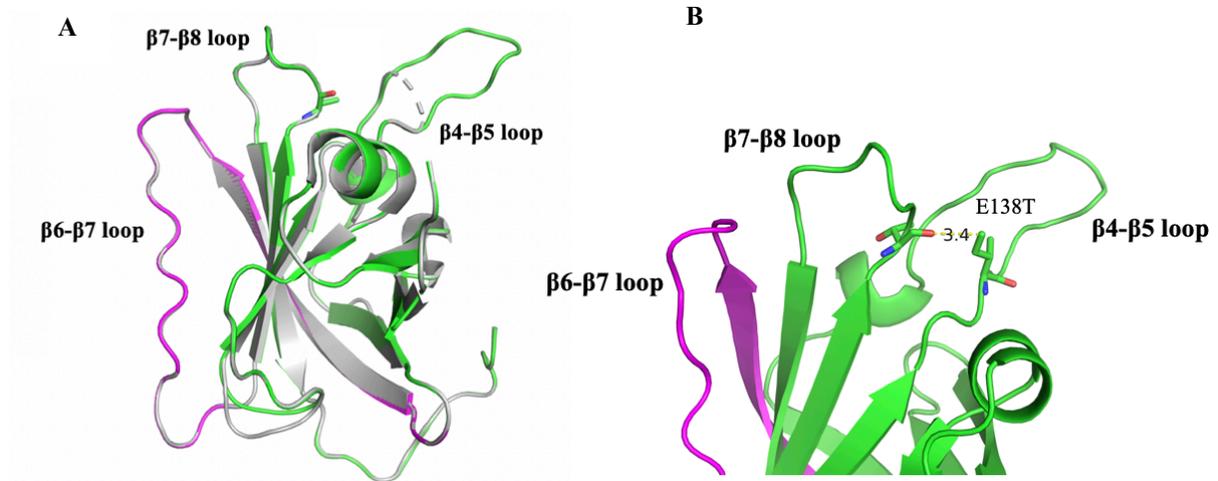


Figure 2-8. SWISS model of Ancstrep and WT SrtA from *S. pneumoniae*. Modeled using 3RCC as template. WT SrtA colored grey, Anc_{strep} colored green, areas of mutated residues in β 6- β 7 loop colored magenta (PDB: 3RCC). **(A)** Mutation in β 7- β 8 loop and mutations in β 6- β 7 loop (magenta). **(B)** Zoomed view of the E138T interaction in the Anc_{strep} with the Ile residue in the β 4- β 5 loop to a distance of 3.4 angstroms.

2.4 Expansion into More Ancestral Relatives

Promising results observed from our initial constructs Anc_{staph} and Anc_{strep}, encouraged further investigation into how these even more ancestral sequences would behave. The sequences of these ancestral SrtA relatives were obtained in a similar fashion to our initial constructs, Anc_{staph} and Anc_{strep}, in which nodes with high statistical support were selected for further testing by our kinetic fluorescence assay. Three new SrtA constructs were tested, termed corresponding to which node was selected, Anc₄₀₈, Anc₅₀₃, and Anc₅₄₇ (Figure 2-9). Node 408 is the most ancestral, corresponding to the branch between the Staph/Strep families and other bacterial

families, node 503 is the branch between Staph and Strep families, and node 547 is the branch between Strep and other families.

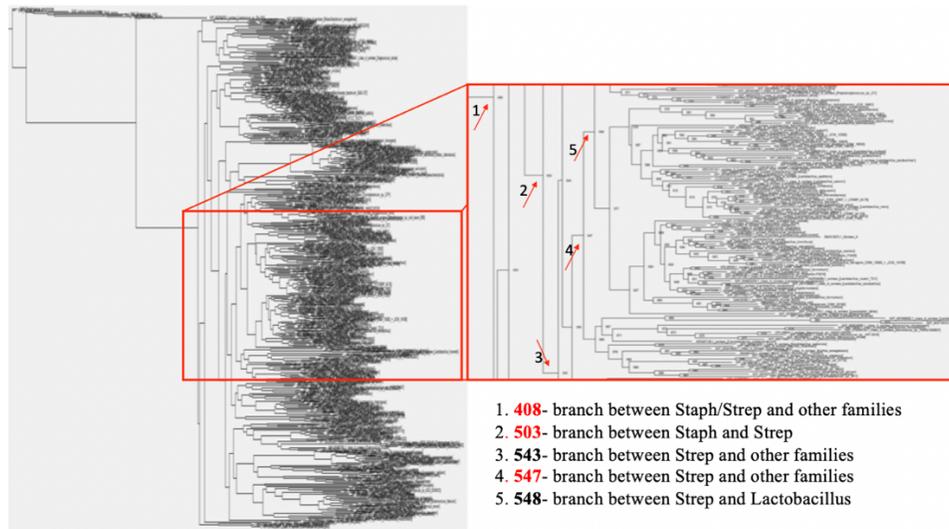


Figure 2-9. Phylogenetic tree of ancestral SrtA sequences. Multiple nodes are displayed (#1-5), further rounds of manual filtering revealed three nodes and their corresponding sequences to be tested, 408, 503, and 547, highlighted in red.

Though these nodes did have higher confidence values compared to other nodes, we acknowledge the limitations of ASR in reconstructing these ancestral sequences compared to the previous Anc_{staph} and Anc_{strep} nodes, though we aimed to select ancestral nodes with high confidence scores, there is a potential for sequence bias and error when utilizing a MSA in order to reconstruct these sequences (52).

By reconstructing the ancestral SrtA sequences of nodes further back on the phylogenetic tree we hoped to explore the natural sequence variation of class A SrtA enzymes and investigate how this variation could enhance or alter substrate recognition motifs. Reconstruction of ancestral constructs serves to fill in the sequence space of class A sortase enzymes, these reconstructed ancestral sequences “fill in” the space between these enzyme sequences, including

extant relatives, offering researchers the opportunity to predict functional sites and in addition, by adding ancestral sequences to a native MSA can improve the detection of new class A SrtA homologues. We hoped to observe an improvement in substrate promiscuity as previous literature indicates that ancestral proteins tend to be more thermally stable and act as generalists, recognizing a broader variety of substrates (35, 46, 51).

