Fall 2019

Isopeptide Ligations Catalyzed by Streptococcus Suis Sortase A

Sarah Bowersox
Western Washington University, sarahbowersox5@gmail.com

Isopeptide Ligations Catalyzed by Streptococcus Suis Sortase A

By

Sarah Bowersox

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

ADVISORY COMMITTEE

Dr. John Antos, Chair

Dr. Gerry Prody

Dr. Sergey Smirnov

GRADUATE SCHOOL

David L. Patrick, Interim 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 that 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 Bowersox
December, 2019

Isopeptide Ligations Catalyzed by Streptococcus Suis Sortase A

A Thesis Presented to The Faculty of Western Washington University

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

By
Sarah Bowersox
December, 2019

Abstract

Chemically modified proteins are critical components of modern therapeutics and basic research. To generate non-natural protein derivatives, bacterial sortase enzymes have been effective due to their ability to catalyze selective ligations between protein targets and functional groups that are uncommon in nature. Thus far, the enzymatic approach using sortase has been limited to modifications at the termini of peptide chains. Here we describe efforts to develop a sortase-mediated strategy for the formation of isopeptide bonds at the side chains of internal lysine residues. To this end, we have identified a sortase A homolog from *Streptococcus suis* ($SrtA_{suis}$) that is capable of generating isopeptide linkages at levels far superior to sortases that are typically used for protein modification. As confirmed by RP-HPLC, ESI-MS, and MS/MS analysis, we have succeeded in generating isopeptide linkages with model peptide substrates and larger biological targets by utilizing $SrtA_{suis}$. Additionally, we have optimized these reactions by using Ni(II) ions to sequester reaction by-products in order to drive the equilibrium toward the desired products. Overall, $SrtA_{suis}$ is uniquely effective at catalyzing isopeptide bond formation and shows promise for future applications such as generating protein oligomers linked together by isopeptide bond, and the formation of protein therapeutics modified at lysine residues.

Table of Contents

7. Appendices

List of Figures and Tables

List of Abbreviations and Acronyms

Abz	Aminobenzoyl
ADC	Antibody-drug conjugate
Ahx	Amino hexanoic acid
Boc	tert-Butyloxycarbonyl
^{Cd}SpaA	surface antigen protein A Corynebacterium diphtheriae
^{Cd}SrtA ^{3M}	Three mutant ^{Cd}SrtA
CPMV	Cowpea chlorotic mottle virus
CRISPR	clustered regularly interspaced short palindromic repeat
^{C}SpaA	C terminal SpaA
DCM	Dichloromethane
DEAC	7-(diethylamino)coumarin-3-carboxylic acid
DIPEA	Diisopropylethylamine
DMSO	Dimethylsulfoxide
Dnp	Dinitrophenyl
EDTA	Ethylenediaminetetraacetic acid
ESI-MS	Electrospray ionization mass spectrometry
Fmoc	Fluorenylmethyloxycarbonyl
FPLC	Fast protein liquid chromatography
HBTU	O-(Benzotriazol-1-yl)-N,N,N’,N’-tetramethyluronium hexafluorophospate
IgH	Immunoglobulin heavy chain
IgL	Immunoglobulin light chain

IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl-beta-D-1-thiogalactopyranoside
LB	Lysogeny broth
LC	Liquid chromatography
mAb	Monoclonal antibodies
MBHA	4-Methylbenzhydrylamine hydrochloride
MeCN	Acetonitrile
NHS	N-hydroxysuccinimide
Ni^{2+}	$NiSO_4$
Ni-NTA	Nickel nitriloacetic acid
NMP	N-methyl-2-pyrrolidinone
^{N}SpaA	N-terminal SpaA
OD600	Optical density at 600nm
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PEG	polyethylene glycol
RP-HPLC	Reverse phase high performance liquid chromatography
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromotography
SMIL	Sortase-mediated isopeptide ligation
SML	Sortase-mediate ligation
SPPS	Solid-phase peptide synthesis

SpaA	Surface protein antigen A
SrtA	Class A sortase
$SrtA_{oralis}$	*Streptococcus oralis* Sortase A
$SrtA_{staph}$	*Staphylococcus aureus* Sortase A
$SrtA_{suis}$	*Streptococcus suis* Sortase A
TFA	Trifluoroacetic acid
TIPS	Triisopropylsilane
Tris	Tris(hydroxymethyl)aminomethane
UV-vis	Ultraviolet/visible spectroscopy

1. Introduction

1.1 Overview of Protein Modification Techniques

Protein modification is a field devoted to altering natural protein structures in order to give them novel functions and properties applicable to problems that arise in academic, industrial, and medical settings.[1,2] Modified proteins may take a variety of forms; however, it is often the case that protein modification involves the creation of a hybrid structure containing a protein of interest and an exogenous moiety that are covalently linked together.[3] The result is a novel protein derivative that contains its own individual properties as well as the properties of the modification endowed upon it. In turn, these properties are useful in producing more efficient therapeutics,[4,5] protein-based materials,[6] strategies for enzyme immobilization,[7] increasing stability,[5] assembling proteins for elucidation of their structures,[3] and many other applications.

Protein modifications can be carried out in a variety of ways, each having their own advantages and limitations. A primary technique, common in molecular biology, modifies proteins at the genetic level by introducing point mutations or deletions and insertions of specific DNA sequences.[1,2,8] There are many biological methods available for gene modification, such as error-prone PCR, transfer-DNA or transposon insertion, as well as techniques utilizing site-specific nucleases such as zinc fingers, transcription activator-like effectors, and clustered regularly interspaced short palindromic repeat (CRISPR) associated protein 9.[8] These techniques have the advantage that they offer precise control over where alterations are made in the protein sequence. However, in the majority of cases these strategies are restricted to the incorporation of naturally occurring amino acids, and thus are limited in the types of modifications that may be

installed. Nonetheless, these strategies are extremely valuable and are routinely used in applications such as gene therapy for genetic disorders or cancer,[9,10] development of cell based therapies,[11] introgression of parasite-resistance genes,[12] and quality enhancement in crops.[13] It should also be noted that genetic code expansion provides a powerful means for adding unnatural amino acids and modifications.[14] However, this strategy continues to require extensive and sometimes expensive screening in order to optimize unnatural amino acid incorporation, and can also have deleterious effects on protein expression levels.[15]

As an alternative to genetic methods for protein modification, chemical protein modification provides an experimentally simple way to introduce useful modifications to proteins. Many amino acid side chains, such as cysteine, lysine, and glutamic/aspartic acid are of particular interest due to their natural ability to participate in reactions that allow for covalent modification. Choosing the right amino acid for protein modification is critical, and is based on its acidity/basicity, electrophilicity/nucleophilicity, oxido-reductive characteristics, or its location and accessibility.[3] Done carefully, the modification of native amino side chains allows a broad range of non-natural moieties to be added to an already folded protein structure. This includes but is not limited to chromophore addition, polyethylene glycol (PEG) addition to increase circulation half-life[16] and increase solubility, and the attachment of drugs for therapeutics.[17,18] Unfortunately, the site-selectivity obtained via chemical protein modification is often not ideal, and if the target protein contains multiple reactive residues then random conjugation might occur at sites that interfere with the protein's native function or stability.

To address the limitations of genetic and chemical methods for protein modification, a more recent approach is to use enzymes that possess innate protein modification activity. This

process is termed chemoenzymatic protein modification, and takes advantage of the inherent protein modification capabilities of certain enzymes and their ability to recognize specific amino acid sequences in protein substrates. When these recognition sequences are incorporated into protein targets via genetic methods, the enzyme is available to perform its modification role at a single, well-defined position. The result is a covalent linkage between protein targets and a modification of interest (**Figure 1**).[19] This technique offers the significant advantage of increased site-selectivity, and allows for easy prediction of the site of modification. It also maintains flexibility since the enzymes can often employ a variety of small molecule reagents tethered to the non-natural modification. The use of enzymes to site-specifically modify proteins also has the advantage that the reactions typically operate under mild aqueous conditions, which is important for preserving the structure and function of the biomolecule target.[20]

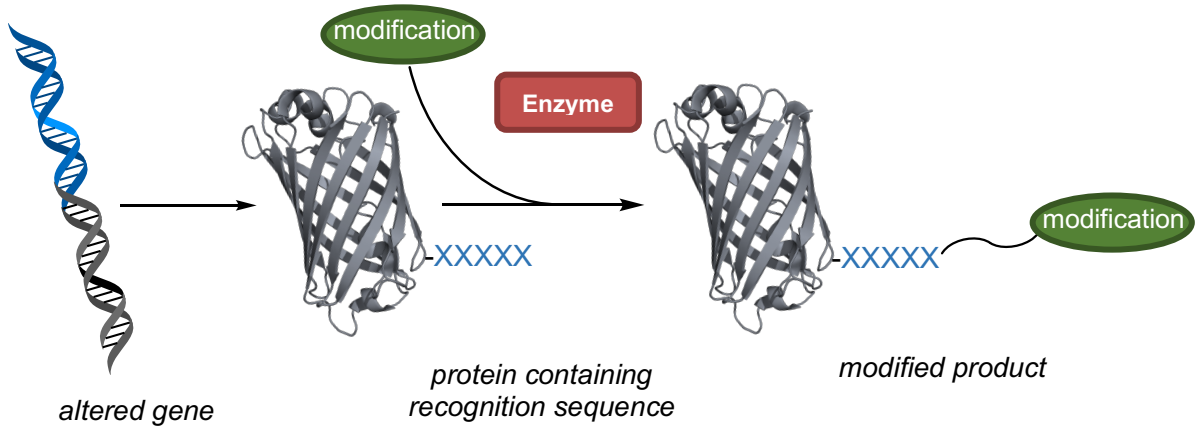


Figure 1. General representation of chemoenzymatic protein modification

Chemoenzymatic approaches combine many of the attractive features of genetic and chemical approaches for protein modification, and as a result these methods have seen increasing use in the literature.[21,22] In addition, the range of enzymes that are compatible with

chemoenzymatic approaches is continually expanding.[21] While a full review of all enzymes that have been utilized for protein modification is beyond the scope of this thesis, noteworthy examples include formyl glycine generating enzyme, which converts cysteine residues to aldehydes that can participate in further bioconjugation with aminooxy and hydrazide molecules, [19,23] transglutaminases that catalyze the ligation of glutamine with lysine side chains or other primary amines,[19,24] biotin and lipoic acid ligases that catalyze the attachment of their respective modification to lysine sidechains,[19] and bacterial sortases which are transpeptidases that are able to ligate together protein or peptide building blocks.[19]

1.2 Using Sortase A for Chemoenzymatic Protein Engineering

Sortase enzymes, first characterized by Schneewind and colleagues in 1999,[25] are found on the cell surface of all gram-positive bacteria and function to 'sort' (i.e. covalently anchor) proteins to the cell wall or polymerize protein subunits in the construction of pili.[25,26,27] These cell wall anchored proteins interact with extracellular matrices or specific host molecules and enable bacterial adherence to tissues, immune evasion, and target cell invasion. [28] Phylogenetic studies have classified sortases into six subclasses denoted A-F, and all sortases characterized to date are cysteine transpeptidases that aid in the joining of proteins containing a specific amino motif (termed a "sorting signal") with an amino group located on the cell surface. [26,27,29,30]

One of the most noteworthy sortase subclasses is Class A (SrtA), containing the archetypal SrtA from *Staphylococcus aureus* ($SrtA_{staph}$).[31] SrtA proteins are often called 'housekeeping' enzymes, due to their responsibility for sorting the majority of cell surface proteins in an organism and anchoring them to the cell wall.[24-27,31] Class B enzymes, found in some gram-

positive bacilli and cocci, are primarily involved in heme-iron acquisition as well as the construction of bacterial pili.[26,27,32,33] Also known for participating in pilus formation are the Class C enzymes, which are also unique in having the capability to form isopeptide bonds.[26,27,34,35] Class D enzymes are found primarily in bacilli and are an important component in anchoring cell wall proteins involved in bacterial sporulation. [26,36] Class E and F are the least characterized, however class E sortases found in organisms such as *Corynebacterium diphtheriae* seem to have 'housekeeping' activity similar to SrtA.[26,37]

Mechanistically for SrtA enzymes, the reaction involves a catalytic transacylation between proteins containing a LPXTG sorting motif (where X can be any amino acid) and the lipid II precursor of peptidoglycan.[26,27] First, the highly conserved active-site cysteine cleaves the peptide bond between threonine and glycine in the LPXTG motif. This results in the formation of an acyl-enzyme intermediate through a thioester bond between the active-site cysteine and the threonine residue, while the target molecule's remaining C-terminus is released (**Figure 2**). Second, the intermediate is resolved by nucleophilic attack from the reactive portion of lipid II, which is the nucleophilic amine terminus in the pentaglycine stretch of the interpeptide bridge.[38] The result is a new cell wall anchored protein linked to lipid II via a standard peptide bond.

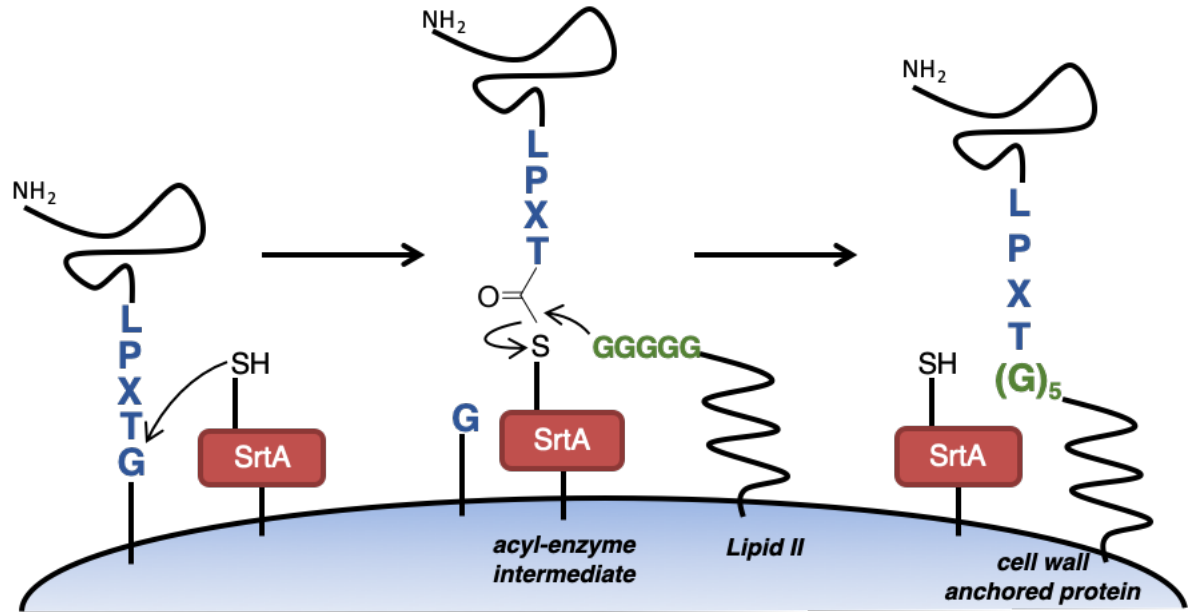


Figure 2. Cell wall anchoring catalyzed by sortase A (*S. aureus*) *in vivo*.

For Class C sortases, also known as pilin sortases, the transacylation reaction follows a similar mechanism that results in the covalent assembling of pili, or long protein fibers, from pilin subunits. For many Class C enzymes, pilin subunits consist of a C-terminal LPXTG motif and a conserved N-terminal lysine-containing motif (VYPKH).[35] Pilin polymerization involves the same intermediary acyl enzyme as observed for SrtA, however it is resolved through nucleophilic attack by the lysine side chain in the pilin motif of another subunit rather than by the amine terminus of a cell wall precursor (**Figure 3**).[35] After repeated transacylation reactions with other pilin subunits, the result is a cell wall anchored pilus.

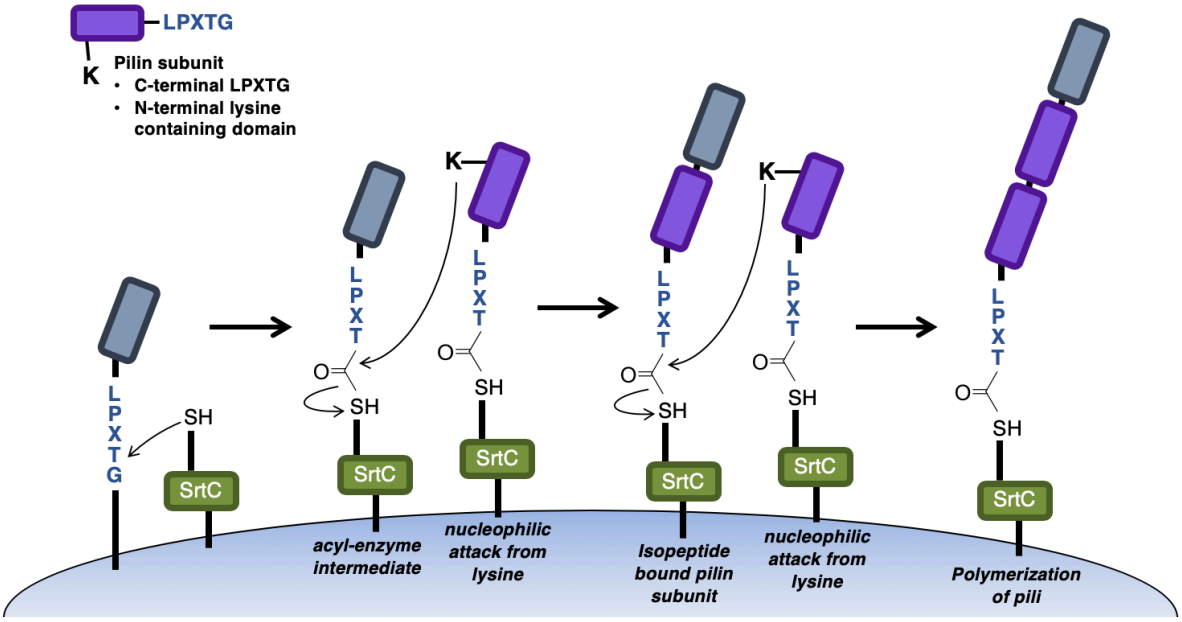


Figure 3. Pilin sortases (SrtC) mode of action for pilin polymerization.

Due to the display of various surface proteins that help in host cell entry, bacterial adhesion to host tissues, and immune evasion, sortases are considered lead virulence factors in pathogenic bacteria and have been the subject of a great deal of research as potential anti-infective targets.[38,27,37] Alternatively, sortases have emerged as useful tools in protein modification, specifically SrtA from *Staphylococcus aureus* ($SrtA_{staph}$).[39,40] Using $SrtA_{staph}$ *in vitro* has become a popular chemoenzymatic approach termed sortase-mediated ligation (SML). The transacylation reaction can be replicated efficiently using proteins or synthetic peptides containing the LPXTG motif and terminal glycine nucleophiles (**Figure 4**).[41] These different ligation partners can be live cells,[42,43,44] solid supports,[45] full sized proteins,[39,45,46,47,48,49] chromophores,[43,44,46,50,51] and a wide range of other modifications.