Results from the kinetic enzyme assay revealed no recognition of any of the 5th position substrate motifs wherein we expected to potentially observe an improvement in substrate

promiscuity (Figure 2-10). The complete

lack of all three enzymes ability to

process any of the substrates indicates

that key contact points and residues

necessary for activity may be absent in

these reconstructed enzymes, similar to

what we saw in the Anc_{staph} construct. A

comparison of the loop sequence identity

reveals that all three of the enzymes

possess shorter β 7- β 8 loops and have

almost no similarity in identity to the

WT or initially constructed ancestral

constructs except for a conserved Asp

near the N-terminus of the β 7- β 8 loop.

The presence of this Asp was also

indicated in our four active SrtA

Abz-LPATXG-K(Dnp):

	<i>S. pneumoniae</i>	<i>S. aureus</i>	Anc _{strep}	Anc _{staph}	Anc408	Anc503	Anc547
G	35	110	77	41	1	0	0
S	31	1	57	1	2	1	1
A	24	0	73	1	1	1	1
T	1	1	4	0	1	1	1
C	2	1	24	0	0	0	0
V	2	0	2	1	1	1	1
L	3	0	1	0	0	--	--
I	0	0	0	0	0	1	0
M	1	0	2	0	0	--	--
P	0	0	1	0	1	1	1
F	2	0	1	0	0	--	--
Y	1	0	3	0	0	--	--
D	0	0	0	0	1	3	3
E	0	0	0	0	1	1	1
N	4	0	20	0	1	1	1
Q	1	1	3	0	1	--	--
H	1	0	2	0	1	1	1
K	4	5	5	4	5	6	6
R	2	1	5	1	1	1	1

Figure 2-10. Heat map of more ancestral SrtA enzymes.

Displays measured catalytic activity of WT and ancestral SrtA enzymes with a 5th position substitutions (LPATXG). Each “hit” corresponds to final percent conversion from starting material to product measured via florescent plate reader assay after 2 hrs. Darker shades of red indicate an enhanced overall percent conversion.

homologues, SPS_{faec}, SPS_{suis}, SPS_{oralis}, and SPS_{lactis} (Figure 2-11, 2-12). This Asp plays an unknown role in catalysis but could be a potential target for mutation.

A			
	408	503	547
SrtA <i>strep</i>	46.8%	46.0%	52.6%
Anc <i>strep</i>	51.0%	55.9%	58.7%
SrtA <i>staph</i>	35.6%	32.5%	33.6%
Anc <i>staph</i>	35.4%	33.3%	33.9%

B		
SrtA Enzyme	β7-β8 Loop Sequence	β4-β5 Loop Sequence
<i>Staphylococcus aureus</i>	DDYEKTGVWEK	TFIDRPNYQ
<i>Streptococcus pneumoniae</i>	EDLAATE	HIFGVDNANKML
<i>Ancstaph</i>	DDYNEQTGVWEK	HTFTDRPHYQ
<i>Ancstrep</i>	TDYEATQ	HIFGVTASDML
<i>Anc408</i>	DTDTDQ	HMREDLL
<i>Anc503</i>	DNTDK	HMREDLL
<i>Anc547</i>	DDDGTN	HMNNPNLL

Figure 2-11. Sequence comparison of ancestral SrtA enzymes and newly constructed nodes. (A) BLAST sequence comparison, value corresponds to percent sequence identity between constructs. (B) β 7- β 8 loop sequences from WT and ancestral sequences.

There are also numerous mutations in the β 6- β 7 region of all of these ancestral enzymes. As described previously, the β 6- β 7 loop plays a role in substrate recognition, and before substrate binding is in a flexible, disordered state but upon substrate binding becomes ordered (22, 27). The mutation of not only this β 6- β 7 loop region as well as the β 7- β 8 loops in all three of the ancestral constructs may be deleteriously impacting the enzymes' ability to recognize and process the substrates.

SrtA Enzyme	β7-β8 Loop Sequence	β4-β5 Loop Sequence
<i>Staphylococcus aureus</i>	DDYEKTGVWEK	TFIDRPNYQ
<i>Streptococcus pneumoniae</i>	EDLAATE	HIFGVDNANKML
<i>Enterococcus faecalis</i>	GDLQATT	HRTEDGVSL
<i>Streptococcus suis</i>	TDYYATQ	HIFGVTGAADV L
<i>Streptococcus oralis</i>	VDYNATE	HIFTAENASQML
<i>Lactococcus lactis</i>	ADAEATH	HNMTGFTSDLSIL
<i>Bacillus anthracis</i>	VSVKDNSK	HNMSKKGVL
<i>Listeria monocytogenes</i>	DKPTETTK	HMRDESML

Figure 2-12. β 7- β 8 loop and β 4- β 5 loop sequences of WT SrtA enzymes and sortase homologues.

Regarding the potential interactions with the β 4- β 5 loop as previously discussed with the Anc_{strep} construct, we observe a pattern in loop length discrepancies between the β 4- β 5 loop and the β 7- β 8 loop in our ancestral constructs, exemplified by the SWISS homology model, which displays a significantly longer β 4- β 5 loop for the Anc_{strep} compared to the other three constructs (Figure 2-11, 2-13). The β 4- β 5 loop boundaries are defined as an N-terminal His and a C-terminal Phe. Though this loop segment does not encompass the entire β 4- β 5 loop length the conserved C-terminal Phe marks the end of the loop area that may be interacting with the β 7- β 8 loop. A combination of a short β 7- β 8 loop and a long β 4- β 5 loops seems to correlate to higher promiscuity and catalytic activity, and we see this pattern with our WT *S. pneumoniae* and the Anc_{strep} enzymes (Figure 2-11, 2-13).

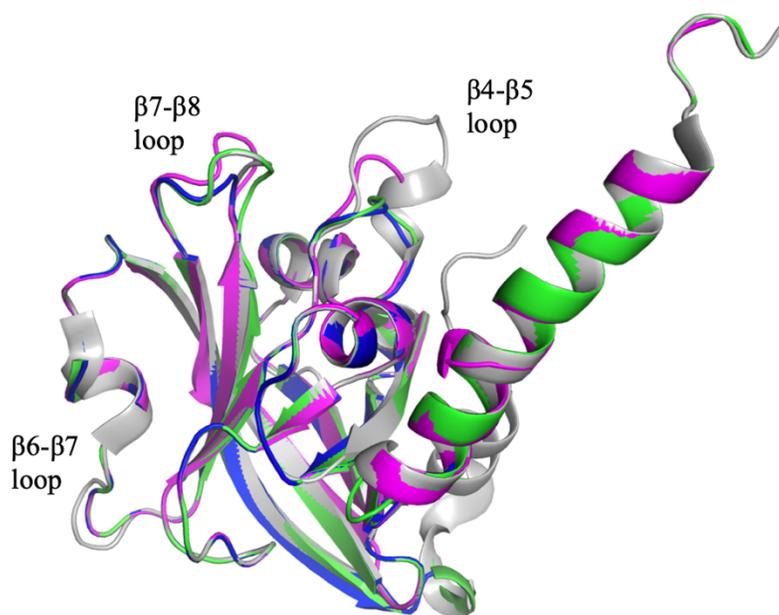


Figure 2-13. SWISS model of Anc408, Anc503, and Anc547 and WT *S. agalactiae*. Modeled using 3RCC as template. WT *S. agalactiae* colored grey, Anc408, green, Anc503, blue, Anc547, magenta.

2.5 β 7- β 8 and β 4- β 5 Loops and Enzymatic Behavior

The behavior of these substrate interacting loops and the interactions between them seems to be the key in modulating the behavior of engineered sortase A constructs. The results from the alignment of these ancestral sequences and the homology models led us back to our original ‘loop swapped’ constructs. Applying the knowledge we have gained from the exploration of the class A sortase sequence space we looked at the loop length variations and potential interactions between the β 7- β 8 and β 4- β 5 loops of the sortase homologues.

Using the WT sortase homologues sequences and the promiscuity data from the 2018 study out of the Antos lab we were unable to observe any overarching trends in loop length discrepancies between the β 7- β 8 and β 4- β 5 loops and substrate promiscuity using this MS data, contrary to the results obtained from the ancestral homology models (Figure 1-1). Though, the

WT sortase homologues that were more promiscuous possessed a hydrophobic residue directly following the catalytic His residue in the β 4- β 5 loop (Figure 2-12). This hydrophobic residue could be necessary for enhanced substrate promiscuity and catalysis. From this MS data we see that the WT *E. faecalis* was highly selective and only recognized the LPATAG substrate, but when engineered onto the *S. pneumoniae* core (SPS_{faec}) this enzyme recognized not only the LPATAG substrate but many other substrates when tested with our kinetic fluorescence assay. Thus, this improvement in activity seen for the SPS_{faec} may be not only due to the identity of the β 7- β 8 loop and the presence of a Gly residue near the catalytic Cys but the also the identity of the β 4- β 5 loop and the presence of a hydrophobic residue near the catalytic His. We were unable to explore site specific mutations in the β 4- β 5 loop regions for this study so our hypothesis regarding specific residue interactions is speculative, but our exploration of the sequence space and the data we have obtained exemplifies an exciting new avenue for exploration. Future studies could explore swapping out β 4- β 5 loop regions, specifically the region near the β 7- β 8 loop, as there seems to be conserved residues near the N and C terminus (His and Phe respectively) and variability between these residues. In addition, site specific mutations of the residues near the catalytic His could reveal loop interactions that may be modulating activity. These substrate interacting loops seem to be more intricately related then we previously thought and the real key to the difference in catalytic behavior and promiscuity.

2.6 Crystallization of AncS_{staph}

Previous published crystallization efforts of SrtA have successfully characterized the WT SrtA_{staph} and the dimer swapped SrtA_{strep} as well as other SrtA constructs but crystallization efforts have not been attempted for any ancestrally reconstructed SrtA constructs. The ability to

successfully crystallize either Anc_{staph} or Anc_{strep} could reveal key structural components and residues that resulted in improved substrate promiscuity and catalytic efficiency, especially with the Anc_{strep} construct where improved activity and substrate promiscuity was observed. Though crystallization attempts were made for both the apo-Anc_{strep} and apo-Anc_{staph} constructs and the apo-Anc₄₀₈, the only crystals that were successfully grown were those from the apo-Anc_{staph} enzyme. The results from our kinetic fluorescence assay showed that Anc_{staph} exhibited reduced catalytic efficiency and unremarkable substrate promiscuity, therefore limiting the conclusions we may make regarding the differing structural elements between WT SrtA_{staph} and Anc_{staph}, and their application to future engineered SrtA constructs.

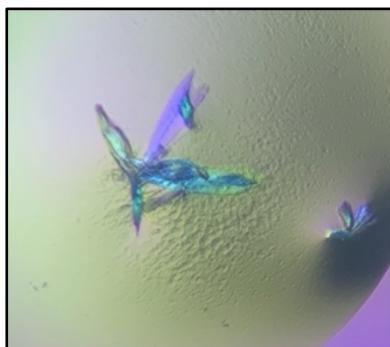


Figure 2-14. Crystallization of Anc_{staph}. Crystallized at 6.76 mg/mL in 0.2 M sodium thiocyanate, 6.9 pH. Crystallized as needle shaped crystal form with nucleation.

Initial crystallization for the Anc_{staph} construct was performed using a commercially available PEG ion 2 screen, 0.2 M sodium thiocyanate, 20% PEG3350, pH 6.9. Crystal trays were set up using the ‘hanging drop’ method at 20 °C. The drop contained 2 μL of well solution and 2 μL of 6.76 mg/mL SrtA enzyme. The N-terminal His tagged protein crystallized as a needle shaped crystal form with noticeable nucleation that had to be broken apart prior to looping (Figure 2-14). Crystals were grown to their maximum size after five months. These crystals

diffracted when analyzed at the synchrotron source to 3 angstroms. Though, the results for this construct were not in line with our research goals as the activity of this construct was unremarkable and there are multiple SrtA_{staph} structures published on the PDB database. Further studies of SrtA proteins should explore a crystallization of Anc_{strep} which could reveal more impactful results due to its overall catalytic behavior.

2.7 Ancestral Sortase and Future Directions

By means of ASR we were able to reconstruct a SrtA enzyme, *Anc_{strep}*, that displayed a slightly improved substrate promiscuity profile, recognizing the LPATCG and LPATNG peptides as well as our standard peptide panel of LPATGG, LPATSG, and LPATAG. Though the catalytic activity was only improved roughly 2-fold this enzyme still offers an option for researchers wanting to process a wider variety of substrates for the purposes of SML. The inactivity of the *Anc_{staph}* illustrated that perturbations in the β 6- β 7 loop region may result in deleterious effects on substrate promiscuity and catalysis.