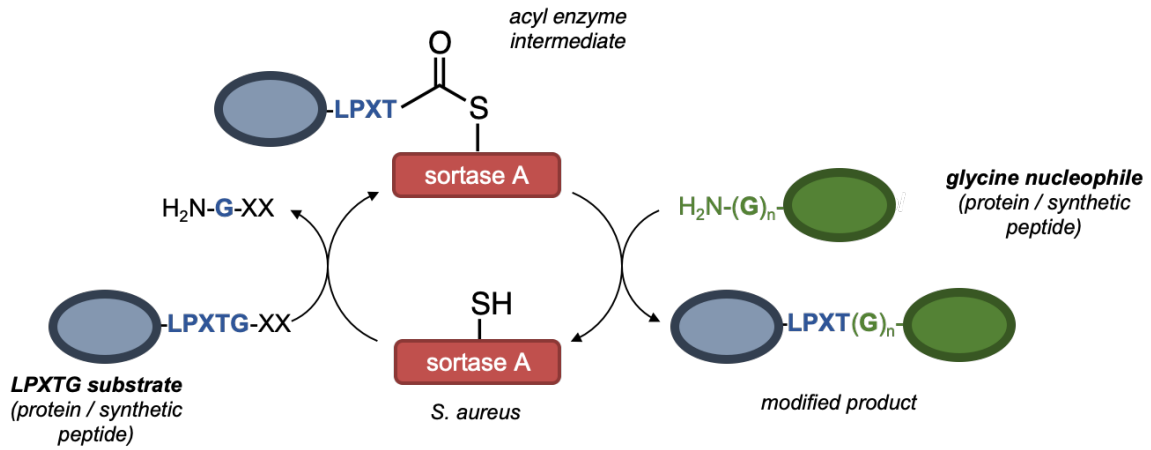


Figure 4. Sortase-Mediated Ligation (SML) *in vitro.*

While the range of SML applications is extensive, one notable example was shown by Beerli *et al.* in 2015, which involved the formation of antibody-drug conjugates (ADCs).[49] In these antibody-drug conjugates the immunoglobulin heavy chains (IgH) and immunoglobulin light chains (IgL) were modified at the C-terminus to contain an LPETG motif. Two potent cytotoxins commonly utilized in ADCs, monomethylauristatin E and maytansin, were then modified to contain the requisite pentaglycine peptide. Using $SrtA_{staph}$ the authors were able to successfully conjugate anti-CD30 and anti-Her-2 antibodies to these potent cytotoxic molecules with >80% conjugation efficiency. These ADCs were further tested for their ability to release payloads, conjugated via a peptide bond to the C-terminal sortase tag, upon internalization and lysosomal degradation inside a targeted cell. Potent effects for tumor killing *in vitro* as well as tumor regression in *in vivo* models were observed, suggesting successful drug release. Therefore, it was concluded that the resulting ADCs conjugated by sortase were found to display tumor killing activity *in vitro* and *in vivo* that was similar to chemically conjugated ADCs, despite lacking the traditional cleavable linker structures commonly employed in ADCs .[49]

In another example, the Park group used SML for labeling live HeLa cells that expressed $SrtA_{staph}$ on their surface.[44] By incubating cells with fluorescent proteins or fluorophores containing the sortase sorting motif (eGFP-$LPETG_5$ and TAMRA-$LPETG_5$), they were able to directly label the cells via acyl enzyme formation between the surface-displayed $SrtA_{staph}$ and the fluorescent label. Controls using non-compatible substrates (eGFP-$LPETA_5$ and TAMRA-$LPATA_5$) did not result in cell labeling, highlighting the exceptional specificity of $SrtA_{staph}$ for the LPXTG motif.[44]

As a final example of SML, Schoonen *et a.l* enzymatically modified the capsid of cowpea chlorotic mottle virus (CPMV), which has applications in the encapsulation of cargoes such as small drug molecules, fluorescent dyes for optical imaging, and Gd^{3+} ions for MRI.[52,53,54] In this case, the N-termini of the viral capsid proteins were located on the inside of the virus-like particle allowing for controlled encapsulation. To render the virus-like particle compatible with SML, the N-terminus of the capsid protein was genetically modified to display a triglycine motif. When combined with substrates containing the LPXTG motif, they were able to achieve up to 58% modification of the virus capsid proteins. In addition, the assembly of the virus particles did not appear to be perturbed as a result of sortase-mediated modification, suggesting that SML was a viable method for loading CPMV with cargoes.[52]

1.3 Current Techniques and Limitations of Sortase-Mediated Ligation (SML) Reactions

The SML examples highlighted in the preceding section demonstrate that the highly selective transacylation reaction catalyzed by $SrtA_{staph}$ can be used to modify a diverse range of proteins, making it a powerful tool for the generation of non-natural polypeptide derivatives.[20]

However, in parallel with the development of new applications that exploit SML, there have also been a number of studies aimed at enhancing the performance of $SrtA_{staph}$ and other sortases for protein modification. Historically, the enzyme typically used in SML is a truncated form (Δ59) of wild-type $SrtA_{staph}$. It is lacking the first 59 amino acid residues that contain a transmembrane domain which normally anchors the sortase to the cell surface. The truncated version allows for expression of the protein in a soluble form, which is critical for *in vitro* ligations. However, it suffers from limitations such as poor reaction rates and Ca^{2+} cofactor dependency. Mutations to improve substrate binding, increase reaction kinetics, or efforts to remove the dependence of a Ca^{2+} cofactor have been reported.[55,56,57] The results are the so-called SrtA pentamutant (SrtA5mut) and SrtA heptamutant (SrtA7mut) that both show an increase in k_{cat}/K_M relative to wild-type $SrtA_{staph}$.[56] The heptamutant is also noteworthy as it does not require Ca^{2+} for activity *in vitro*.

Other SML limitations exist, notably limited ligation efficiency due to reaction reversibility.[58,59] After cleavage of the threonine and glycine peptide bond, the target protein's remaining C-terminus, hereafter referred to as the 'excised fragment', is released. The newly formed N-terminal glycine residue can return to act as a nucleophile resulting in the formation of the original substrate. Having one of the reactants in excess will drive the reaction forward towards production of desired products, however if the preparation of one of these reactants is labor intensive or costly this can become problematic. Presented by Row *et al.*, one method to avoid this issue is to expand the LPXTG recognition sequence to LPXTGGH, resulting in a GGH excised fragment following substrate cleavage.[60] In the presence of Ni^{2+} ions, the GGH excised fragment coordinates with Ni^{2+} forming a metal-peptide complex anchored through the histidine

residue (**Figure 5**). The nucleophilic lone pair of the N-terminal glycine is sequestered from the reaction thus impeding reversibility and improving yields.[60] Importantly, the Ni^{2+} strategy is not the only option for improving SML efficiency, and alternate approaches have been described that similarly block reversibility by sequestering or deactivating certain reaction components.[40,61,62]

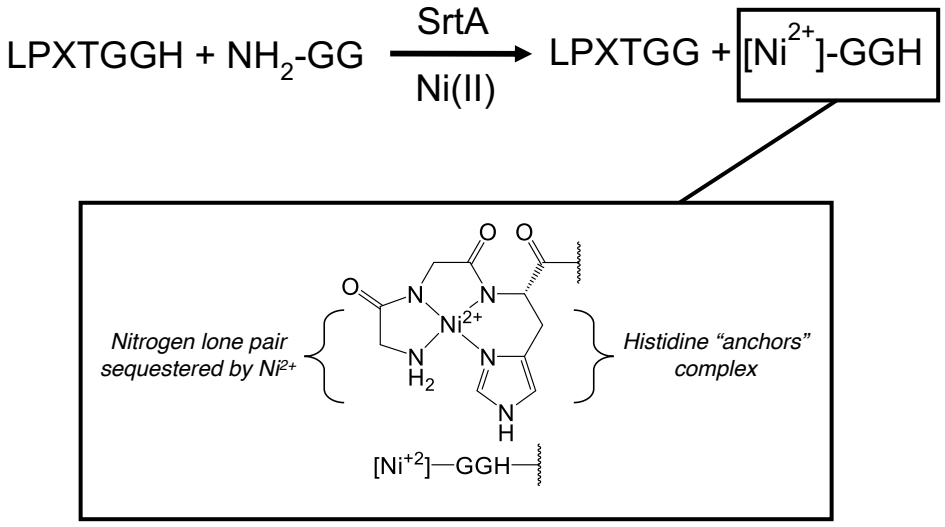


Figure 5. Nickel coordination decreases reaction reversibility.

A final limitation of SML is that the site of modification is generally restricted to the protein C- or N-terminus. This arises from the fact that standard SML reactions occur between C-terminal LPXTG and N-terminal glycine motifs, resulting in standard peptide linkages (**Figure 6**). While modification of protein termini is often useful, this still represents a significant limitation of SML and makes this technique unsuitable for protein targets where the termini are critically important for function.[5] With this in mind, the goal of this work was to adapt SML for the modification of internal sites in proteins, namely the side chain of lysine residues (**Figure 6**). Since the ε-amino groups of lysine residues are typically solvent exposed, they represent an attractive

target for generating unique modified protein derivatives using SML.[3] While chemical methods for lysine modification are well known, they are notoriously nonselective and typically result in a distribution of products involving modifications at multiple lysines as well as the N-terminus.[3,4] The use of an enzyme-mediated process has the potential to improve site-selectivity due to the complex environment of the enzyme active site, which may allow the reaction to be tuned to target specific lysines within a large protein target.

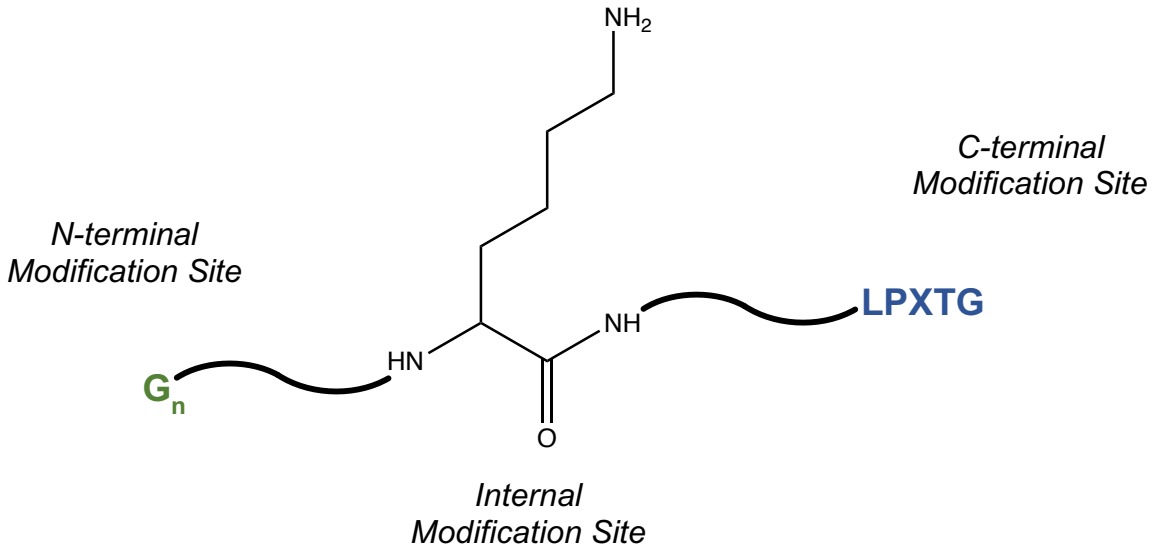


Figure 6. Potential site of internal modification at the side chain of lysine residues.

1.4 Previous Work on Sortase-Mediated Isopeptide Ligation (SMIL)

The primary amine side chain of a lysine residue has the potential to act similarly to the glycine residue as a nucleophile for SML reactions. Rather than forming a standard peptide bond, the lysine would instead form an isopeptide linkage (**Figure 7**).

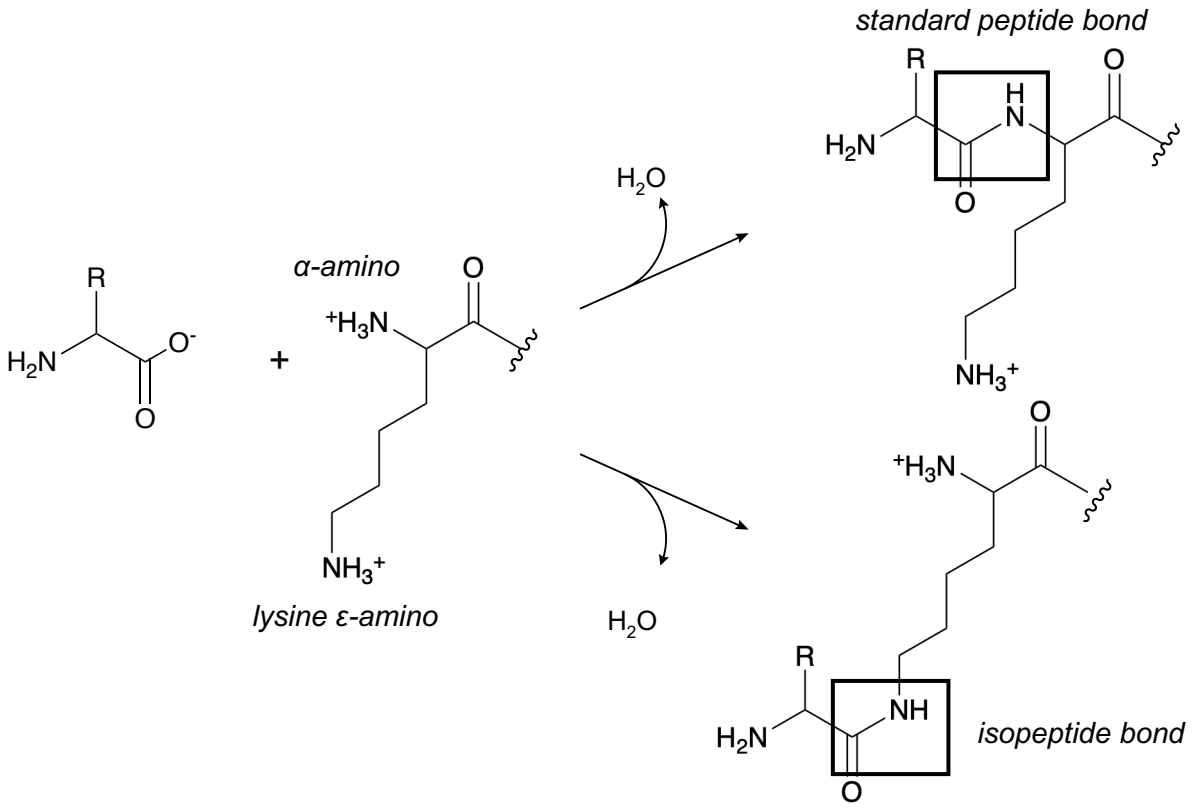


Figure 7. Standard peptide bond versus an isopeptide bond.

This possibility has been explored to a limited extent in the literature, however sortase-mediated isopeptide ligation (SMIL) is often inefficient and is not widely used. Beginning in 2011, isopeptide ligation was actually observed as an unwanted side reaction of SML involving an antibody Fab fragment.[63] An LPETG motif was incorporated on the C-terminus of the light chain and a triglycine fluorophore acted as the nucleophile to form a modified light chain. After the reaction with sortase, the desired SML product was observed along with a protein band at ~50 kDa.[62] Appearance of the 50 kDa protein was accompanied with the vanishing of the Fab light and heavy chain bands, and it was ultimately determined that this polypeptide corresponded to a heavy and light chain covalently linked via an isopeptide bond.[63] This side reaction was noticed to increase under basic conditions likely due to increased deprotonation of the lysine side chain.

This was further confirmed through model SMIL reactions with increasing concentrations of free lysine at pH 7.5 and 9.5.[63] At higher pH there was increased conjugation of the light chain to lysine. The overall yield of this product was still low compared to standard SML involving the triglycine N-terminus, and could be outcompeted with higher concentrations of triglycine and lower pH.[63] Although undesirable, the observed isopeptide ligation indicated that $SrtA_{staph}$ had some capacity to accept the ε-amino group on the side chain of lysine as a nucleophile.

Also in 2011, $SrtA_{staph}$ was purposefully used to form branched oligomers of indolicidin connected together by isopeptide bonds.[31] Expanding on the concept of pilin sortase reactions, this study utilized indolicidin that naturally contains a lysine residue and engineered it to contain a C-terminal LPXTG motif. With both of these components, branched indolicidin oligomers were created with up to six repeat units observed. The reaction was likely between an acyl-enzyme intermediate and free monomeric or oligomeric indolicidin units.[31] However, while successful isopeptide ligation was observed, the efficiency of the reaction was modest and large amounts of starting materials or by-products not involving isopeptide ligation were observed.[31]

Moving to 2014, anti-Her2 monoclonal antibodies (mAb) was genetically modified to contain a pilin domain (VGGS**WLQDVHVYPKH**GGSGR) at the C-terminus of the heavy chain. When reacted with a biotin modified LPXTG peptide, an isopeptide ligation product was created.[64] Confirmed by anti-biotin western blot analysis, ligation only occurred when the pilin domain was present. Furthermore, when the antibody interchain disulfide bonds were reduced, conjugation was exclusive to the heavy chain, suggesting that modification was targeted to the inserted pilin motif. The conversion of unmodified to modified mAb in this case was determined to be ~90%,

however this required a vast excess (100-fold molar excess) of the labeled LPXTG peptide relative to the mAb.[64]

The most recent SMIL study was published in 2018, and utilized a pilus-specific SrtA from *Corynebacterium diphtheriae* (^{Cd}SrtA) to attach a peptide to a protein through an isopeptide bond.[65] In order to do this the authors needed to improve the ligation activity of ^{Cd}SrtA by creating a triple mutant form (^{Cd}SrtA^{3M}).[65] This mutant was able to produce ~10-fold more isopeptide product than ^{Cd}SrtA WT in a model reaction in which the C-terminus of SpaA (^{C}SpaA) protein containing an LPLTG motif was linked to a second copy of SpaA (^{N}SpaA) protein containing the pilin motif (WxxxVxVYPK).[65] While promising, the ^{Cd}SrtA^{3M} was not able to recognize the lysine as a nucleophile when it was present in a peptide mimicking the pilin motif in reactions with either ^{C}SpaA or a peptide containing the LPLTG motif. Given that it was only successful in producing >95% yield when ligating the entire ^{N}SpaA domain with the LPLTG-containing peptide, it suggested that there were additional tertiary elements present in ^{N}SpaA that were required for ^{Cd}SrtA^{3M} to recognize lysine as a nucleophile. Nonetheless, the authors were able to show peptide fluorophore attachment, as well as joining green fluorescent protein containing the C-terminal LPLTG motif to ^{N}SpaA.[65]

In summary, previous efforts have been made to adapt sortase-mediated methods for the construction of isopeptide bonds. However, these reactions have yet to exhibit the same efficiency and generality as SML involving WT SrtA$_{staph}$, and only limited examples of isopeptide ligation in the context of protein modification have been described.

1.5 Project Goals and Overview

SMIL is a potentially powerful addition to the existing set of modification strategies based on bacterial sortase enzymes (**Figure 8**). This approach will expand the types of polypeptide derivatives that could be made using a sortase-based approach, and will allow modification of sites other than the protein termini. It was therefore made the goal of this project to develop an improved strategy for SMIL that is experimentally straightforward, efficient, and compatible with a broad range of polypeptide targets.