When analyzing our more ancestral constructs the inactivity observed for all three of our ancestral constructs (*Anc₄₀₈*, *Anc₅₀₃*, *Anc₅₄₇*) could be related to not just the differences in the loop length and loop identity of β 7- β 8 loops but also the β 4- β 5 loops and the interactions between them. We identified potential residue interactions that may be causing this inactivity in not only our ancestral constructs but their extant relatives, specifically the presence of a hydrophobic residue near the catalytic His in the β 4- β 5 loop, and future work will investigate site specific mutations to test this hypothesis. A future study should explore if and/or where the interactions between these substrate interacting loops is occurring as we think that these loops and the interactions between them seem to be the key in modulating the behavior of engineered sortase A constructs.

Materials and Methods

Materials and methods for the ancestral sortase enzymes are the same as those detailed in ‘Chapter 1: ‘Loop Swapped’ Engineered Sortase A. Refer to this chapter for information regarding protein purification, peptide purification, fluorescence plate reader assays, and instrumentation information. Tricine gel images showing purified samples of all of our ancestral proteins is pictured (Figure 2-15), molecular weights listed in Table 3 in Appendix 1. Crystallization methods are detailed in Section 2.6 *Crystallization of Anc_{staph}*.

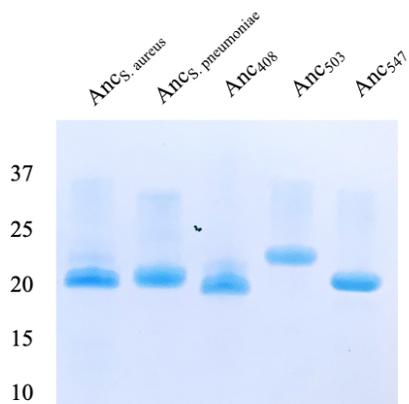


Figure 2-15. Gel image of SrtA enzymes. Tricine gel showing purified SrtA protein samples of ancestral sortase enzymes.

References

1. Suree, N., Liew, C. K., Villareal, V. A., Thieu, W., Fadeev, E. A., Clemens, J. J., Jung, M. E., and Clubb, R. T. (2009) The structure of the *Staphylococcus aureus* sortase-substrate complex reveals how the universally conserved LPXTG sorting signal is recognized. *J. Biol. Chem.* **284**, 24465–24477
2. Liew, C. K., Smith, B. T., Pilpa, R., Suree, N., Ilangovan, U., Connolly, K. M., Jung, M. E., and Clubb, R. T. (2004) Localization and mutagenesis of the sorting signal binding site on sortase A from *Staphylococcus aureus*. *FEBS Lett.* **571**, 221–226
3. Marraffini, L. A., DeDent, A. C., and Schneewind, O. (2006) Sortases and the Art of Anchoring Proteins to the Envelopes of Gram-Positive Bacteria. *Microbiol. Mol. Biol. Rev.* **70**, 192–221
4. Ton-That, H., Mazmanian, S. K., Alksne, L., and Schneewind, O. (2002) Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. Cysteine 184 and histidine 120 of sortase form a thiolate-imidazolium ion pair for catalysis. *J. Biol. Chem.* **277**, 7447–7452
5. Antos, J. M., Truttmann, M. C., and Ploegh, H. L. (2016) Recent advances in sortase-catalyzed ligation methodology. *Curr. Opin. Struct. Biol.* **38**, 111–118
6. CDC: hospitals “need to do more” to control MRSA. AHRQ report disheartening. (2008) *Hosp. Peer Rev.* **33**, 5–6, 11
7. Clancy, K. W., Melvin, J. A., and McCafferty, D. G. (2010) Sortase transpeptidases: insights into mechanism, substrate specificity, and inhibition. *Biopolymers* **94**, 385–396
8. Spirig, T., Weiner, E. M., and Clubb, R. T. (2011) Sortase enzymes in Gram-positive bacteria. *Mol. Microbiol.* **82**, 1044–1059
9. Khare, B., Krishnan, V., Rajashankar, K. R., I-Hsiu, H., Xin, M., Ton-That, H., and Narayana, S. V. (2011) Structural differences between the *Streptococcus agalactiae* housekeeping and pilus-specific sortases: SrtA and SrtC1. *PLoS One* **6**, e22995
10. Jacobitz, A. W., Wereszczynski, J., Yi, S. W., Amer, B. R., Huang, G. L., Nguyen, A. V., Sawaya, M. R., Jung, M. E., McCammon, J. A., and Clubb, R. T. (2014) Structural and computational studies of the *Staphylococcus aureus* sortase B-substrate complex reveal a substrate-stabilized oxyanion hole. *J. Biol. Chem.* **289**, 8891–8902
11. Dorr, B. M., Ham, H. O., An, C., Chaikof, E. L., and Liu, D. R. (2014) Reprogramming the specificity of sortase enzymes. *Proc. Natl. Acad. Sci. USA* **111**, 13343–13348
12. Beerli, R. R., Hell, T., Merkel, A. S., and Grawunder, U. (2015) Sortase Enzyme-Mediated Generation of Site-Specifically Conjugated Antibody Drug Conjugates with High In Vitro and In Vivo Potency. *PLoS One* **10**, e0131177
13. Elleuche, S., and Pöggeler, S. (2010) Inteins, valuable genetic elements in molecular biology and biotechnology. *Appl. Microbiol. Biotechnol.* **87**, 479–489
14. Chen, I., Dorr, B. M., and Liu, D. R. (2011) A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci. USA* **108**, 11399–11404
15. Hirakawa, H., Ishikawa, S., and Nagamune, T. (2012) Design of Ca²⁺-independent *Staphylococcus aureus* sortase A mutants. *Biotechnol. Bioeng.* **109**, 2955–2961
16. Jeong, H.-J., Abhiraman, G. C., Story, C. M., Ingram, J. R., and Dougan, S. K. (2017) Generation of Ca²⁺-independent sortase A mutants with enhanced activity for protein and cell surface labeling. *PLoS One* **12**, e0189068