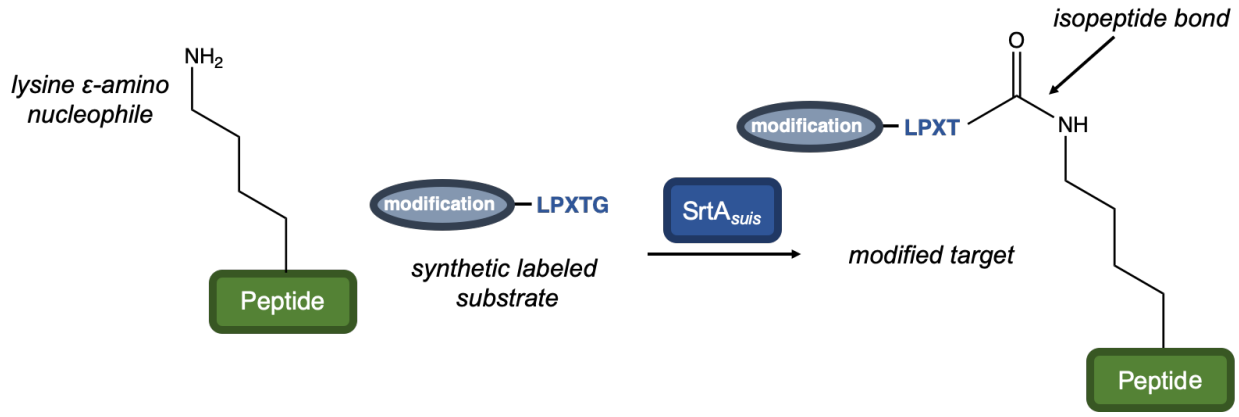


Figure 8. Approach to applying isopeptide ligation to the modification of larger peptide targets.

To this end, we have identified a naturally occurring SrtA homolog from *Streptococcus suis* ($SrtA_{suis}$) and characterized its ability to generate isopeptide bonds *in vitro* using a variety of model peptide substrates. Specifically, recombinant $SrtA_{suis}$ was expressed and purified from *E. coli* and shown to be active under mild aqueous conditions using standard LPXTG substrates. Building from this result, we have succeeded in optimizing a strategy for *in vitro* SMIL that relies on the unique reactivity of $SrtA_{suis}$. Notably, the success of this process relies on applying our previously reported Ni^{2+} coordination strategy for controlling reaction equilibrium. Using this

approach, we have shown that SrtA$_{suis}$ is superior to SrtA$_{staph}$ for SMIL involving model peptides modelled after the *in vivo* substrates of these enzymes. We have further demonstrated that SMIL using SrtA$_{suis}$ allows for the modification of lysines at internal positions of large commercially-available peptides without the need for prior modification or engineering of these targets.

Overall, this work represents a new approach for SMIL that exploits the reactivity of a naturally occurring SrtA homolog that has not been used previously in the context of protein or peptide modification. This reaction proceeds under mild reaction conditions, and provides one of the most efficient sortase-mediated isopeptide ligations reported to date. It is envisioned that this work will be a significant addition to the ever growing bioconjugation toolbox, and will complement existing chemoenzymatic methods for protein modification.

2. Preparation of Sortase Enzymes, Model Peptide, and Larger Targets for SMIL Reactions

2.1 Strategy for SMIL Using Naturally Occurring SrtA Homologs

The evolved role of $SrtA_{staph}$ is to generate standard peptide linkages at the termini polypeptides. This provides a rationale for why it has been so successful in imitating the reaction *in vitro,* and is also the probable reason for why it is relatively unsuccessful in producing isopeptide bonds. Therefore, we sought to identify other naturally occurring sortase homologs to replace $SrtA_{staph}$ for building isopeptide products, which ideally had the ability to form isopeptides *in vivo*. To that end, we identified a SrtA homolog from *Streptococcus suis* ($SrtA_{suis}$) that appeared to have the ability to build isopeptides *in vivo*.[38,66,67] $SrtA_{suis}$, which is presumed to serve a housekeeping role similar to $SrtA_{staph}$, has been shown in the literature to catalyze a transacylation reaction similar to $SrtA_{staph}$ involving LPXTG substrates. However, evidence in the literature suggested that the *in vivo* nucleophile for $SrtA_{suis}$ was significantly different. Specifically, the *S. suis* lipid II subunit lacks a pentaglycine interpeptide bridge and instead contains a single lysine residue where the reactive portion is presumably the nucleophilic side chain (**Figure 9**).[66,38] Due to this, ligation should occur at the ε-amino group, resulting in the formation of an isopeptide bond instead of a standard peptide bond. Given this potential ability of $SrtA_{suis}$ to generate isopeptide bonds *in vivo*, we postulated that it would therefore be a superior tool for generating isopeptide linkages *in vitro*, thereby significantly expanding the capabilities of sortase-based methods. As an additional advantage over $SrtA_{staph}$, $SrtA_{suis}$ activity has also been observed to be Ca^{2+}-independent.[68] This may be a useful property for applications in different cellular environments that contain a broad range of physiological calcium concentrations, such as in the study of cell-cell interactions.[68]

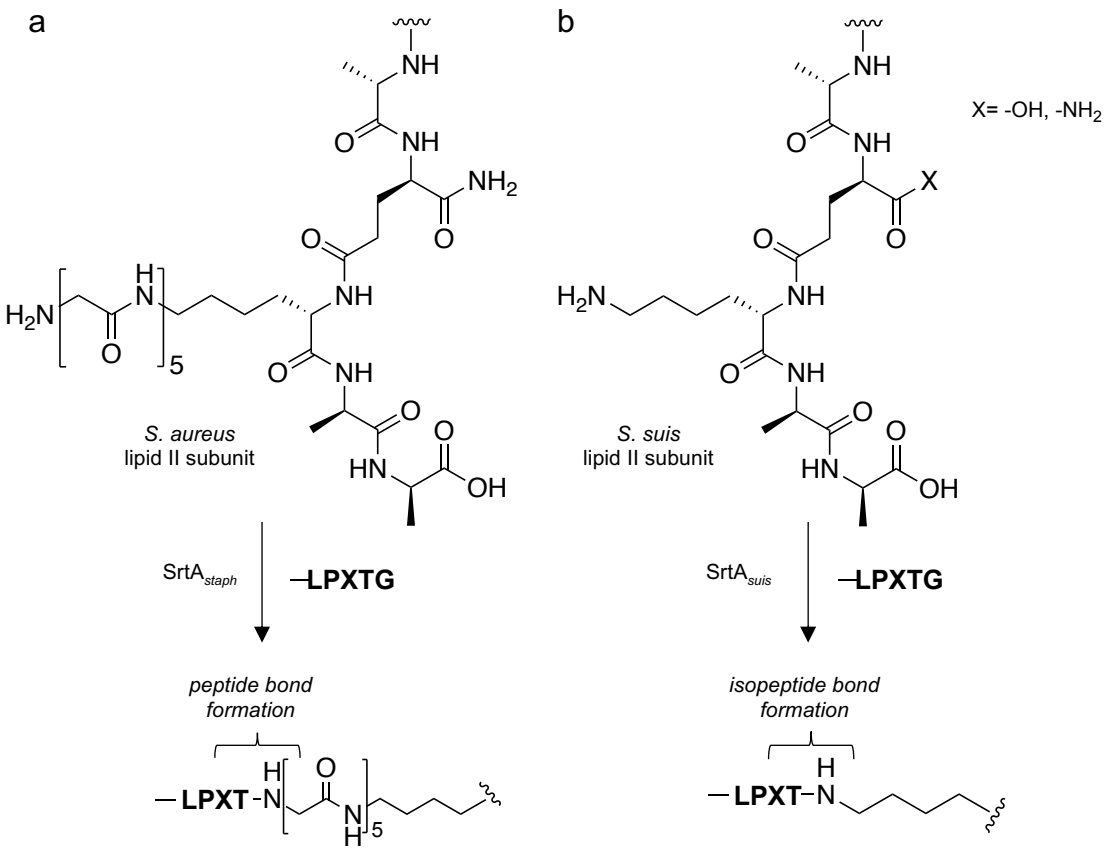


Figure 9. Comparison of the lipid II subunit structures present in $SrtA_{staph}$ (a) and $SrtA_{suis}$ (b) that are postulated to participate in sortase-catalyzed transacylation.

2.2 Expression and Purification of Sortase Enzymes

In order to begin examining the ability of $SrtA_{suis}$ to catalyze *in vitro* SMIL reactions, it was first necessary to produce the requisite enzymes. Prior work from our lab and others had shown that two variants of $SrtA_{suis}$ with truncations at the N-terminus (Δ79 and Δ23) were active *in vitro*. Therefore, bacterial expression plasmids for these enzymes, along with N-terminal His_6 tags, were obtained via commercial gene synthesis. Both $SrtA_{suis}$ clones were transformed into competent BL21(DE3) *E. coli* cells, and protein expression was performed using standard isopropyl β-D-thiogalactopyranoside (IPTG) induction. The enzymes were then purified by

immobilized metal affinity chromatography (IMAC) on a gravity flow-based nickel-nitrilotriacetic acid (Ni-NTA) column to remove most of the contaminants and isolate the desired enzymes possessing N-terminal His_6-tags. The protein solutions were then desalted using a fast protein liquid chromatography (FPLC) system. The enzymes were then characterized by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), electrospray ionization mass spectrometry (ESI-MS), and size exclusion chromatography (SEC) as seen in **Figure 10**. Overall, this revealed enzymes of the appropriate molecular weight, with SEC also suggesting that both (Δ79 and Δ23) clones were predominantly monomeric in solution.

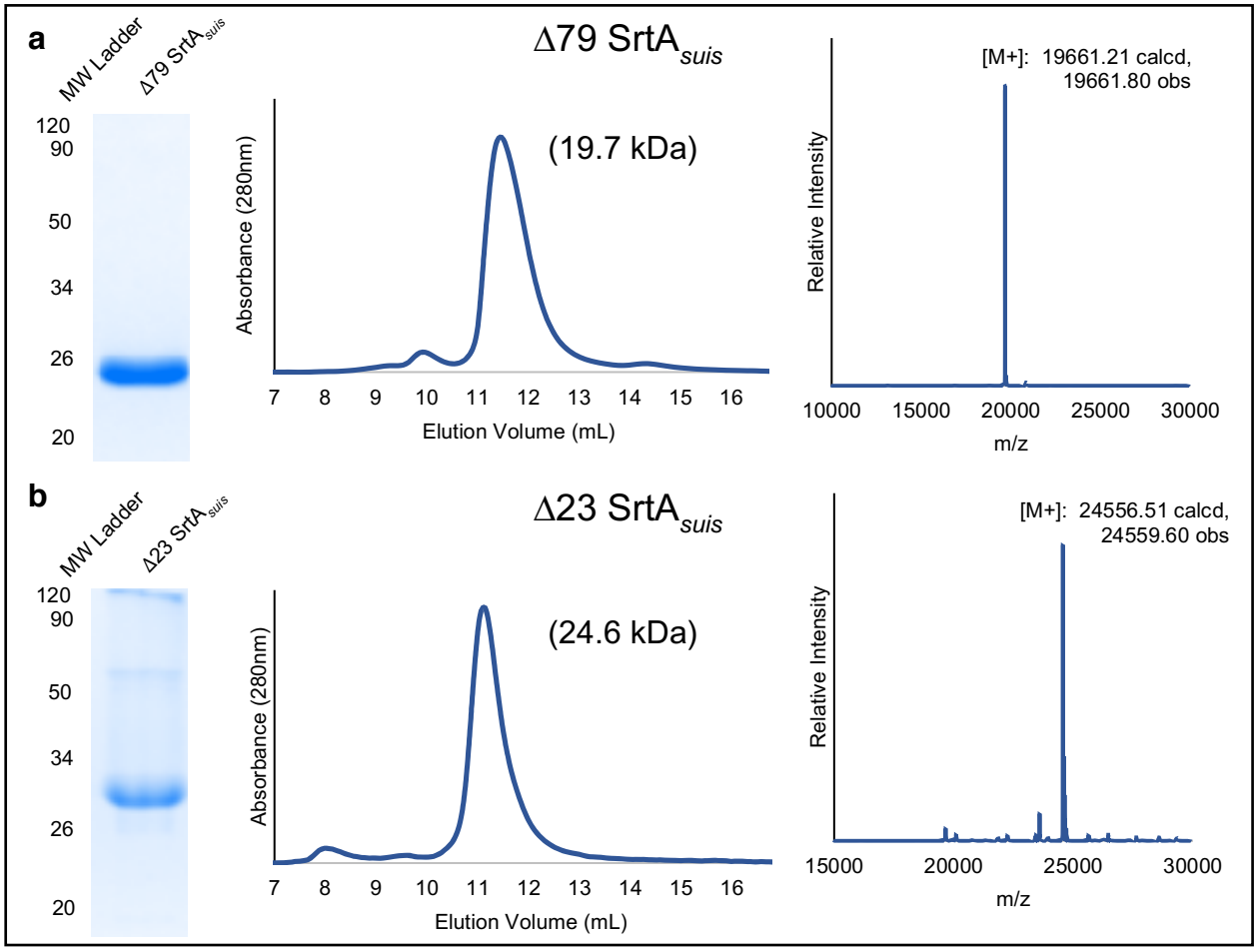


Figure 10. SDS-PAGE, SEC, and ESI-MS data for sortase A clones from *Streptococcus suis*: **a**) Δ79 $SrtA_{suis}$ and **b**) Δ23 $SrtA_{suis}$

2.3 Synthesis of SMIL-Compatible Peptides

As initial model systems, a set of peptides were synthesized for this research that were modeled after LPXTG substrates and lysine-containing nucleophiles that $SrtA_{suis}$ was predicted to utilize *in vivo* for isopeptide ligation reaction. Certain fluorophores and chromophores were also included to assist in reaction monitoring and for quantifying peptide concentrations in solution. Representative reversed phase high performance liquid chromatography (RP-HPLC) chromatograms, ESI-MS data and full structures of all LPXTG-containing substrates utilized in this work are illustrated in **Figure 11**. The first two, Abz-LPATGG-K(Dnp) (**Figure 11a**) and Abz-LPATGGH-K(Dnp) (**Figure 11b**), have a N-terminal aminobenzoyl (Abz) fluorophore and a C-terminal 2,4-dinitrophenyl K(Dnp) chromophore. When Abz and Dnp are conjugated with limited distance between them, as in these peptides, Dnp will quench Abz fluorescence.[69,70] When the threonine-glycine bond is cleaved by sortase, the Abz and Dnp are separated and Abz fluorescence increases. Therefore, monitoring of enzyme cleavage activity can be measured by tracking increases in fluorescence. Additionally, substrate (**b**) contained the extended LPXTGGH sequence for Ni^{2+}-assisted reactions. In addition to the first two substrates, a third substrate (DEAC-Ahx-LPRTGGH) was designed for isopeptide ligations on larger biomolecular targets (**Figure 11c**). It included an N-terminal 7-(diethylamino)coumarin-3-carboxylic acid (DEAC) fluorophore with a strong blue fluorescence (absorbance maximum at 425 nm) to assist with reaction monitoring and to represent a useful protein modification. Additionally, DEAC-Ahx-LPRTGGH also contained the extended LPXTGGH sequence and an arginine residue to promote solubility in water.

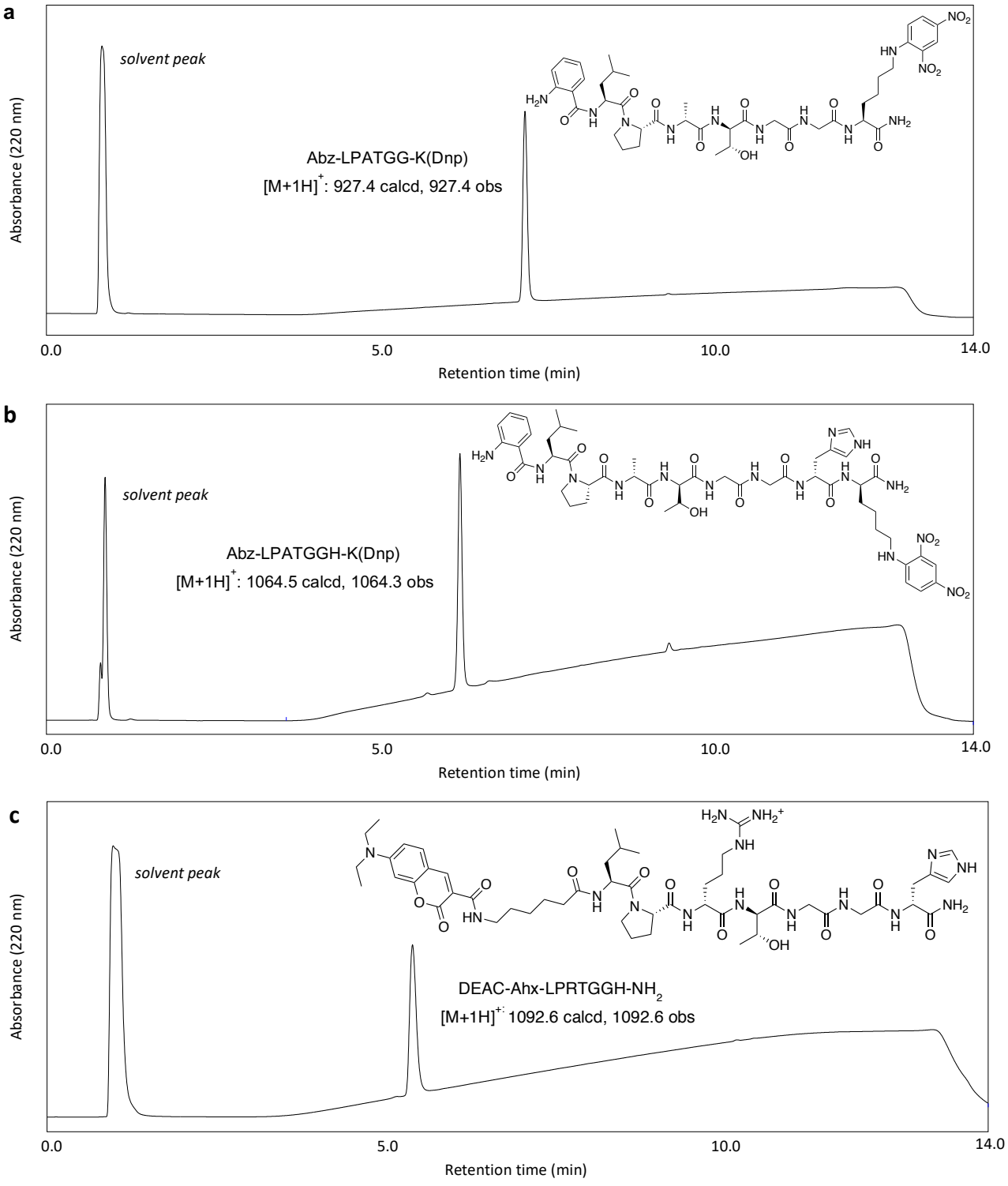


Figure 11. Structures, analytical RP-HPLC, and ESI-MS data for model peptide substrates used in this study: **a**) Abz-LPATGG-K(Dnp), **b**) Abz-LPATGGH-K(Dnp), and **c**) DEAC-Ahx-LPRTGGH-NH_2.

A complementary set of peptides were generated to serve as model lysine nucleophiles. The full structures, pure RP-HPLC traces, and ESI-MS data are illustrated in **Figure 11**. The first nucleophile, Ac-K(Dnp)-AisoQKaa-NH_2 (**Figure 12d**), mimicked the lipid II structure of *S. suis,* including the presence of D-ala residues (aa) and iso-glutamine (iso-Q). The second nucleophile, Ac-K(Dnp)-GGKGG-NH_2 (**Figure 12e**), was designed to be a simplified version of lipid II in order to observe the effects that amino acids neighboring the lysine residue had on nucleophile reactivity in isopeptide ligation. Therefore, this peptide simply contained an internal lysine residue surrounded by glycines. Both lysine nucleophiles contained an N-terminal K(Dnp) chromophore for quantification and to assist with reaction monitoring. They also were acetylated at the N-terminus to prohibit competing standard peptide ligations.

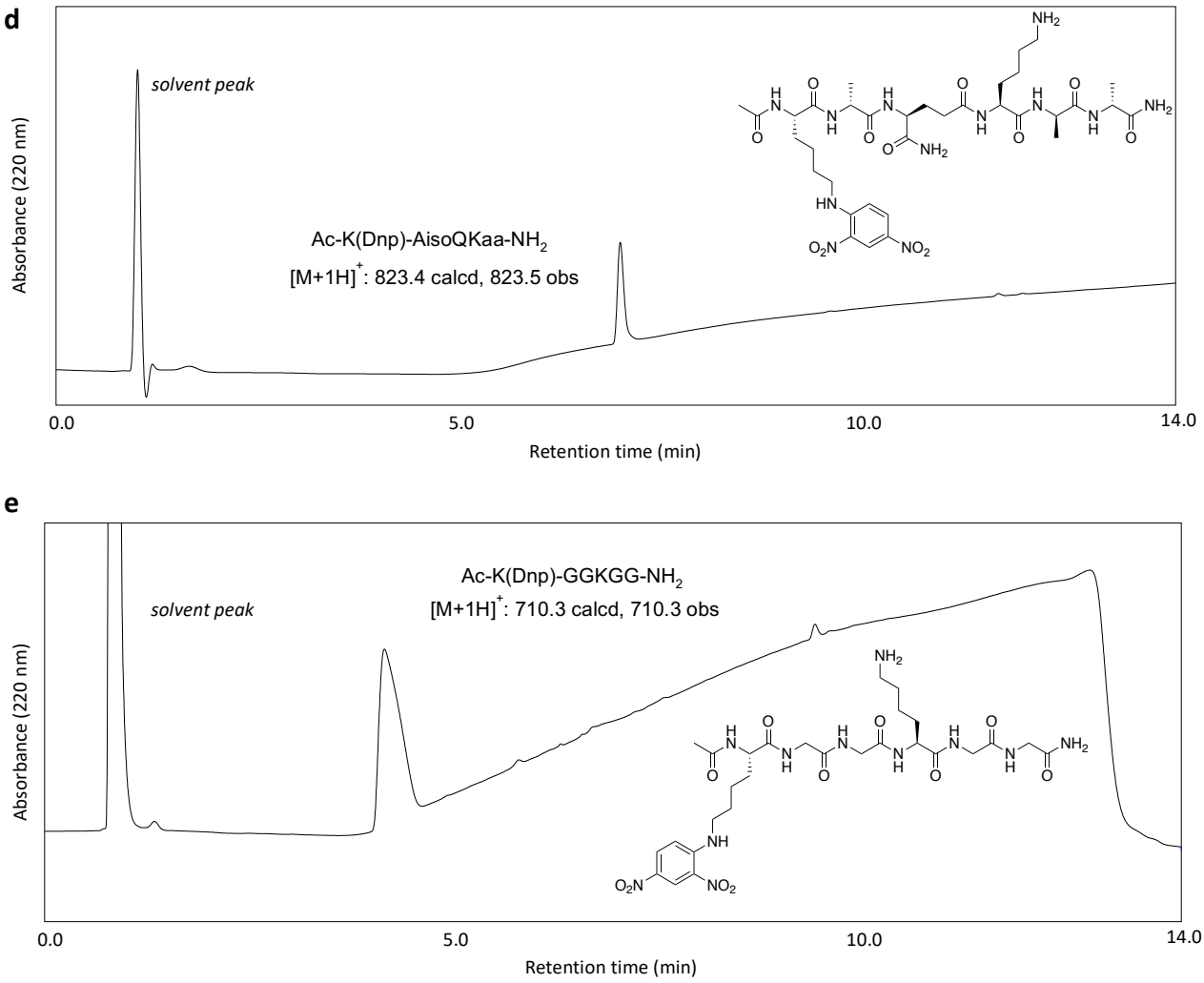


Figure 12. Structures, analytical RP-HPLC, and ESI-MS data for model lysine nucleophiles used in this study: **d**) Ac-K(Dnp)-AisoQKaa-NH_2 and **e**) Ac-K(Dnp)-GGKGG-NH_2.

Preparation of all peptides was achieved through manual fluorenylmethyloxycarbonyl (Fmoc) solid phase peptide synthesis (SPPS). This followed the general strategy represented in **Figure 13**. Briefly, Fmoc protected Rink amide resin was deprotected by treatment with piperidine, followed by HBTU-promoted coupling of the first Fmoc protected amino acid. After repeated deprotection and amino acid coupling cycles, the final products were cleaved from the resin and purified to at least 98%. Stock solutions of all peptides were prepared, and solution

concentrations were estimated using the absorbance of the Dnp chromophore. All stocks were adjusted to ~1 mM final concentration prior to use in SMIL reactions (**Appendix I.**)

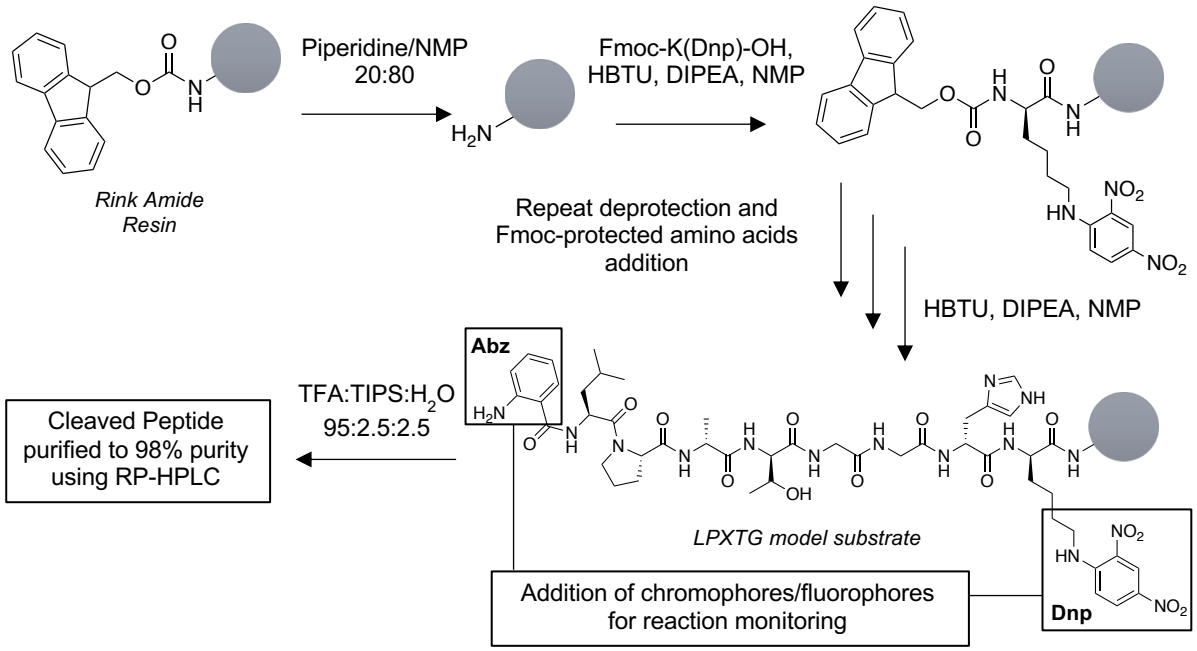


Figure 13. Representative scheme for synthesis of model peptides.

2.4 Larger Peptide Targets

In addition to peptide model systems, we also identified a variety of larger commercially available peptides containing internal lysine residues to serve as nucleophiles in SMIL reactions. The panel of chosen peptides included the following hormone or antimicrobial peptides: indolicidin, glucagon, exendin, and β-endorphin.[71,72,73,74] The criteria used to select these targets were that they displayed varying numbers of internal lysine residue, and also possessed N-terminal residues (Ile, His, Phe) that prior work from our lab had suggested were unlikely to participate in SML.[67] As in the case of the smaller synthetic peptides, the identity and purity of all

larger targets was confirmed by RP-HPLC and ESI-MS (shown along with the full structures of each target in **Figures 14-18).**

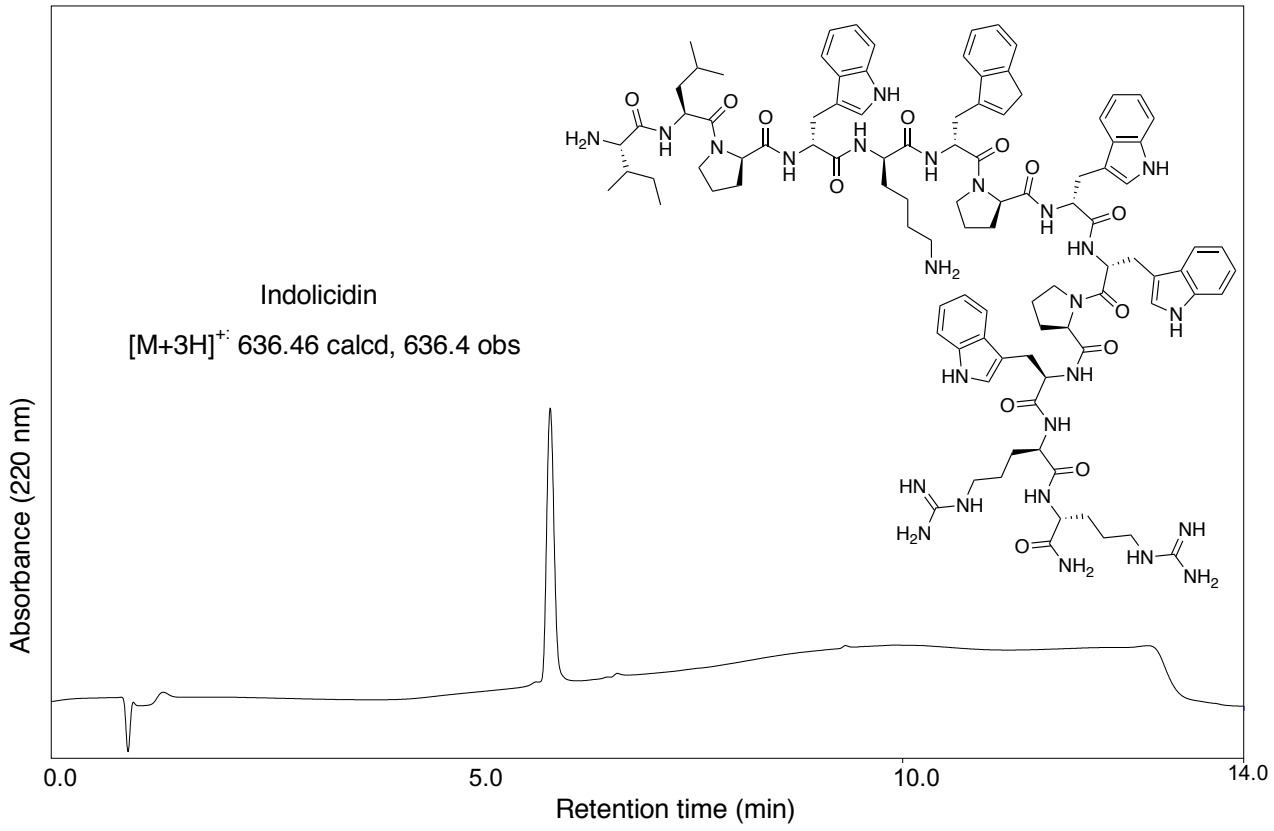


Figure 14. Structure, analytical RP-HPLC, and ESI-MS data for the antimicrobial peptide, indolicidin. Full sequence: *ILPWKWPWWPWRR - NH_2.*

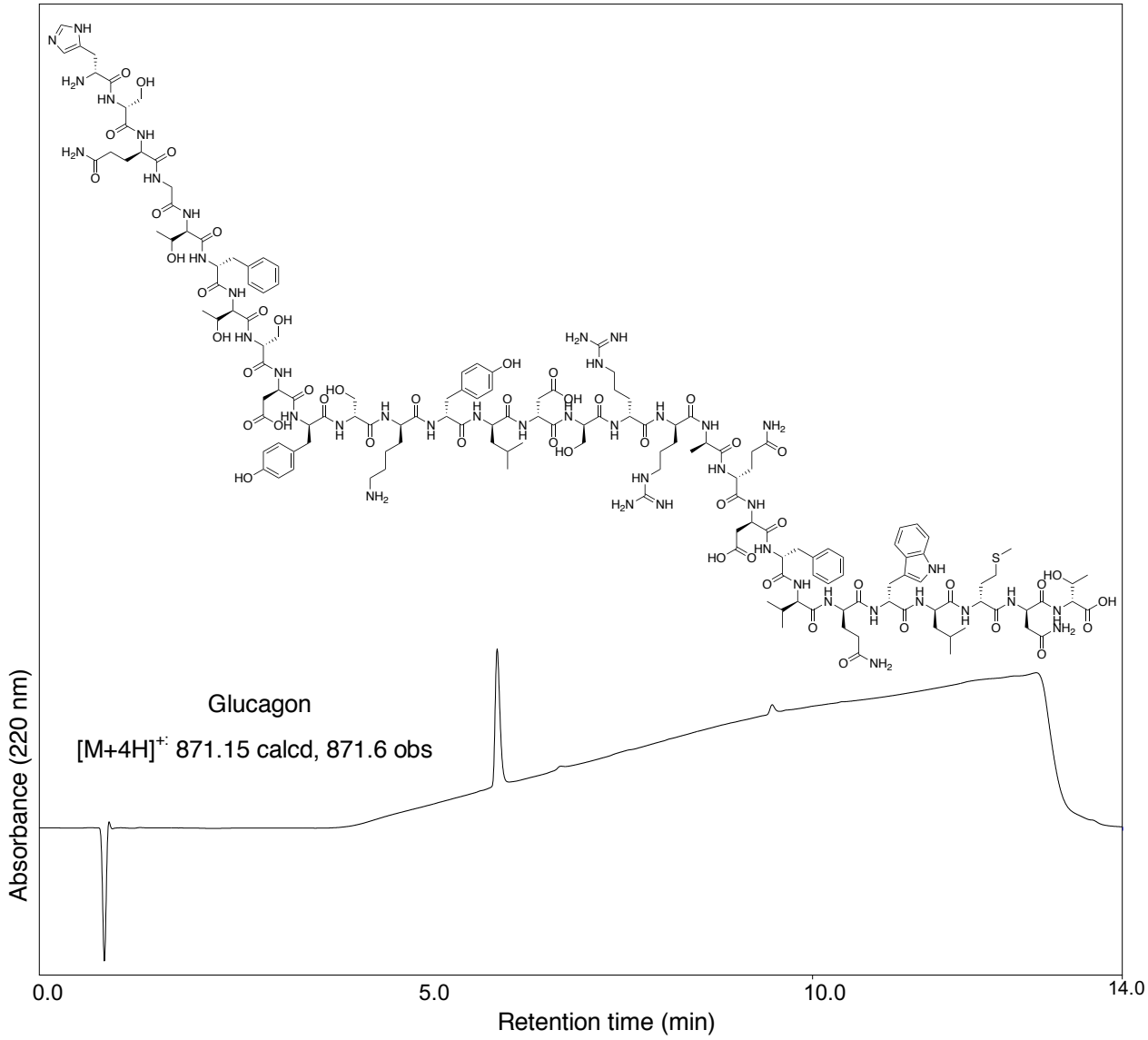


Figure 15. Structure, analytical RP-HPLC, and ESI-MS data for the hormone, glucagon. Full sequence: *HSQGTFTSDYSKYLDSRRAQDFVQWLMNT.*

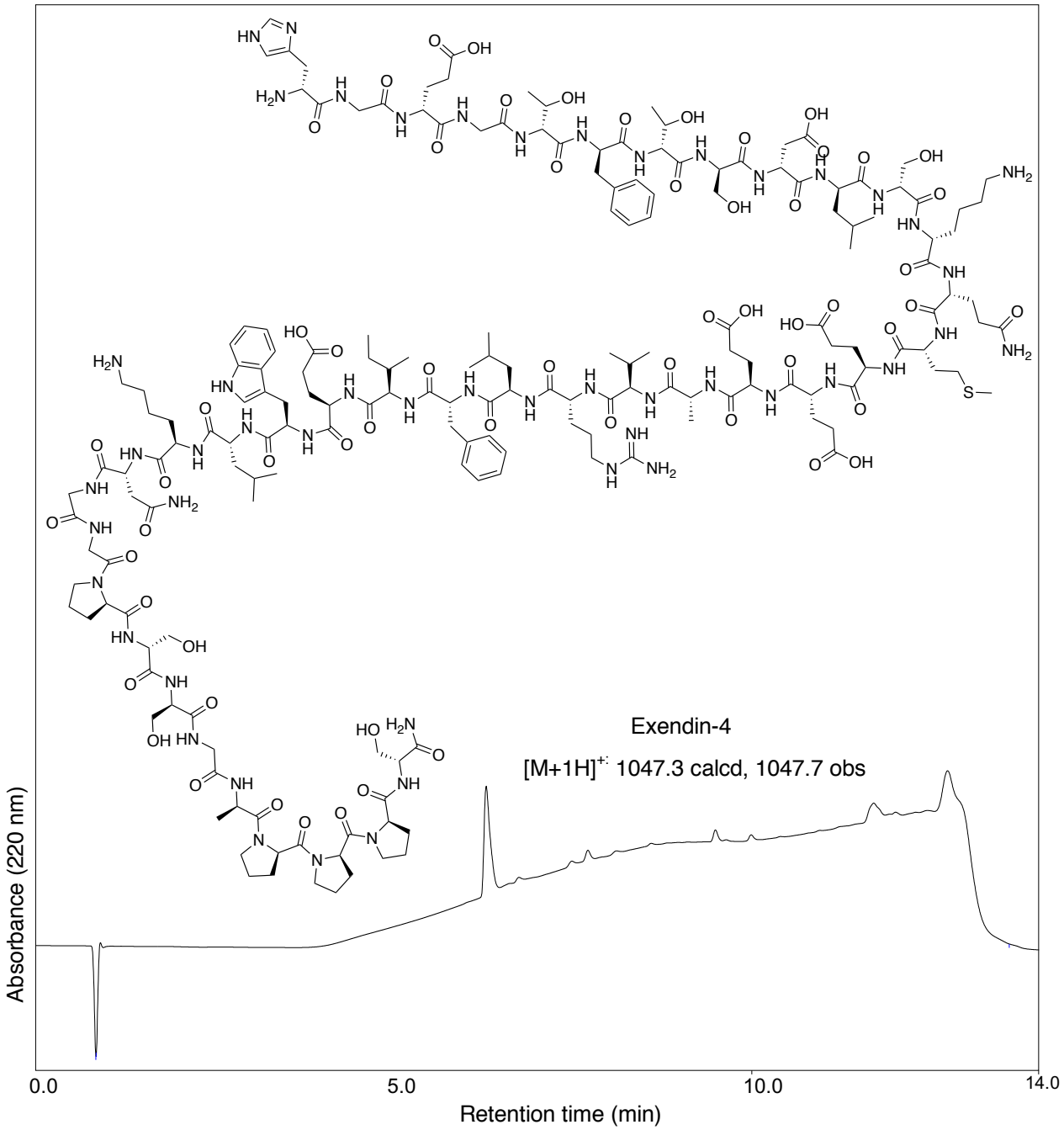


Figure 16. Structure, analytical RP-HPLC, and ESI-MS data for the hormone, exendin-4. Full sequence: *HGEGTFTSDLSKQMEEEAVRLFIEWLKNGGPSSGAPPPS - NH_2*.