17. Schmohl, L., Bierlmeier, J., Gerth, F., Freund, C., and Schwarzer, D. (2017) Engineering sortase A by screening a second-generation library using phage display. *J Pept Sci* **23**, 631–635
18. Cobb, R. E., Chao, R., and Zhao, H. (2013) Directed Evolution: Past, Present and Future. *AIChE J.* **59**, 1432–1440
19. Tobin, P. H., Richards, D. H., Callender, R. A., and Wilson, C. J. (2014) Protein engineering: a new frontier for biological therapeutics. *Curr Drug Metab* **15**, 743–756
20. Dods, R. L., and Donnelly, D. (2015) The peptide agonist-binding site of the glucagon-like peptide-1 (GLP-1) receptor based on site-directed mutagenesis and knowledge-based modelling. *Biosci. Rep.* **36**, e00285
21. Ilangovan, U., Ton-That, H., Iwahara, J., Schneewind, O., and Clubb, R. T. (2001) Structure of sortase, the transpeptidase that anchors proteins to the cell wall of *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. USA* **98**, 6056–6061
22. Bentley, M. L., Gaweska, H., Kielec, J. M., and McCafferty, D. G. (2007) Engineering the substrate specificity of *Staphylococcus aureus* Sortase A. The beta6/beta7 loop from SrtB confers NPQTN recognition to SrtA. *J. Biol. Chem.* **282**, 6571–6581
23. Marraffini, L. A., Ton-That, H., Zong, Y., Narayana, S. V. L., and Schneewind, O. (2004) Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. A conserved arginine residue is required for efficient catalysis of sortase A. *J. Biol. Chem.* **279**, 37763–37770
24. Weiner, E. M., Robson, S., Marohn, M., and Clubb, R. T. (2010) The Sortase A enzyme that attaches proteins to the cell wall of *Bacillus anthracis* contains an unusual active site architecture. *J. Biol. Chem.* **285**, 23433–23443
25. Frankel, B. A., Tong, Y., Bentley, M. L., Fitzgerald, M. C., and McCafferty, D. G. (2007) Mutational analysis of active site residues in the *Staphylococcus aureus* transpeptidase SrtA. *Biochemistry* **46**, 7269–7278
26. Jacobitz, A. W., Kattke, M. D., Wereszczynski, J., and Clubb, R. T. (2017) Sortase transpeptidases: structural biology and catalytic mechanism. *Adv. Protein Chem. Struct. Biol.* **109**, 223–264
27. Naik, M. T., Suree, N., Ilangovan, U., Liew, C. K., Thieu, W., Campbell, D. O., Clemens, J. J., Jung, M. E., and Clubb, R. T. (2006) *Staphylococcus aureus* Sortase A transpeptidase. Calcium promotes sorting signal binding by altering the mobility and structure of an active site loop. *J. Biol. Chem.* **281**, 1817–1826
28. Pang, X., and Zhou, H.-X. (2015) Disorder-to-Order Transition of an Active-Site Loop Mediates the Allosteric Activation of Sortase A. *Biophys. J.* **109**, 1706–1715
29. Kappel, K., Wereszczynski, J., Clubb, R. T., and McCammon, J. A. (2012) The binding mechanism, multiple binding modes, and allosteric regulation of *Staphylococcus aureus* Sortase A probed by molecular dynamics simulations. *Protein Sci.* **21**, 1858–1871
30. Ugur, I., Schatte, M., Marion, A., Glaser, M., Boenitz-Dulat, M., and Antes, I. (2018) Ca²⁺ binding induced sequential allosteric activation of sortase A: An example for ion-triggered conformational selection. *PLoS One* **13**, e0205057
31. Nikghalb, K. D., Horvath, N. M., Prelesnik, J. L., Banks, O. G. B., Filipov, P. A., Row, R. D., Roark, T. J., and Antos, J. M. (2018) Expanding the Scope of Sortase-Mediated Ligations by Using Sortase Homologues. *Chembiochem* **19**, 185–195
32. Piotukh, K., Geltinger, B., Heinrich, N., Gerth, F., Beyermann, M., Freund, C., and Schwarzer, D. (2011) Directed evolution of sortase A mutants with altered substrate

- selectivity profiles. *J. Am. Chem. Soc.* **133**, 17536–17539
33. Wójcik, M., Szala, K., van Merkerk, R., Quax, W. J., and Boersma, Y. L. (2020) Engineering the specificity of *Streptococcus pyogenes* sortase A by loop grafting. *Proteins*
 34. Gumulya, Y., and Gillam, E. M. J. (2017) Exploring the past and the future of protein evolution with ancestral sequence reconstruction: the “retro” approach to protein engineering. *Biochem. J.* **474**, 1–19
 35. Merkl, R., and Sterner, R. (2016) Reconstruction of ancestral enzymes. *Perspectives in Science* **9**, 17–23
 36. Risso, V. A., Sanchez-Ruiz, J. M., and Ozkan, S. B. (2018) Biotechnological and protein-engineering implications of ancestral protein resurrection. *Curr. Opin. Struct. Biol.* **51**, 106–115
 37. Jolliffe, I. T., and Cadima, J. (2016) Principal component analysis: a review and recent developments. *Philos. Trans. A, Math. Phys. Eng. Sci.* **374**, 20150202
 38. Bentley, M. L., Lamb, E. C., and McCafferty, D. G. (2008) Mutagenesis studies of substrate recognition and catalysis in the sortase A transpeptidase from *Staphylococcus aureus*. *J. Biol. Chem.* **283**, 14762–14771
 39. Harrison, R. A., and Engen, J. R. (2016) Conformational insight into multi-protein signaling assemblies by hydrogen-deuterium exchange mass spectrometry. *Curr. Opin. Struct. Biol.* **41**, 187–193
 40. Chalmers, M. J., Busby, S. A., Pascal, B. D., West, G. M., and Griffin, P. R. (2011) Differential hydrogen/deuterium exchange mass spectrometry analysis of protein-ligand interactions. *Expert Rev Proteomics* **8**, 43–59
 41. Englander, S. W., and Mayne, L. (2014) The nature of protein folding pathways. *Proc. Natl. Acad. Sci. USA* **111**, 15873–15880
 42. Masson, G. R., Burke, J. E., Ahn, N. G., Anand, G. S., Borchers, C., Brier, S., Bou-Assaf, G. M., Engen, J. R., Englander, S. W., Faber, J., Garlish, R., Griffin, P. R., Gross, M. L., Guttman, M., Hamuro, Y., Heck, A. J. R., Houde, D., Iacob, R. E., Jørgensen, T. J. D., Kaltashov, I. A., Klinman, J. P., Konermann, L., Man, P., Mayne, L., Pascal, B. D., Reichmann, D., Skehel, M., Snijder, J., Strutzenberg, T. S., Underbakke, E. S., Wagner, C., Wales, T. E., Walters, B. T., Weis, D. D., Wilson, D. J., Wintrode, P. L., Zhang, Z., Zheng, J., Schriemer, D. C., and Rand, K. D. (2019) Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nat. Methods* **16**, 595–602
 43. Konermann, L., Pan, J., and Liu, Y.-H. (2011) Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.* **40**, 1224–1234
 44. Harms, M. J., and Thornton, J. W. (2010) Analyzing protein structure and function using ancestral gene reconstruction. *Curr. Opin. Struct. Biol.* **20**, 360–366
 45. Yang, Z., Kumar, S., and Nei, M. (1995) A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* **141**, 1641–1650
 46. Gardner, J. M., Biler, M., Risso, V. A., Sanchez-Ruiz, J. M., and Kamerlin, S. C. L. (2020) Manipulating conformational dynamics to repurpose ancient proteins for modern catalytic functions. *ACS Catal.*
 47. Thornton, J. W. (2004) Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.* **5**, 366–375
 48. Cai, W., Pei, J., and Grishin, N. V. (2004) Reconstruction of ancestral protein sequences and its applications. *BMC Evol. Biol.* **4**, 33