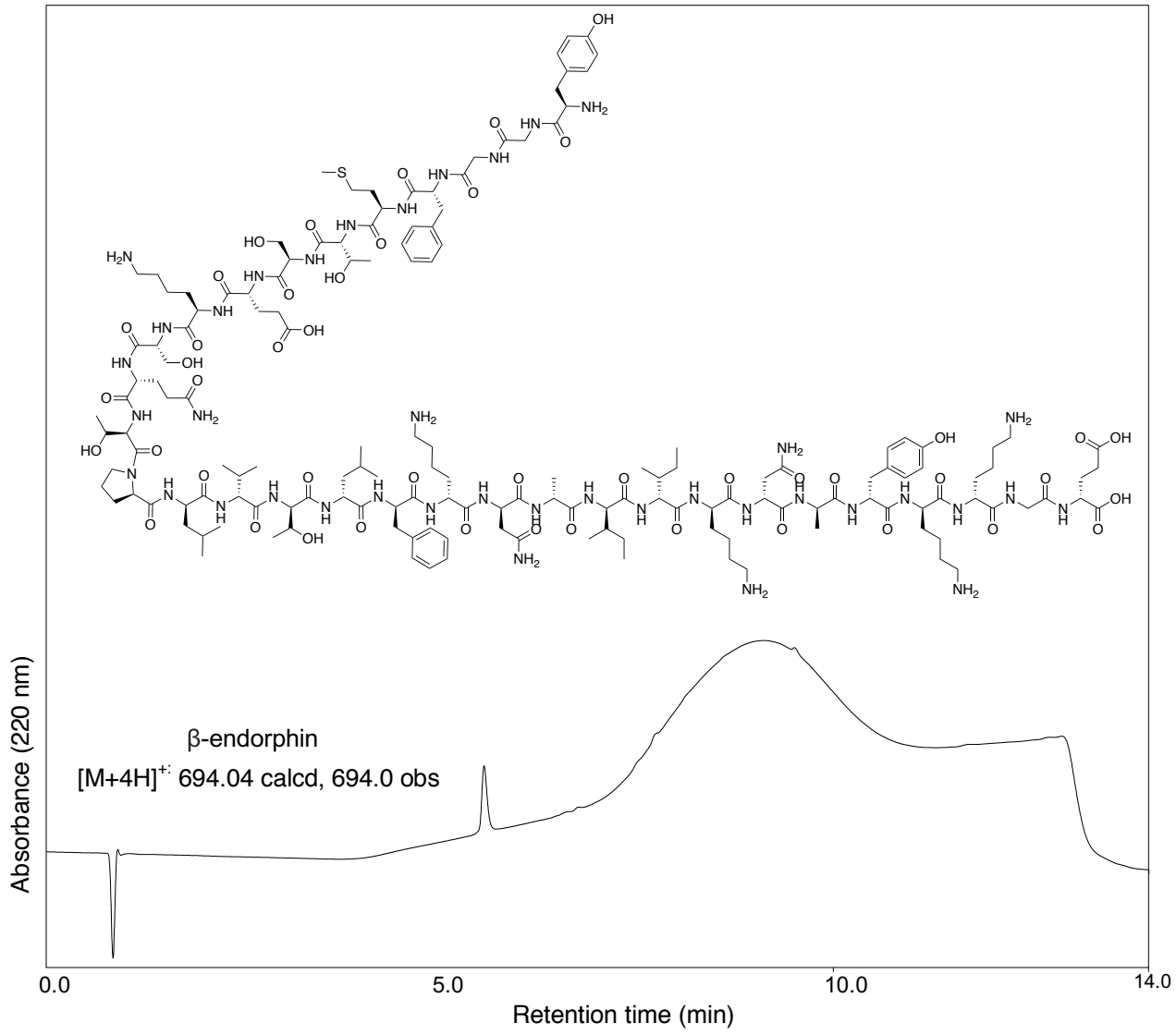


Figure 17. Structure, analytical RP-HPLC, and ESI-MS data for the hormone, β-endorphin. Full sequence: *YGGFMTSEKSQTPLVTLFKNAIIKNAYKKGE*.

3. Sortase-Mediated Isopeptide Ligation

3.1 In Vitro Comparison of LPXTG Cleavage Efficiency

With all the necessary reagents in hand, we began by first verifying the *in vitro* activity of both $SrtA_{suis}$ constructs (Δ79 and Δ23). As described in the previous chapter, the two $SrtA_{suis}$ clones are lacking the first 79 or 23 amino acid residues, which corresponded to the transmembrane portion of these enzymes. These truncations were also similar to the Δ59 derivative of $SrtA_{staph}$ which is known to be active *in vitro*. Thus, in addition to determining whether the different clones of $SrtA_{suis}$ exhibited different activity, we also sought to compare their activity to the well-studied $SrtA_{staph}$.

To that end, the ability of all SrtA clones to cleave the Abz-LPATGG-K(Dnp) model substrate was monitored by fluorescence. Upon successful cleavage of the threonine-glycine peptide bond, the excised fragment containing Dnp (GG-K(Dnp)) was released, no longer quenching the fluorescence of Abz. Therefore, as the reaction progressed and the acyl enzyme intermediate was resolved in the presence of excess hydroxylamine, an increase in fluorescence could be measured. Reactions for each SrtA clone were set up with the same parameters (50 μM Abz-LPATGG-K(Dnp) substrate, 5 mM NH_2OH, and 5 μM SrtA clone). All reactions with $SrtA_{suis}$ were conducted in the absence of Ca^{2+}, while $SrtA_{staph}$ reactions were supplemented with 10 mM Ca^{2+}. The reactions were incubated at room temperature for 1 hour then quenched with the addition of 1:1 dimethylsulfoxide (DMSO):H_2O. Additionally, reactions lacking SrtA were performed as controls to provide a background level of fluorescence for the uncleaved Abz-LPATGG-K(Dnp) probe. As shown in **Figure 18a**, the increase in fluorescence relative to controls was similar for all three enzymes, suggesting that each had the capacity to recognize and cleave

the model LPXTG substrate. To complement these fluorescence studies, the same reactions were set up for Δ79 $SrtA_{suis}$ and Δ23 $SrtA_{suis}$ and analyzed every 30 min via RP-HPLC. Based on the relative peak areas in the 360 nm chromatogram, both Δ79 $SrtA_{suis}$ and Δ23 $SrtA_{suis}$ cleaved >80% of the Abz-LPATGG-K(Dnp) within the first 1-2 hours (**Figure 18b**). Additionally, the peaks observed via RP-HPLC were also analyzed by ESI-MS, which verified that the enzymes were cleaving the Abz-LPATGG-K(Dnp) at the expected site between the threonine and glycine residues (**Appendix II**).

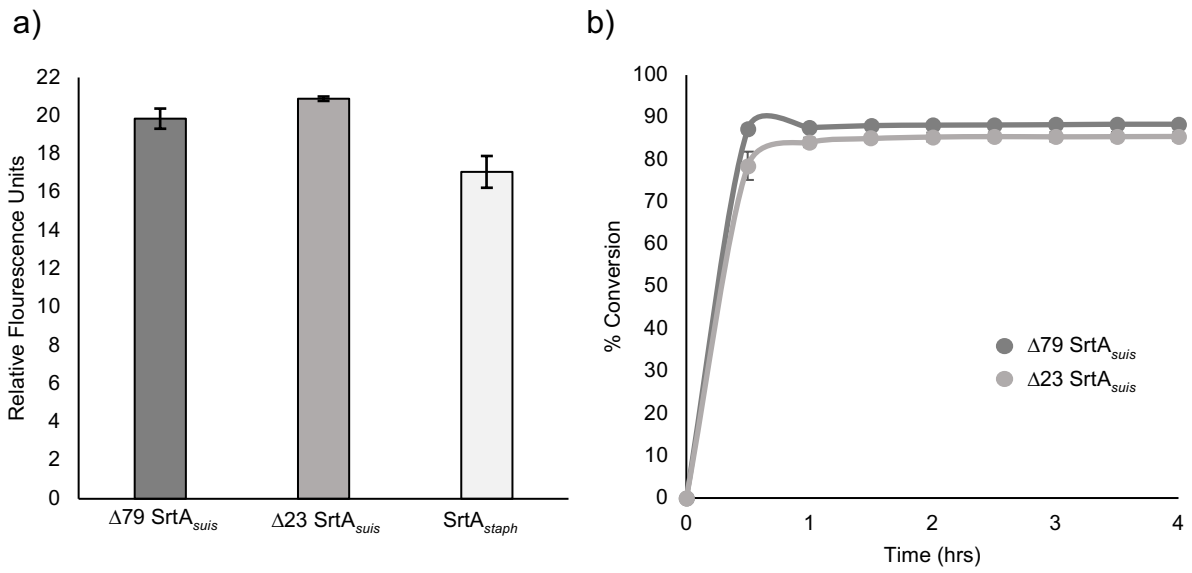


Figure 18. a) Comparison of *in vitro* activity for Δ79 $SrtA_{suis}$, Δ23 $SrtA_{suis}$ and Δ59 $SrtA_{staph}$ based on the increase in observed fluorescence following cleavage of a model Abz-LPATGG-K(Dnp) substrate. Fluorescence increases reported relative to control reactions lacking enzyme. **b)** Comparison of Δ79 $SrtA_{suis}$ and Δ23 $SrtA_{suis}$ activity based on RP-HPLC analysis of model reactions. Reaction conditions: 50 μM Abz-LPATGG-K(Dnp), 5 mM NH_2OH, 5 μM $SrtA_{suis}$, 50 mM Tris pH 7.5, 150 mM NaCl, < 0.6% (v/v) glycerol, 0.6% (v/v) DMSO. Reactions with Δ59 $SrtA_{staph}$ also contained 10 mM $CaCl_2$.

3.2 Model Isopeptide Ligation Studies

Having confirmed that both clones of $SrtA_{suis}$ were active *in vitro*, we proceeded to test their ability to conjugate the LPXTG motif containing peptides with the Lipid II peptide models (Ac-K(Dnp)-AisoQKaa-NH_2, Ac-K(Dnp)-GGKGG-NH_2) displaying internal lysine nucleophiles. For these reactions, a modest twofold excess of LPXTG substrate was employed relative to the Lipid II peptide models in an attempt to consume all of the lysine-containing nucleophiles. Representative RP-HPLC chromatograms for model isopeptide ligations involving the Ac-K(Dnp)-AisoQKaa-NH_2 nucleophile are shown in **Figure 19**. In the presence of either Δ79 $SrtA_{suis}$ and Δ23 $SrtA_{suis}$, peaks consistent with the substrates (**S**), nucleophile (**N**), excised fragment (**EF**), and the desired isopeptide ligation product (**LP**) were present as determined by ESI-MS (**Appendix II**). While we were encouraged to obtain evidence of the desired isopeptide ligation, overall production of this material was limited. Specifically, the reactions catalyzed by Δ23 $SrtA_{suis}$ gave only ~15-16% conversion of Ac-K(Dnp)-AisoQKaa-NH_2 to the ligation product, with maximum amounts observed after 16 h incubations at room temperature. Reactions with Δ79 $SrtA_{suis}$ showed modest improvements, exhibiting maximum reaction conversions of 18-21% after a 2 h reaction time. For all reactions, the choice of substrate (LPATGG versus LPATGGH) appeared to have little impact on reaction efficiency.

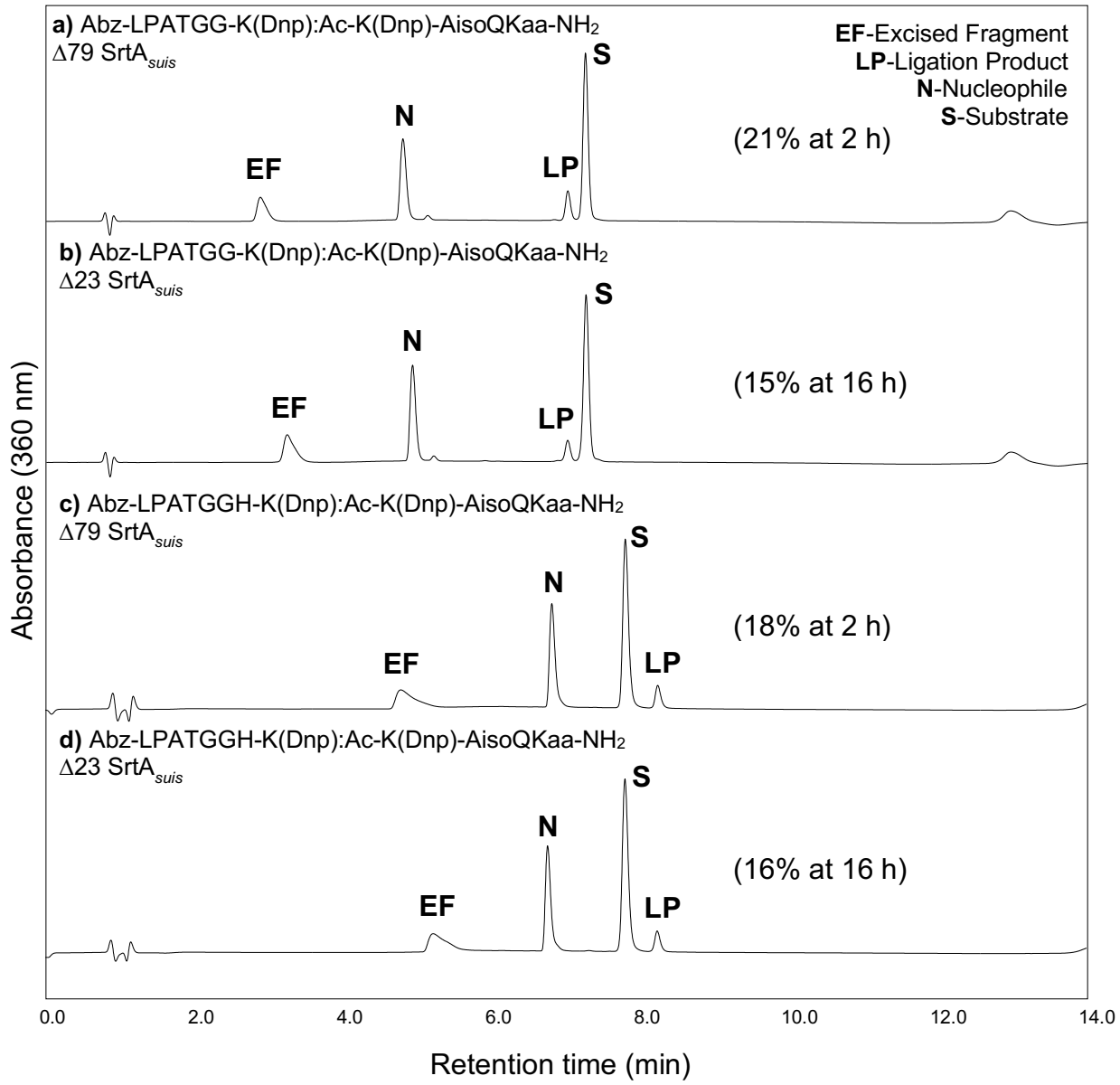


Figure 19. Model isopeptide ligation catalyzed by Δ79 SrtA$_{suis}$ and Δ23 SrtA$_{suis}$. RP-HPLC chromatograms represent the timepoints of highest reaction conversion utilizing the Abz-LPATGG-K(Dnp) (**a-b**) or Abz-LPATGGH-K(Dnp) (**c-d**) substrates. Peak Labels: **EF** = GG-K(Dnp) or GGH-K(Dnp) excised fragment, **LP** = Abz-LPAT-(Ac-K(Dnp)-AisoQ**K**aa-NH_2) ligation product, **N** = Ac-K(Dnp)-AisoQKaa-NH_2 nucleophile, **S** = Abz-LPATGG-K(Dnp) or Abz-LPATGGH-K(Dnp) substrate. Reaction conditions: 100 μM **S**, 50 μM **N**, 5 μM SrtA$_{suis}$, 50 mM Tris pH 7.5, 150 mM NaCl, < 0.6% (v/v) glycerol, < 1.5% (v/v) DMSO. Reactions **a-b** were analyzed by RP-HPLC (**Method E** (**Appendix III**), 360 nm) and reactions **c-d** were analyzed by RP-HPLC (**Method D** (**Appendix III**), 360 nm). The extent of ligation product formation was estimated from peak areas derived from the 360 nm chromatogram.

In addition to reaction utilizing Ac-K(Dnp)-AisoQKaa-NH_2, we also conducted a similar screen employing the structurally simplified Ac-K(Dnp)-GGKGG-NH_2 nucleophile. As shown in **Figure 20**, both clones of $SrtA_{suis}$ appeared to yield small amounts of the expected isopeptide ligation product, as verified by ESI-MS characterization of the corresponding RP-HPLC peak (**Appendix II**). However, reaction progress was generally lower than that obtained with Ac-K(Dnp)-AisoQKaa-NH_2, suggesting that Ac-K(Dnp)-GGKGG-NH_2 was an even less effective nucleophile. Once again, the Δ79 $SrtA_{suis}$ construct proved to be more successful than Δ23 $SrtA_{suis}$, however reaction conversion never exceeded 18%.

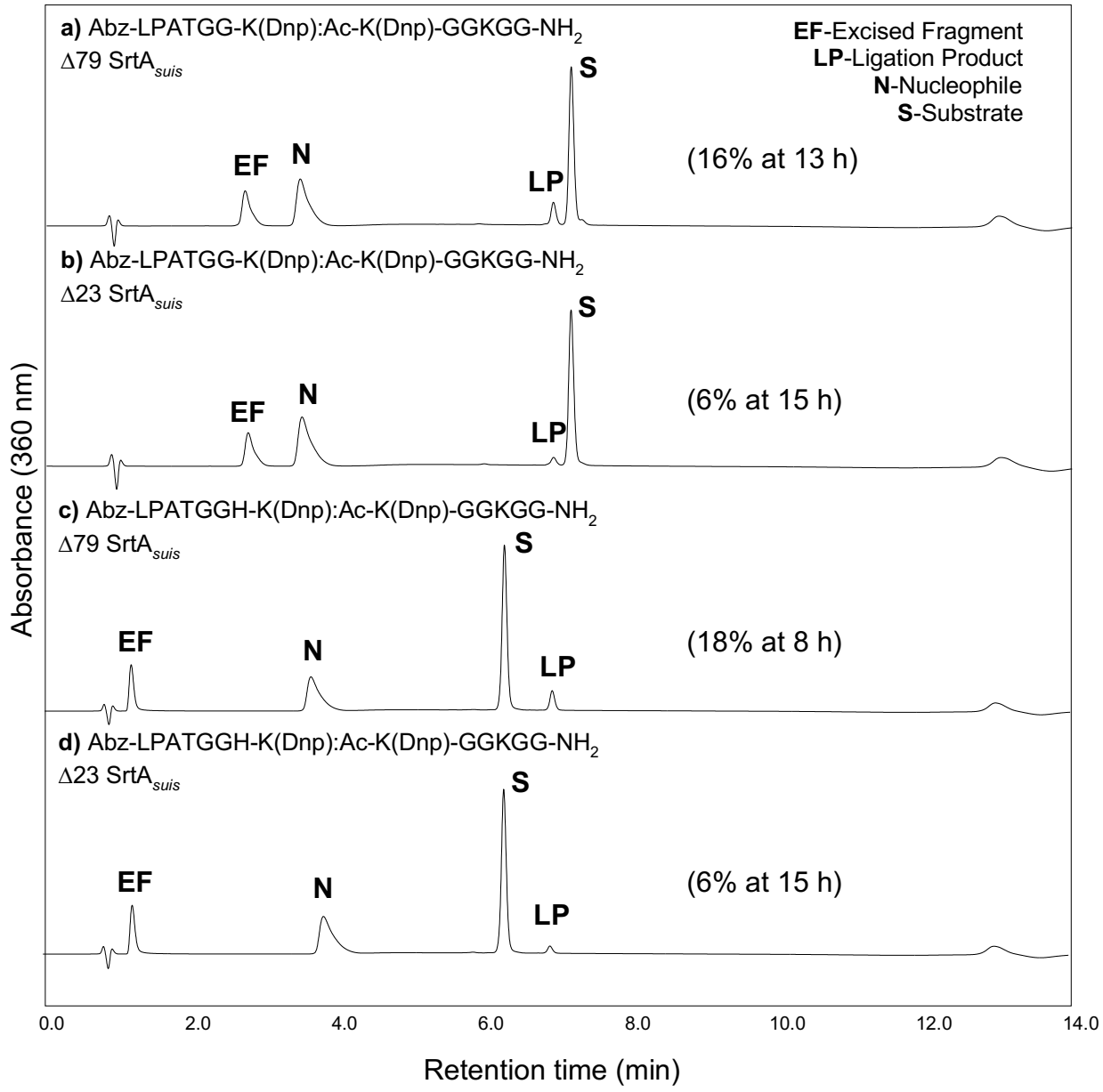


Figure 20. Model isopeptide ligation catalyzed by Δ79 SrtA$_{suis}$ and Δ23 SrtA$_{suis}$. RP-HPLC chromatograms represent the timepoints of highest reaction conversion utilizing the Abz-LPATGG-K(Dnp) (**a**-**b**) or Abz-LPATGGH-K(Dnp) (**c**-**d**) substrates. Peak Labels: **EF** = GG-K(Dnp) or GGH-K(Dnp) excised fragment, **LP** = Abz-LPAT-(Ac-K(Dnp)-GG**K**GG-NH_2) ligation product, **N** = Ac-K(Dnp)-GGKGG-NH_2 nucleophile, **S** = Abz-LPATGG-K(Dnp) or Abz-LPATGGH-K(Dnp) substrate. Reaction conditions: 100 μM **S**, 50 μM **N**, 5 μM SrtA$_{suis}$, 50 mM Tris pH 7.5, 150 mM NaCl, < 0.6% (v/v) glycerol, < 2.5% (v/v) DMSO. Reactions were analyzed by RP-HPLC and ESI-MS (**Method E** (**Appendix III**), 360 nm. The extent of ligation product formation was estimated from peak areas derived from the 360 nm chromatogram.

In total, our peptide model systems did suggest that $SrtA_{suis}$ was capable of catalyzing the formation of isopeptide bonds *in vitro*. Additionally, our results revealed that the choice of substrate (LPATGG versus LPATGGH) had little impact on the success of the reaction, while the choice of enzyme and nucleophile was potentially more significant. Specifically, the Ac-K(Dnp)-AisoQKaa-NH_2 nucleophile, which more closely mirrored the *in vivo* lipid II structure of *S. suis*, provided higher levels of product formation than the structurally simpler Ac-K(Dnp)-GGKGG-NH_2 nucleophile. In addition, the shortened clone of $SrtA_{suis}$ (Δ79) gave the highest levels of product formation, and was able to achieve these levels in significantly shorter reaction times than those observed for Δ23 $SrtA_{suis}$. Therefore, we elected to focus on the Δ79 $SrtA_{suis}$ for further reaction optimization.

3.3 Metal Assisted Model Peptide Studies

In an attempt to improve isopeptide ligation efficiency, we hypothesized that reducing the reversibility of the reaction could lead to higher levels of product formation. As described in Chapter 1, reaction reversibility is a known issue for many applications of sortase-mediated ligation, and circumventing this problem often provides a means for pushing the reaction forward towards the production of desired product. To that end, we sought to explore the Ni^{2+} coordination strategy previously developed in our lab.

First, the model Δ79 $SrtA_{suis}$ catalyzed ligation of Ac-K(Dnp)-GGKGG-NH_2 and Abz-LPATGGH-K(Dnp) was repeated in the presence of $NiSO_4$ (**Figure 21c**). As in the case of our previous model studies, product conversion was monitored by RP-HPLC and estimated from peak areas derived from the 360 nm chromatogram. It should be noted that with the Ni^{2+} additive a

small shift in retention time was observed in HPLC traces. The identity of all peaks was also assessed by ESI-MS (**Appendix II**). In contrast to reactions lacking Ni^{2+} (**Figure 21a,b**), the system employing the metal additive resulted in a 2-3 fold increase in product formation, reaching a maximum of 50% after 5 h at room temperature. To ensure that this increase in ligation efficiency was specific to Δ79 $SrtA_{suis}$, we conducted a control using $SrtA_{staph}$ (**Figure 21d**). Consistent with previous reports, as well as our hypothesis about the isopeptide capabilities of this enzyme, we observed that $SrtA_{staph}$ was significantly less effective, producing only 7% product after 16 h in the presence of the Ni^{2+} additive.

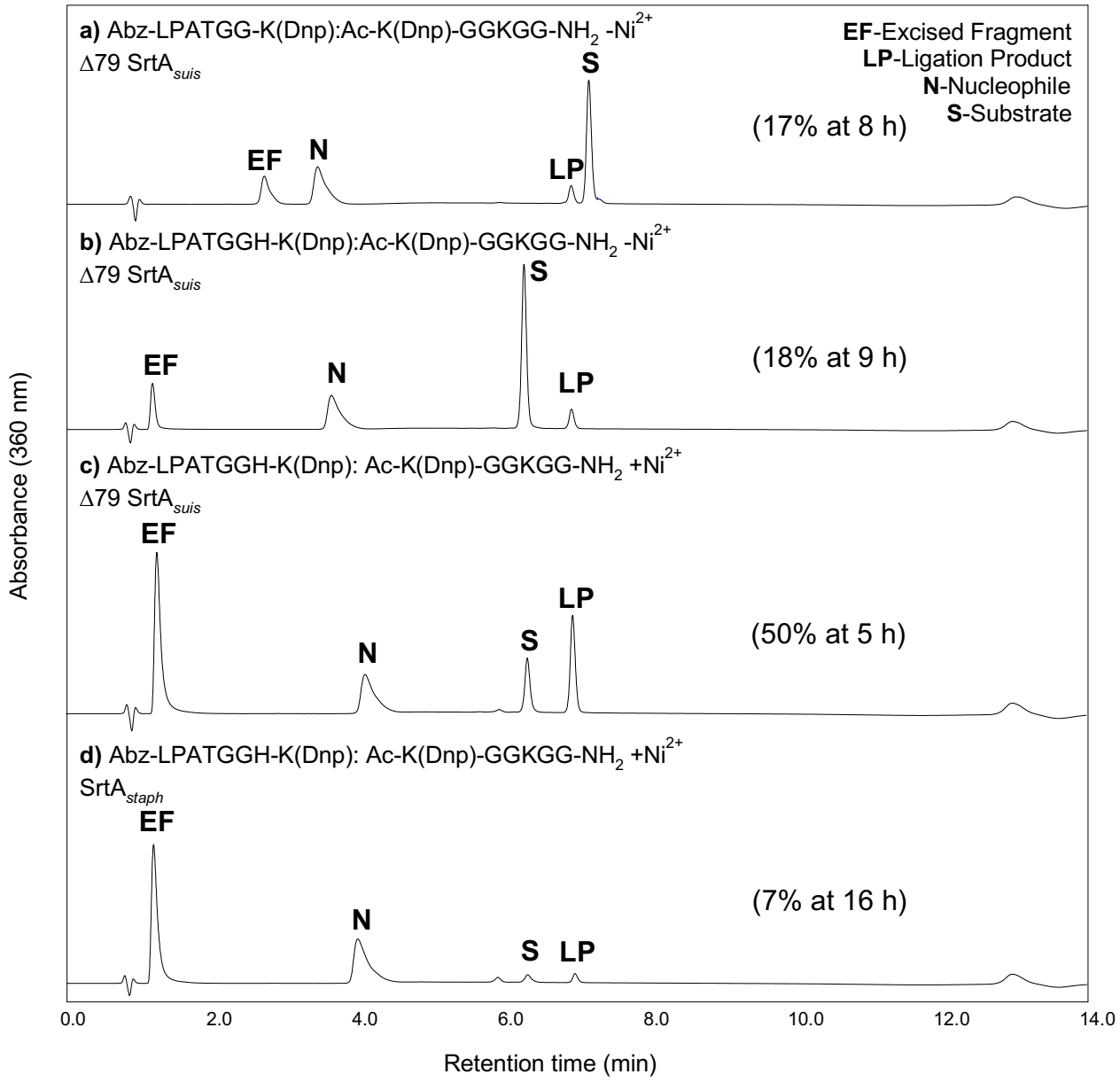


Figure 21. Effects of Ni^{2+} addition on ligation efficiency between Ac-K(Dnp)-GGKGG-NH_2 and Abz-LPATGG-K(Dnp) (**a**) or Abz-LPATGGH-K(Dnp) (**b-d**). Peak Labels: **EF** = GG-K(Dnp) or GGH-K(Dnp) excised fragment, **LP** = Abz-LPAT-(Ac-K(Dnp)-GG**K**GG-NH_2) ligation product, **N** = Ac-K(Dnp)-GGKGG-NH_2 nucleophile, **S** = Abz-LPATGG-K(Dnp) (**a**) or Abz-LPATGGH-K(Dnp) (**b-d**) substrate. Reaction conditions: 100 μM **S**, 50 μM **N**, 5 μM Δ79 $SrtA_{suis}$ (**a-c**) or 5 μM $SrtA_{staph}$ (**d**), 50 mM Tris pH 7.5, 150 mM NaCl, 200 μM $NiSO_4$, 0.1% (v/v) glycerol, < 2.5% (v/v) DMSO. Reactions with Δ59 $SrtA_{staph}$ also contained 10 mM $CaCl_2$. Reactions were analyzed by RP-HPLC and ESI-MS (**Method E** (**Appendix III**), 360 nm).

A similar series of reactions were then conducted using the Ac-K(Dnp)-AisoQKaa-NH_2 nucleophile (**Figure 22**). Once again, the use of Ni^{2+} proved extremely effective when combined with the Abz-LPATGGH-K(Dnp) substrate. A 3-fold improvement over reactions lacking Ni^{2+} was observed, with a maximum of 65% conversion after 2 h under the most successful reaction conditions (**Figure 22c**). As in the case of the Ac-K(Dnp)-GGKGG-NH_2 nucleophile, all reaction species were identified by ESI-MS (**Appendix II**). Additionally, a control was performed using $SrtA_{staph,}$ and as expected this enzyme gave far lower levels of ligation product formation (~5% after 16 h), indicating that Δ79 $SrtA_{suis}$ was uniquely capable of generating isopeptide bonds *in vitro.*

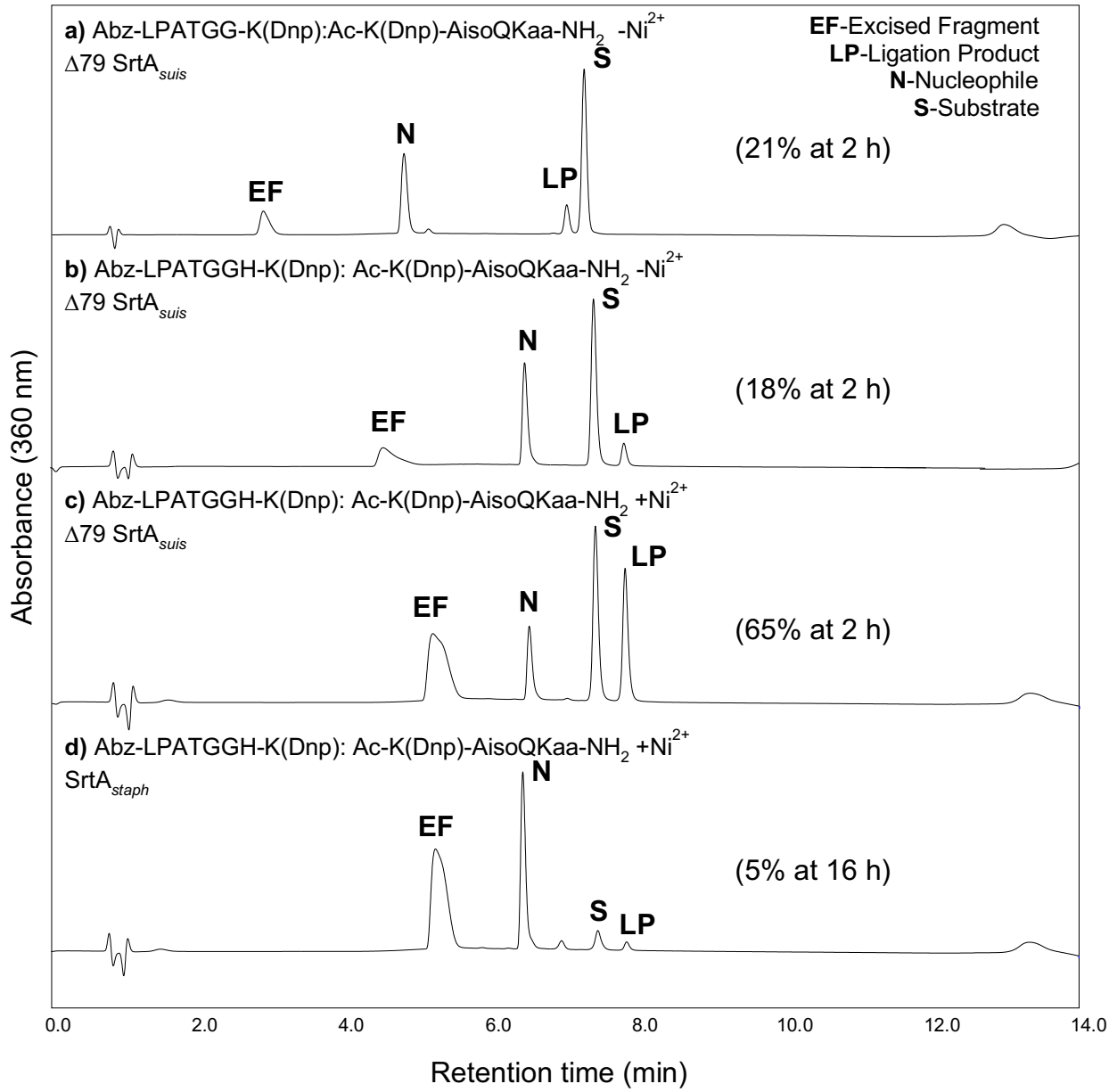


Figure 22. Effect of Ni^{2+} addition on ligation efficiency between Ac-K(Dnp)-AisoQKaa-NH_2 and Abz-LPATGG-K(Dnp) (**a**) or Abz-LPATGGH-K(Dnp) (**b-d**). Peak Labels: **EF** = GG-K(Dnp) or GGH-K(Dnp) excised fragment, **LP** = Abz-LPAT-(Ac-K(Dnp)-AisoQ**K**aa-NH_2) ligation product, **N** = Ac-K(Dnp)-AisoQKaa-NH_2 nucleophile, **S** = Abz-LPATGG-K(Dnp) (**a**) or Abz-LPATGGH-K(Dnp) (**b**) substrate. Reaction conditions: 100 μM **S**, 50 μM **N**, 5 μM $SrtA_{suis}$ (**a-c**) or 5 μM $SrtA_{staph}$ (**d**), 50 mM Tris pH 7.5, 150 mM NaCl, 200 μM $NiSO_4$, 0.1% (v/v) glycerol, < 1.5% (v/v) DMSO. Reactions with Δ59 $SrtA_{staph}$ also contained 10 mM $CaCl_2$. Reactions were analyzed by RP-HPLC and ESI-MS (**Method D** (**Appendix III**), 360 nm) with the exception of reaction (**a**) which used **Method E** (**Appendix III**), 360 nm.

3.4 Confirmation of Isopeptide Ligation Using MS/MS Characterization

Having succeeded in producing significant quantities of a ligation product with a molecular weight appropriate for isopeptide ligation, we next sought to gather additional evidence that the reaction was indeed occurring at the lysine side chain. Specifically, in order to confirm the position of the covalent modification, the putative isopeptide ligation products were collected via RP-HPLC and characterized by MS/MS analysis. The predicted b and y fragmentation pattern for the Ac-K(Dnp)-GG**K**GG-NH_2 isopeptide product is shown in **Figure 23**, and b and y ions observed after MS/MS analysis are summarized in **Table 1**. The same fragmentation pattern and observed b and y ions for the Ac-K(Dnp)-AisoQ**K**aa-NH_2 isopeptide product are shown in **Figure 24** and **Table 2**.