49. Hochberg, G. K. A., and Thornton, J. W. (2017) Reconstructing ancient proteins to understand the causes of structure and function. *Annu. Rev. Biophys.* **46**, 247–269
50. Voordeckers, K., Brown, C. A., Vanneste, K., van der Zande, E., Voet, A., Maere, S., and Verstrepen, K. J. (2012) Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. *PLoS Biol.* **10**, e1001446
51. Siddiq, M. A., Hochberg, G. K., and Thornton, J. W. (2017) Evolution of protein specificity: insights from ancestral protein reconstruction. *Curr. Opin. Struct. Biol.* **47**, 113–122
52. Vialle, R. A., Tamuri, A. U., and Goldman, N. (2018) Alignment modulates ancestral sequence reconstruction accuracy. *Mol. Biol. Evol.* **35**, 1783–1797
53. Khersonsky, O., and Tawfik, D. S. (2010) Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* **79**, 471–505
54. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., and Schwede, T. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303

Appendix A

Table 1. Standard deviation values from 5th position kinetic enzyme assays. Dashed lines indicate that only one trial was performed.

Abz-LPATXG-K(Dnp):

	<i>S. pneumoniae</i>	<i>S. aureus</i>	SASPneumoniae	SPSAureus	SPSLactis	SPSAnthracis	SPSFaecalis	SPSOrahis	SPSSuis	SPSMonocytogenes	SPSFaecalisG145E	SASAEK1TG	SASSuis	SASPneumoniaeT194W	SPSAureusW194T	Ancstaph	Ancstrep	Anc408	Anc503	Anc547
G	2	6	0	1	12	1	6	13	11	2	0	0	0	5	2	9	0	0	0	0
S	1	0	0	0	4	0	4	4	8	1	3	0	0	1	0	13	0	0	0	0
A	2	0	0	2	7	0	10	7	12	0	4	0	0	1	0	4	0	0	0	0
T	0	0	0	0	1	0	2	0	0	1	--	--	--	0	0	0	0	0	--	--
C	0	0	0	0	1	1	2	2	1	0	--	--	--	0	0	1	0	--	--	--
V	1	0	0	0	4	0	3	1	1	0	--	--	--	1	1	0	0	0	--	--
L	0	0	0	1	0	--	2	0	1	0	--	--	--	--	0	0	--	--	--	--
I	0	0	0	0	1	0	1	0	0	0	--	--	--	0	0	0	0	0	--	--
M	0	0	0	0	--	--	--	--	--	--	--	--	--	--	0	0	--	--	--	--
P	0	0	0	0	0	0	0	0	0	0	--	--	--	0	0	0	0	0	--	--
F	0	0	0	1	6	--	2	3	2	1	--	--	--	--	0	0	--	--	--	--
Y	0	0	0	0	--	--	--	--	--	--	--	--	--	--	0	0	--	--	--	--
D	0	0	0	0	1	1	1	1	1	1	--	--	--	1	1	0	0	1	--	--
E	0	0	0	0	0	0	0	0	0	0	--	--	--	0	0	0	0	0	--	--
N	0	0	0	0	2	0	2	1	1	0	--	--	--	1	0	1	0	--	--	--
Q	0	0	0	0	1	--	1	0	1	0	--	--	--	--	0	0	--	--	--	--
H	0	0	0	4	0	0	0	0	1	0	--	--	--	0	0	0	0	0	--	--
K	0	0	0	0	--	--	--	--	--	--	--	--	--	--	0	0	1	--	--	--
R	0	0	0	0	1	0	1	0	1	0	--	--	--	0	0	0	0	0	--	--

Table 2. Standard deviation values from 4th position kinetic enzyme assays. Dashed lines indicate that only one trial was performed.

Abz-LPAXGG-K(Dnp):

	<i>S. pneumoniae</i>	<i>S. aureus</i>	SASPneumoniae	SPSAureus	SPSLactis	SPSAnthracis	SPSFaecalis	SPSOrahis	SPSSuis	SPSMonocytogenes	SPSFaecalisG145E	SASAEK1TG	SASSuis	SASPneumoniaeT194W	SPSAureusW194T
A	2	3	0	10	8	0	7	3	7	1	0	0	0	0	5
E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0

Table 3. Molecular Weights and Extinction Coefficients of Sortase A Enzymes

Identity	Molecular Weight	Ext. Coefficient
SrtA _{Staph}	18760.13	0.770
SrtA _{Strep}	20144.73	0.865
SAS _{Pneumoniae}	20235.65	0.442
SPS _{Aureus}	23788.64	1.089
Anc _{Staph}	20954.16	0.760
Anc _{Strep}	22447.98	0.597
Anc ₄₀₈	20183.88	0.443
Anc ₅₀₃	18256.63	0.408
Anc ₅₄₇	17768.22	0.419
SAS _{PneumoniaeT194W}	20320.76	0.711
SPS _{AureusW194T}	23912.74	0.853
SPS _{Suis}	21139.78	1.035
SPS _{Oralis}	21089.72	0.967
SPS _{Monocytogenes}	21197.91	0.892
SPS _{Faecalis}	20983.64	0.901
SPS _{Lactis}	20992.61	0.901
SPS _{Anthraxis}	21154.89	0.894
SAS _{Suis}	18041.32	0.661
SA _{ΔEKTG}	18475.88	0.782

Table 4. Sortase A Enzyme Sequences**>WT Staph_aureus_D59-sortase_A (SrtA_{Staph})**

MGSSHHHHHHSSGLVPRGSHMQAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLN
 RGVSF AEENESLDDQNISIAGHTFIDRPNYQFTNLKAAKKGSMVYFKVGNETRYKMTSI
 RDVKPTDVGVLDEQK GKDKQLTLITCDDYNEKTGVWEKRKIFVATEVK

>WT Strep_pneumoniae_D80-sortase_A (SrtA_{Strep})

MESSHHHHHHAVLTSQWDAQKLPVIGGIAIPELEMNLPFKGLDNNVLFY GAGTMK
 REQVMGEGNYSLASHHIFGVDNANKMLFSPLDNAKNGMKIYLTDKNKVYTYEIREVKRVT
 PDRVDEVDDR DGVNEITLVTCEDLAATERIIVKGD LKETKDYSQTSDEILTAFNQPYKQFY

>SrtA_Streptococcus_pneumoniae-Swap-Aureus (SPS_{aureus})

MGSSHHHHHHSSGLVPRGSHMESSHHHHHHENLYFQSAVLTSQWDAQKLPVIGGIAIPELEMN
 LPIFKGLDNNVLFY GAGTMKREQVMGEGNYSLASHHIFGVDNANKMLFSPLDNAKNGMKIYLTD
 KNKVYTYEIREVKRVT PDRVDEVDDR DGVNEITLVT CDDYNEKTGVWEKR IIVKGD LKETKDY
 SQTSDEILTAFNQPYKQFY

>SrtA_staphylococcus_aureus-Swap-Pneumoniae (SAS_{pneumoniae})

MGSSHHHHHHSSGLVPRGSHMESSHHHHHHENLYFQSAKQIPKDKSKVAGYIEIPDADIKEPV
 YPGPATPEQLNRGVSF AEENESLDDQNISIAGHTFIDRPNYQFTNLKAAKKGSMVYFKVGNETRY
 YKMTSIRDVKPTDVGVLDEQK GKDKQLTLITCEDLAATERKIFVATEVK

>SrtA_staph_aureus-Swap-Pneumoniae_T194W (SAS_{pneumoniaeT194W})

MESSHHHHHHSSGLVPRGSHMESSHHHHHENLYFQSQAKPQIPKDKSKVAGYIEIPDADIKEPV
YPGPATPEQLNRGVSF AEENESLDDQNISIAGHTFIDRPNYQFTNLKAAKKGSMVYFKVGNETRK
YKMTSIRDVKPTDVGVLDEQK GKDKQLTLITCEDLAAWERKIFVATEVK

>SrtA_Strep_pneumoniae-Swap-Aureus_W194T (SPS_{aureusW194T})

MESSHHHHHHSSGLVPRGSHMESSHHHHHENLYFQSAVLTSQWDAQKLPVIGGIAIPELEMNLP
FKGLDNVNLFYGAGTMKREQVMGEGNYSLASHHIFGVDNANKMLFSPLDNAKNGMKIYLT
DKNKVYTYEIREVKRVT PDRVDEVDDRDRDGVNEITLVTCDDYNEKTGVTEKRIIVKGD
LKETKDYSQTSDEILTA FNQPYKQFY

>SrtA_Strep_pneumoniae-Swap-suis (SPS_{suis})

MESSHHHHHHHENLYFQSAVLTSQWDAQKLPVIGGIAIPELEMNLP
PIFKGLDNVNLFYGAGTMKREQVMGEGNYSLASHHIFGVDNANKMLFSPLDNAKNGMKIY
LTDKNKVYTYEIREVKRVT PDRVDEVDDRDRDGVNEITLVTCDDYATQRIIVKGD
LKETKDYSQTSDEILTA FNQPYKQFY

>SrtA_Strep_pneumoniae-Swap-oralis (SPS_{oralis})

MESSHHHHHHHENLYFQSAVLTSQWDAQKLPVIGGIAIPELEMNLP
PIFKGLDNVNLFYGAGTMKREQVMGEGNYSLASHHIFGVDNANKMLFSPLDNAKNGMKIY
LTDKNKVYTYEIREVKRVT PDRVDEVDDRDRDGVNEITLVTCDDYATQRIIVKGD
LKETKDYSQTSDEILTA FNQPYKQFY

>SrtA_Strep_pneumoniae-Swap-monocytogenes (SPS_{monocytogenes})

MESSHHHHHHHENLYFQSAVLTSQWDAQKLPVIGGIAIPELEMNLP
PIFKGLDNVNLFYGAGTMKREQVMGEGNYSLASHHIFGVDNANKMLFSPLDNAKNGMKIY
LTDKNKVYTYEIREVKRVT PDRVDEVDDRDRDGVNEITLVTCDDYATQRIIVKGD
LKETKDYSQTSDEILTA FNQPYKQFY

>SrtA_Strep_pneumoniae-Swap_faecalis (SPS_{faecalis})

MESSHHHHHHHENLYFQSAVLTSQWDAQKLPVIGGIAIPELEMNLP
PIFKGLDNVNLFYGAGTMKREQVMGEGNYSLASHHIFGVDNANKMLFSPLDNAKNGMKIY
LTDKNKVYTYEIREVKRVT PDRVDEVDDRDRDGVNEITLVTCDDYATQRIIVKGD
LKETKDYSQTSDEILTA FNQPYKQFY

>SrtA_Strep_pneumoniae-Swap_lactis (SPS_{lactis})

MESSHHHHHHHENLYFQSAVLTSQWDAQKLPVIGGIAIPELEMNLP
PIFKGLDNVNLFYGAGTMKREQVMGEGNYSLASHHIFGVDNANKMLFSPLDNAKNGMKIY
LTDKNKVYTYEIREVKRVT PDRVDEVDDRDRDGVNEITLVTCDDYATQRIIVKGD
LKETKDYSQTSDEILTA FNQPYKQFY

>SrtA_Strep_pneumoniae-Swap_anthraxis (SPS_{anthracis})