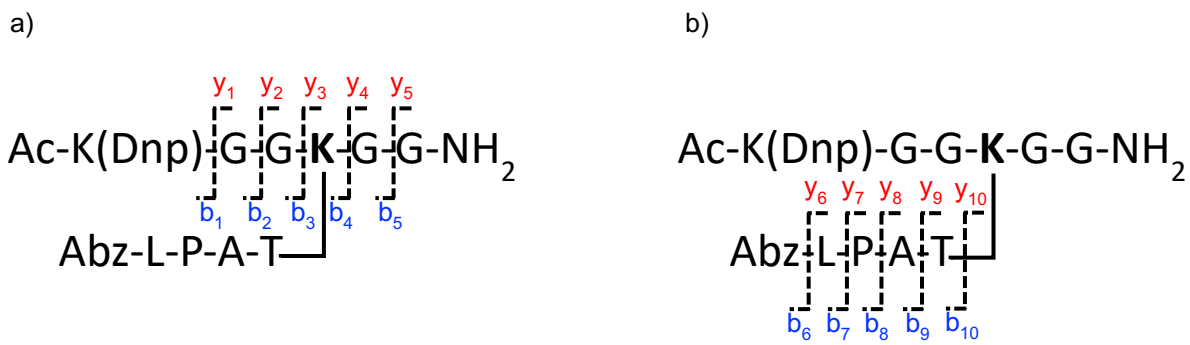


Figure 23. Visual representation of b and y ions from MS/MS analysis of Abz-LPAT modified Ac-K(Dnp)-GG**K**GG-NH_2. Fragmentation occurring in the Ac-K(Dnp)-GG**K**GG-NH_2 portion (a) and the Abz-LPAT modification (b). **K** = isopeptide ligation to Abz-LPAT.

Table 1. Expected and observed b and y ions from MS/MS analysis of Abz-LPAT labeled Ac-K(Dnp)-GG**K**GG-NH_2. Ions created from fragmentation in the Ac-K(Dnp)-GG**K**GG-NH_2 peptide (1) and the Abz-LPAT modification (2) from parent mass m/z = 1211.56. **K** = isopeptide ligation to Abz-LPAT.

1) Ions Produced From Ac-K(Dnp)-GG**K**GG-NH_2 Peptide Chain					
Ion	Expected	Observed	Ion	Expected	Observed
b_1	337.11	337.02	y_1	875.46	875.20
b_2	394.14	394.04	y_2	818.44	818.20
b_3	451.16	451.04	y_3	761.42	761.22
b_4	1080.51	1080.19	y_4	132.06	132.05
b_5	1137.53	1137.20	y_5	75.04	75.05
2) Ions Produced From Abz-LPAT Modification					
Ion	Expected	Observed	Ion	Expected	Observed
b_6	120.4	120.02	y_6	1092.53	N/O*
b_7	233.13	233.08	y_7	979.44	979.18
b_8	330.18	N/O*	y_8	882.39	882.16
b_9	401.22	401.11	y_9	811.35	811.14
b_{10}	502.27	502.14	y_{10}	710.31	710.13

*Not Observed

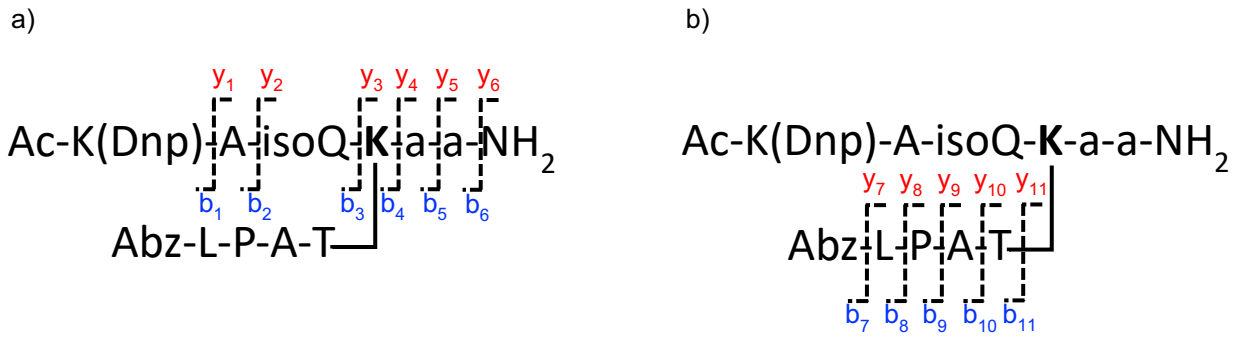


Figure 24. Visual representation of b and y ions from MS/MS analysis of Abz-LPAT labeled Ac-K(Dnp)-AisoQKaa-NH_2. Fragmentation occurring in the Ac-K(Dnp)-AisoQKaa-NH_2 peptide (a) and the Abz-LPAT modification (b). **K** = isopeptide ligation to Abz-LPAT.

Table 2. Expected and observed b and y ions from MS/MS analysis of Abz-LPAT labeled Ac-K(Dnp)-AisoQ**K**aa-NH_2. Fragmentation occurring in the Ac-K(Dnp)-AisoQ**K**aa-NH_2 peptide (1) and the Abz-LPAT modification (2) from parent mass m/z = 1324.25. **K** = isopeptide ligation to Abz-LPAT.

1) Ions Produced From Ac-K(Dnp)-AisoQ**K**aa-NH_2 Peptide Chain					
Ion	Expected	Observed	Ion	Expected	Observed
b_1	337.11	N/O*	y_1	988.54	988.27
b_2	408.15	N/O*	y_2	917.51	917.25
b_3	536.21	535.14	y_3	789.45	N/O*
b_4	1165.56	1165.21	y_4	160.09	N/O*
b_5	1236.60	1236.23	y_5	89.06	N/O*
b_6	1307.64	1307.24	y_6	18.02	N/O*
2) Ions Produced From Abz-LPAT Modification					
Ion	Expected	Observed	Ion	Expected	Observed
b_7	120.04	N/O*	y_7	1205.61	N/O*
b_8	233.13	N/O*	y_8	1092.53	1092.21
b_9	330.18	N/O*	y_9	995.48	995.11
b_{10}	401.22	N/O*	y_{10}	924.44	924.18
b_{11}	502.27	502.13	y_{11}	823.39	823.17

*Not Observed

As shown in **Figures 23-24** and **Tables 3-4**, a multitude of y and b ions were observed that supported the formation of the desired isopeptide ligation product. In particular, ions produced from fragmentation at sites adjacent to the lysine residue provided the most insight. In the case of the Ac-K(Dnp)-GG**K**GG product, observation of the b_3/y_3 ions along with the b_4/y_4 ions produced from the peptide chain are key in validating isopeptide ligation. As represented by MS/MS spectra in **Figure 25**, the ions correspond to masses that are specific to fragments that could only be produced if ligation is occurring on the side chain of the lysine residue. See **Appendix IV-V** for the MS/MS spectra of all additional observed ions.

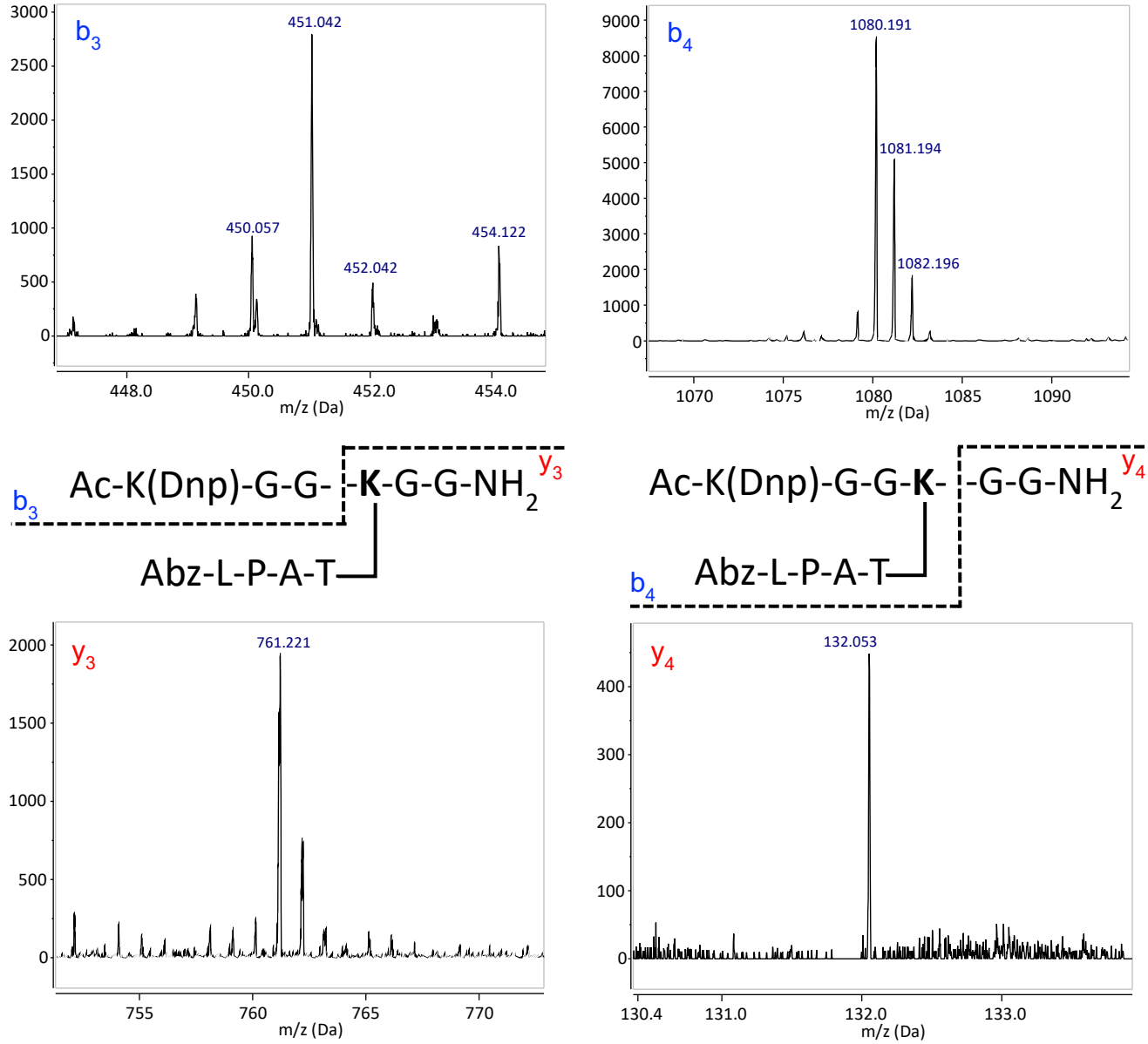


Figure 25. MS/MS data for significant y/b ions confirming isopeptide ligation in Ac-K(Dnp)-GGKGG-NH_2. Ion labels: b_3: Ac-K(Dnp)-GG (m/z: 451.16 calcd), y_3: **K**GG-NH_2 (m/z: 761.42 calcd), b_4: Ac-K(Dnp)-GG**K** (m/z: 1080.51 calcd), and y_4: GG-NH_2 (m/z: 132.06 calcd). **K** = isopeptide ligation to Abz-LPAT.

Similarly, the ions produced for the Ac-K(Dnp)-AisoQ**K**aa product, shown in **Figure 26**, aid in confirming isopeptide ligation. The b_3 ion corresponds to all residues N-terminal to the modified lysine and the adjacent b_4 ion increases by a mass that is specific to the addition of Abz-LPAT-**K**. Unfortunately, the corresponding y ions were not detected, however, the b ions gave

sufficient evidence to support isopeptide ligation occurring on the lysine residue. As shown in **Table 2**, multiple other ions were observed that were consistent with the desired isopeptide ligation and MS/MS spectra for these ions are represented in **Appendix VI-VII**.

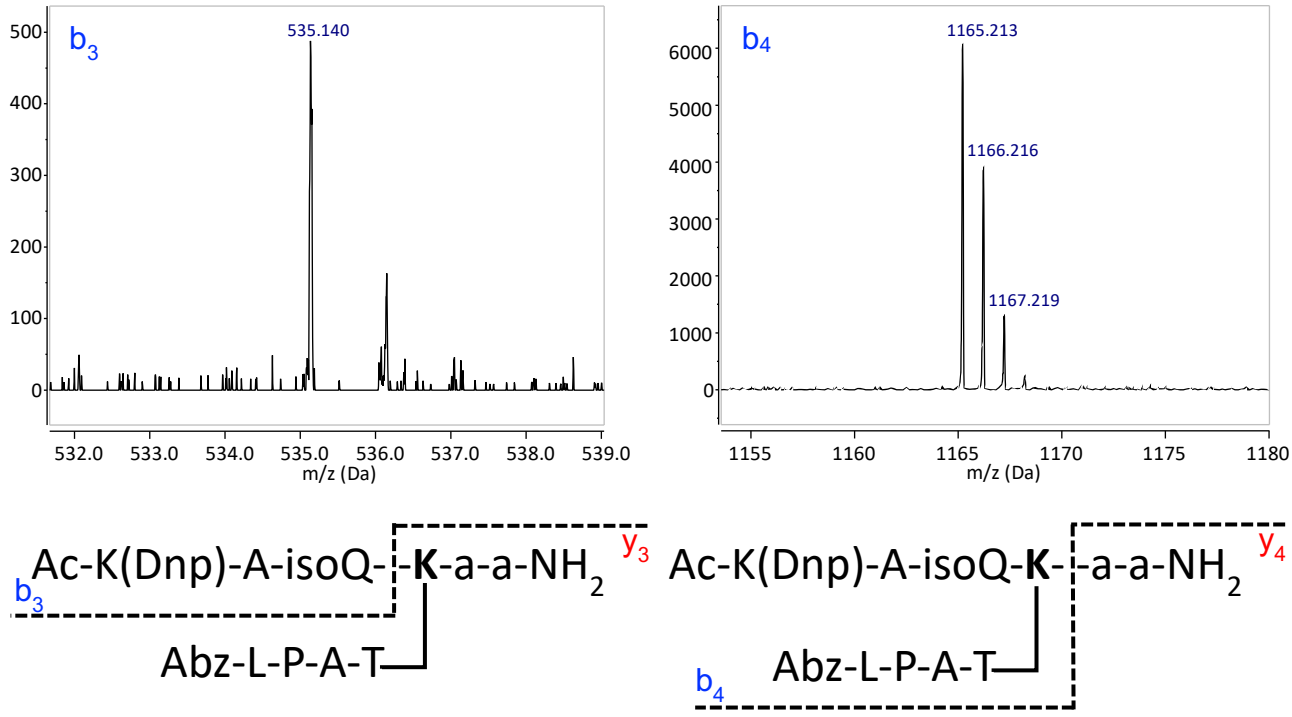


Figure 26. MS/MS data for significant b ions confirming isopeptide ligation in Ac-K(Dnp)-AisoQKaa-NH_2. Ion labels: b_3: Ac-K(Dnp)-AisoQ (m/z: 536.21 calcd), b_4: Ac-K(Dnp)-AisoQ**K** (m/z: 1165.56 calcd) **K** = isopeptide ligation to Abz-LPAT.

3.5 Addressing Production of Side Products

Tris/Glycerol ligation products. While characterizing the model peptide reaction components, a few trace side products were observed by RP-HPLC and ESI-MS. Amine-reactive reagents are not always entirely selective for amines.[3] Potentially, they can react with other nucleophiles that are present and accessible, especially with the depletion of the desired nucleophiles involved in the reaction.[3] For example, as depletion of amine nucleophiles occurs

over the course of sortase-catalyzed reactions, hydrolysis of the acyl-enzyme intermediate becomes inevitable in aqueous solution. Therefore, observation of a hydrolysis product in sortase-mediated chemistry is a common side-product and difficult to entirely avoid.[3] Interestingly, in our model SMIL reactions we also observed that Tris buffer and glycerol, which have potential nucleophilic sites, also contributed to the formation of non-desirable adducts (**Figure 27**).[3] In both cases, these species were identified based solely on molecular weight and therefore the precise connectivity between atoms was not determined. However, we propose that adducts have the structures shown in **Figure 27**.

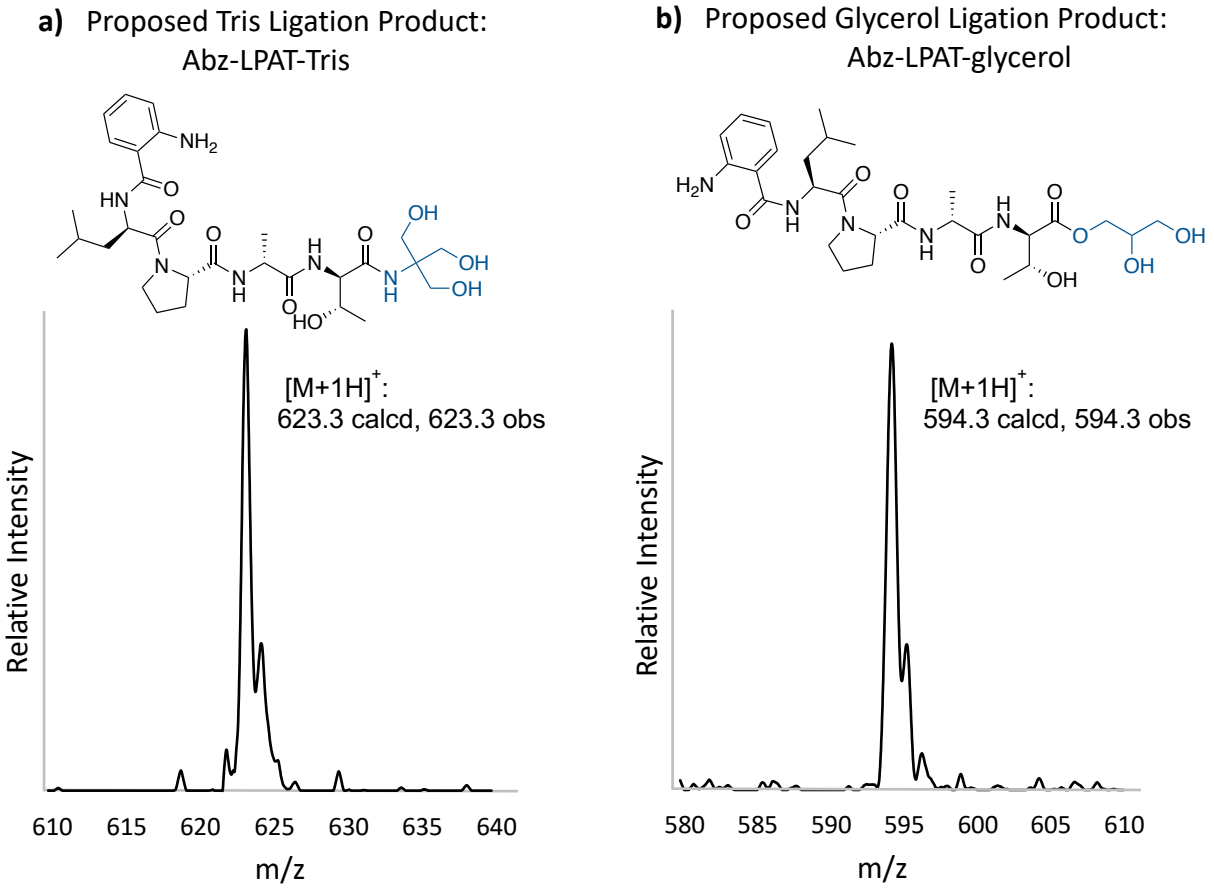


Figure 27. Representative mass spectra highlighting the proposed "Tris"(**a**) and "glycerol" (**b**) by-products.

Recognizing that these side products might complicate future SMIL studies, we modified reaction conditions to eliminate their formation. Specifically, Tris was replaced with phosphate buffered saline (PBS) (100 mM Na_3PO_4 (pH 7.5) and 150 mM NaCl). In addition, glycerol was removed from Δ79 $SrtA_{suis}$ stock solutions by a desalting purification method on the FPLC system using a mobile phase buffer consisting of PBS without glycerol. Importantly, both the replacement of Tris and the removal of glycerol were found to have no detrimental impact on isopeptide ligation (see section 3.6 below).