MESSHHHHHHHENLYFQSAVLTSQWDAQKLPVIGGIAIPELEMNLP
PIFKGLDNVNLFYGAGTMKREQVMGEGNYSLASHHIFGVDNANKMLFSPLDNAKNGMKIY
LTDKNKVYTYEIREVKRVT PDRVDEVDDRDRDGVNEITLVTCDDYATQRIIVKGD
LKETKDYSQTSDEILTA FNQPYKQFY

>SrtA_Strep_pneumoniae-Swap_faecalisG145E (SPS_{faecalisG145E})

MESSHHHHHHHENLYFQSAVLTSQWDAQKLPVIGGIAIPELEMNLP
PIFKGLDNVNLFYGAGTMKREQVMGEGNYSLASHHIFGVDNANKMLFSPLDNAKNGMKIY
LTDKNKVYTYEIREVKRVT PDRVDEVDDRDRDGVNEITLVTCDDYATQRIIVKGD
LKETKDYSQTSDEILTA FNQPYKQFY

>SrtA_Staph_aureus-Swap-suis (SAS_{suis})

MGSSHHHHHHSSGLVPRGSHMQAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLN
RGVSFAEENESLDDQNISIAGHTFIDRPNYQFTNLKAAKKGSMVYFKVGNETRKYKMTSI
RDVKPTDVGVLDEQK GKDKQLTLITCTDYATQRKIFVATEVK

>SrtA_Staph_aureus-ΔEKTG (SA_{ΔEKTG})

MGSSHHHHHSSGLVPRGSHMQAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLN
RGVSFAEENESLDDQNIAGHTFIDRPNYQFTNLKAAKKGSMVYFKVGNETRYKMTSI
RDVKPTDVGVLDEQKQKDKQLTLITCDDYDYNVWEKRKIFVATEVK

>ancSrcA_staph (Anc_{staph})

MESSHHHHHSSGLVPRGSHMESSHHHHHHENLYFQSQPPEIPKDKSKMAGYISVPDADIKEP
VYPGPATPEQLNRGVSFAEEDSLDDQNIAGHTFTDRPHYQFTNLKAAKKGSKVYFKVGNET
RYKMTSIRDVNPDDVEVLDEQGEKNQLTLITCDDYNEQTGVWEKRKIFVAEQVK

>ancSrtA_strep (Anc_{strep})

MESSHHHHHSSGLVPRGSHMESSHHHHHHENLYFQSI SLVAQAQSNLPVIGGIAIPELGINLPI
FKGVGNTSLLYGAGTMKEDQVMGEGNYALASHHIFGV TASDMLFSPLERAKNGMKIYLTDKD
NVYTYTITSVEVVTPERVDVIDDTEGKKEITLVCTDYEATQRIIVKGELEETTPYNEASEDILNAF
NQSYNQF

>ancSrtA_408 (Anc₄₀₈)

MESSHHHHHHENLYFQSIQPPSLSAKVDKSAIGQIAIPSVGLNLPFKGTTNENLLAGAGTMSPDQ
KMGEGNYVLAGHHMREDLLFGPLMKVKKGDKIYLTQNEVYTYKVTETKVVHETDTSVLDDT
GEPRLTLITCDTDTDQRFVVTAELVEKEPMKEESQEVKYQQKNQFILLLL

>ancSrtA_503 (Anc₅₀₃)

MESSHHHHHHENLYFQSEPPSLASAKMDKQVIGQIAIPSVNINLPILKGTNENLLAGAATMKPD
QKMGKGNVYVLAGHHMREDLLFSPLHNKVGDKIYLTDNKHVYTYKVTETKVVDPPTETDVLDD
TGEPQITLITCDNTDKRLVVTGELVETTPFEEEEQVK

>ancSrtA_547 (Anc₅₄₇)

MESSHHHHHHENLYFQSSLSLAKARMDDLHVIGAIIPSVNMNLPILKGVSNENLAVGAGTMKP
DQKMGKGNVYVLAGHHMNNPNLLFSPLHRVKKGDKIYLTDMKHVYTYKVTSTKVVDPTEVDVI
DDTGEPLITLITCDDDGTNRLIVQGELVETTPFDA

List of Abbreviations:

SrtA: Sortase A

SPS_X: Strep pneumoniae Swap X (X=Indicates any of the 6 new sortase homologues)

CWSS: Cell wall sorting signal

ASR : Ancestral Sequence Reconstruction

WT: Wild Type

Anc: Ancestral

Enzyme Abbreviations;

SrtA_{strep}: Wild type Sortase A from *Streptococcus pneumoniae*

SrtA_{staph}: Wild type Sortase A from *Staphylococcus aureus*

SPS_{Aureus}: SrtA *Streptococcus pneumoniae* swap *S. aureus* β7-β8 loop

SAS_{Pneumoniae}: SrtA *Staphylococcus aureus* swap *S. pneumoniae* β7-β8 loop

Anc_{Staph}: Ancestrally reconstructed *Staphylococcus aureus*
Anc_{Strep}: Ancestrally reconstructed *Streptococcus pneumoniae*
Anc₄₀₈: Ancestrally reconstructed SrtA at node 408
Anc₅₀₃: Ancestrally reconstructed SrtA at node 503
Anc₅₄₇: Ancestrally reconstructed SrtA at node 547
SAS_{PneumoniaeT194W}: SrtA *Staphylococcus aureus* swap *S. pneumoniae* β 7- β 8 loop, T194W mutation
SPS_{AureusW194T}: SrtA *Streptococcus pneumoniae* swap *S. aureus* β 7- β 8 loop, W194T mutation
SPS_{Suis}: SrtA *Streptococcus pneumoniae* swap *S. suis* β 7- β 8 loop
SPS_{Oralis}: SrtA *Streptococcus pneumoniae* swap *S. oralis* β 7- β 8 loop
SPS_{Mono}: SrtA *Streptococcus pneumoniae* swap *L. monocytogenes* β 7- β 8 loop
SPS_{Faec}: SrtA *Streptococcus pneumoniae* swap *E. faecalis* β 7- β 8 loop
SPS_{Lactis}: SrtA *Streptococcus pneumoniae* swap *L. lactis* β 7- β 8 loop
SPS_{Anth}: SrtA *Streptococcus pneumoniae* swap *B. anthracis* β 7- β 8 loop
SAS_{Suis}: SrtA *Staphylococcus aureus* swap *S. suis* β 7- β 8 loop
SA Δ EKTG: SrtA *Staphylococcus aureus* swap with truncated EKTG residues