Excised Fragment Differences. Although generally selective for cleavage between the T and G residues of the LPXTG motif, when the substrate was extended by GH-K(Dnp) for Ni(II) coordination, we observed that Δ79 $SrtA_{suis}$ was capable of forming trace amounts of alternative cleavage products shown in **Figure 28** and reported in **Appendix II**. Specifically, Δ79 $SrtA_{suis}$ appeared to be capable of cleaving the peptide bond between the two glycine residues resulting in an excised fragment of GH-K(Dnp) and hydrolysis product of Abz-LPATG-OH. As illustrated in **Figure 28**, the amount of these alternate cleavage products was quite low in relation to the standard sortase cleavage site. In addition, we observed no evidence that species related to the alternate cleavage pathway were participating in subsequent isopeptide. Therefore, no additional steps were taken to minimize the formation of these side products.

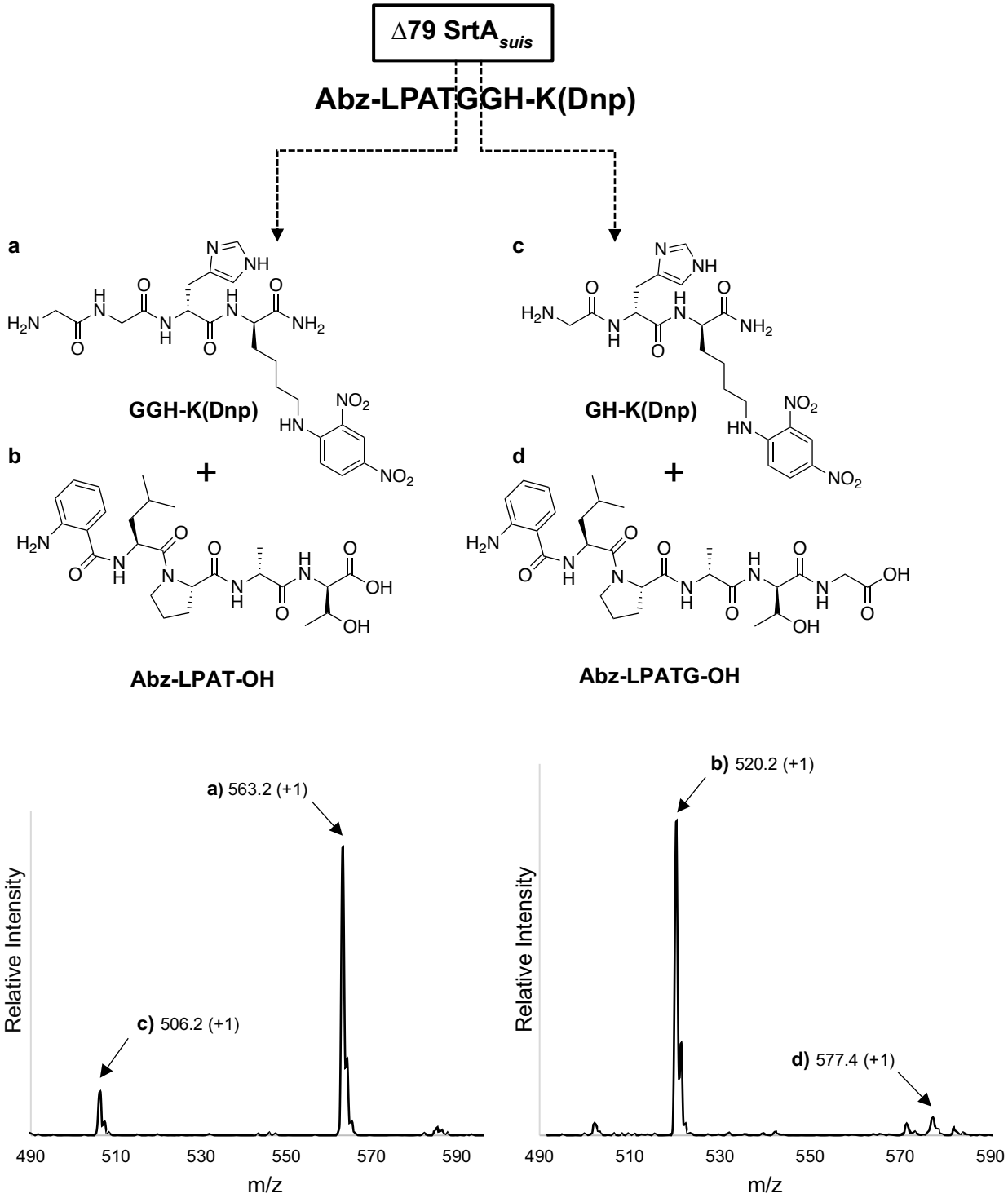


Figure 28. Representative mass spectra highlighting alternative excised fragments and hydrolysis products. Peak Labels: **a** = GGH-K(Dnp) excised fragment ($[M+1H]^+$: 563.22 calcd), **b** = Abz-LPAT-OH hydrolysis product ($[M+1H]^+$: 520.27 calcd), **c** = GH-K(Dnp) alternative excised fragment ($[M+1H]^+$: 506.2 calcd), **d** = Abz-LPATG-OH alternative hydrolysis product ($[M+1H]^+$: 577.29 calcd).

3.6 Larger Peptide Target Studies

Having demonstrated isopeptide ligation with small model peptides using Δ79 $SrtA_{suis}$, as well as optimizing the ligation efficiency with Ni^{2+} coordination, switching to PBS, and eliminating glycerol in enzyme stock solutions, we next extended our approach to the modification of exposed lysine residues in larger peptide targets. For this work, we employed a diethylamino coumarin-modified substrate (DEAC-Ahx-LPRTGGH-NH_2) in order to demonstrate that isopeptide ligation could be utilized to install a meaningful non-natural modification. In addition, the substrate possessed the masked GGH Ni^{2+}-binding motif to boost ligation efficiency. As described in Chapter 2, targets for this work included the peptides indolicidin, glucagon, exendin-4, and β-endorphin, which in all cases contained one or more native lysines.

Modification of Indolicidin. For modification of indolicidin, which contained a single internal lysine, we found that the use of 5 molar equivalents of DEAC-Ahx-LPRTGGH-NH_2 in the presence of Ni^{2+} and Δ79 $SrtA_{suis}$ resulted in successful modification of 90% of the starting target in only 1.5 h at room temperature. We note here that the estimation of reaction conversion was based on peak areas in the 280 nm chromatogram, which in the case of the modified species were corrected for the absorbance of the coumarin at 280 nm.

As shown in **Figure 29**, a number of peaks were observed in the RP-HPLC chromatogram of the crude reaction mixture, which were subsequently identified by ESI-MS analysis (**Appendix VIII**). These included unreacted DEAC-Ahx-LPRTGGH-NH_2 and indolicidin, a by-product resulting from substrate hydrolysis (DEAC-Ahx-LPRT-OH), a major singly-modified species, smaller amounts of a second singly-modified compound, and a doubly-modified compound. Given that

indolicidin contains one internal lysine residue, having a doubly-modified species suggested that in addition to the lysine side chain, ligation was occurring at another site. Based on the reactivity of most other characterized sortase A homologs, we hypothesized that the second modification site was at the N-terminal isoleucine.

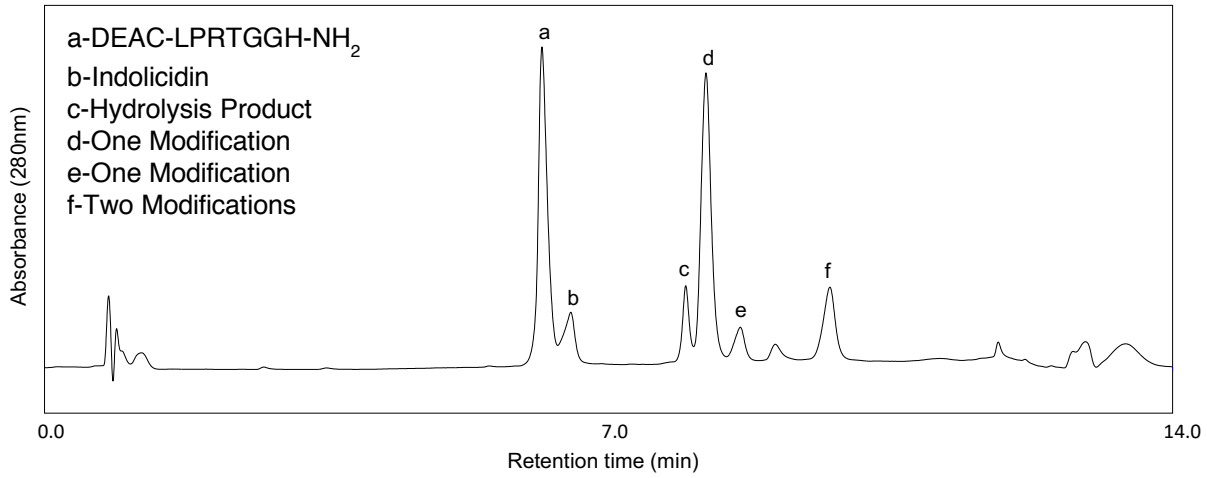


Figure 29. Isopeptide ligation involving Indolicidin. Peak Labels: **a**: DEAC-LPRTGGH-NH2, **b**: Indolicidin, **c**: hydrolysis product, **d**: one modification (Isopeptide) $[M+4H]^+$: 683.6 m/z calcd, 683.5 m/z obs, **e**: one modification (N-terminal) $[M+4H]^+$: 683.6 m/z calcd, 683.4 m/z obs, **f**: two modifications $[M+4H]^+$: 889.8 m/z calcd, 889.5 m/z obs. Reaction conditions: 60 μM DEAC-LPRTGGH-NH2, 20 μM Indolicidin, 5 μM Δ79 $SrtA_{suis}$, 100 mM Na_3PO_4 (pH 7.5), 150 mM NaCl, 120 μM $NiSO_4$, 1.04% (v/v) DMSO, 1.5 h at RT. Analyzed by RP-HPLC and ESI-MS (**Method G** (**Appendix III**), 280 nm). The extent of product formation was estimated from peak areas derived from the 425 nm and 280 nm chromatogram.

Given that multiple modified species were produced, we next wanted to confirm that the major singly-modified product was indeed the desired isopeptide ligation. In order to do so, the putative isopeptide product was first separated and isolated by RP-HPLC. The collected product then underwent chymotrypsin digestion, which was predicted to primarily cleave at the C-terminus of tryptophan, tyrosine and phenylalanine residues. In this case, cleavage occurred C-terminal to tryptophan residues present in the modified Indolicidin, and produced predictable

peptide fragments that were identified by ESI-MS. We were pleased to find two key fragments that were consistent with isopeptide ligation (**Figure 30**). Specifically, in addition to undigested, modified indolicidin (**Figure 30a**), we observed a fragment consisting of the first four residues containing no modification (**Figure 30b**). A final fragment of interest lacked the N-terminal segment of indolicidin, but included the internal lysine residue in modified form **(Figure 30c**). Overall, the observation of these fragments was consistent with isopeptide ligation being the major product of the reaction, as opposed to modification of the indolicidin N-terminus. Unsuccessful attempts were made to characterize the minor singly-modified species using a similar approach, which we attribute to the low quantities of this material that were produced.

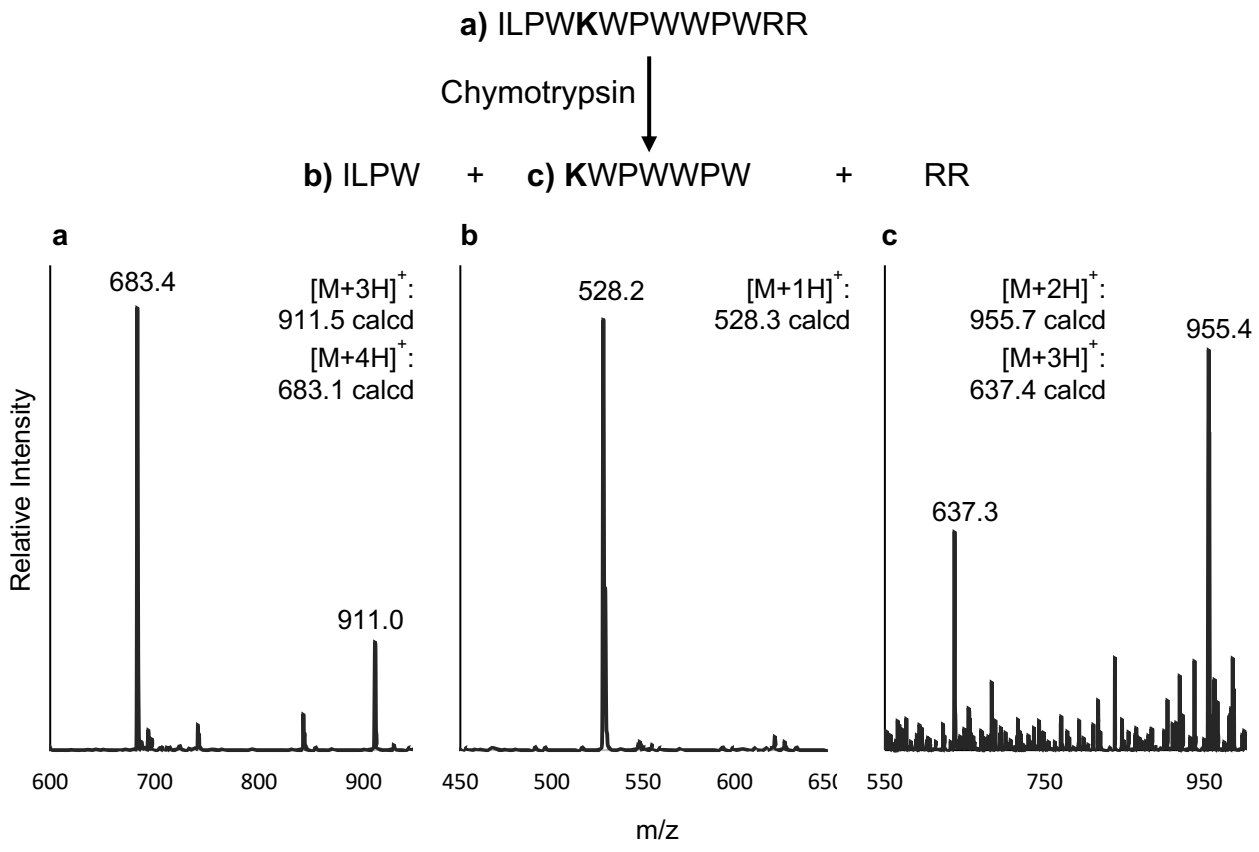


Figure 30. Mass spectra for fragments (**b,c**) produced during chymotrypsin digest of singly-modified indolicidin (**a**).

Modification of Glucagon. Building from our work with indolicidin, we next explored the modification of the target Glucagon. Like indolicidin, glucagon contains a single internal lysine and we found that 5 molar equivalents of DEAC-Ahx-LPRTGGH-NH_2 in the presence of Ni^{2+} and Δ79 $SrtA_{suis}$ provided excellent modification of the starting target, 98% after 5.75 h at room temperature. As before, the estimation of reaction conversion was determined by the peak areas at 280 nm in the chromatogram, with the modified species corrected for the absorbance of coumarin at 280 nm.

Glucagon contains a histidine at its N-terminus, which was one of the previously stated N-terminal residues that was unlikely to participate in SML. So as predicted, the reaction only generated one major singly-modified compound along with the substrate hydrolysis by-product (DEAC-Ahx-LPRT-OH). A representative RP-HPLC for the reaction is provided in **Figure 31**, and all peaks were identified by ESI-MS analysis (**Appendix VIII**). As in the case of indolicidin, we wanted to verify the site of covalent modification. Therefore, the singly-modified compound was separated and isolated by RP-HPLC. Next, it underwent proteolytic digestion using the serine protease Glu-C, which cleaves C-terminal to the aspartic acid residues present in glucagon. Shown in **Figure 32** are the generated fragments analyzed by ESI-MS. We observed two fragments characteristic of containing an isopeptide modification (**Figure 32a-b**). Notably, one of these fragments lacked the first nine and last fourteen residues of glucagon, yet retained an internal segment containing the lysine residue in modified form (**Figure 32b**). The last fragment contained the final fourteen residues containing no modification (**Figure 32c**). Overall, these fragments were consistent with isopeptide ligation. Furthermore, these data suggested that histidine at the N-terminus decreased undesired modification to the N-terminus.

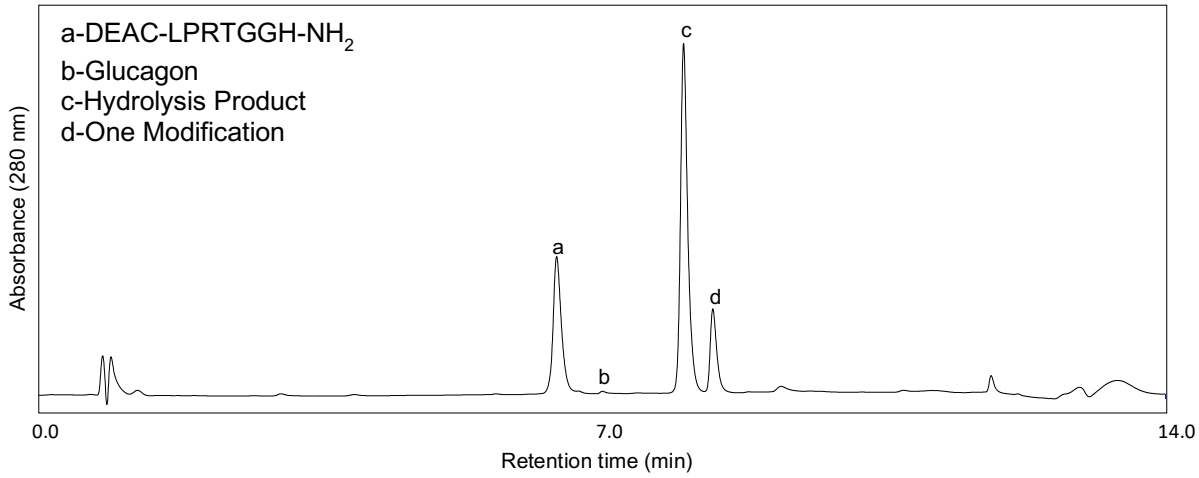


Figure 31. Isopeptide ligation of glucagon. Peak Labels: **a**: DEAC-LPRTGGH-NH_2, **b**: Glucagon, **c**: hydrolysis product, **d**: one modification (Isopeptide), $[M+3H]^+$: 1436.7 m/z calcd, 1436.3 m/z obs. Reaction conditions: 100 μM DEAC-LPRTGGH-NH2, 20 μM Glucagon, 5 μM Δ79 $SrtA_{suis}$, 100 mM Na_3PO_4 (pH 7.5), 150 mM NaCl, 200 μM $NiSO_4$, 1.04% (v/v) DMSO, 8 h at RT. Analyzed by RP-HPLC and ESI-MS (**Method G** (**Appendix III**), 280 nm). Extent of product formation was estimated from peak areas derived from the 425 nm and 280 nm chromatogram.

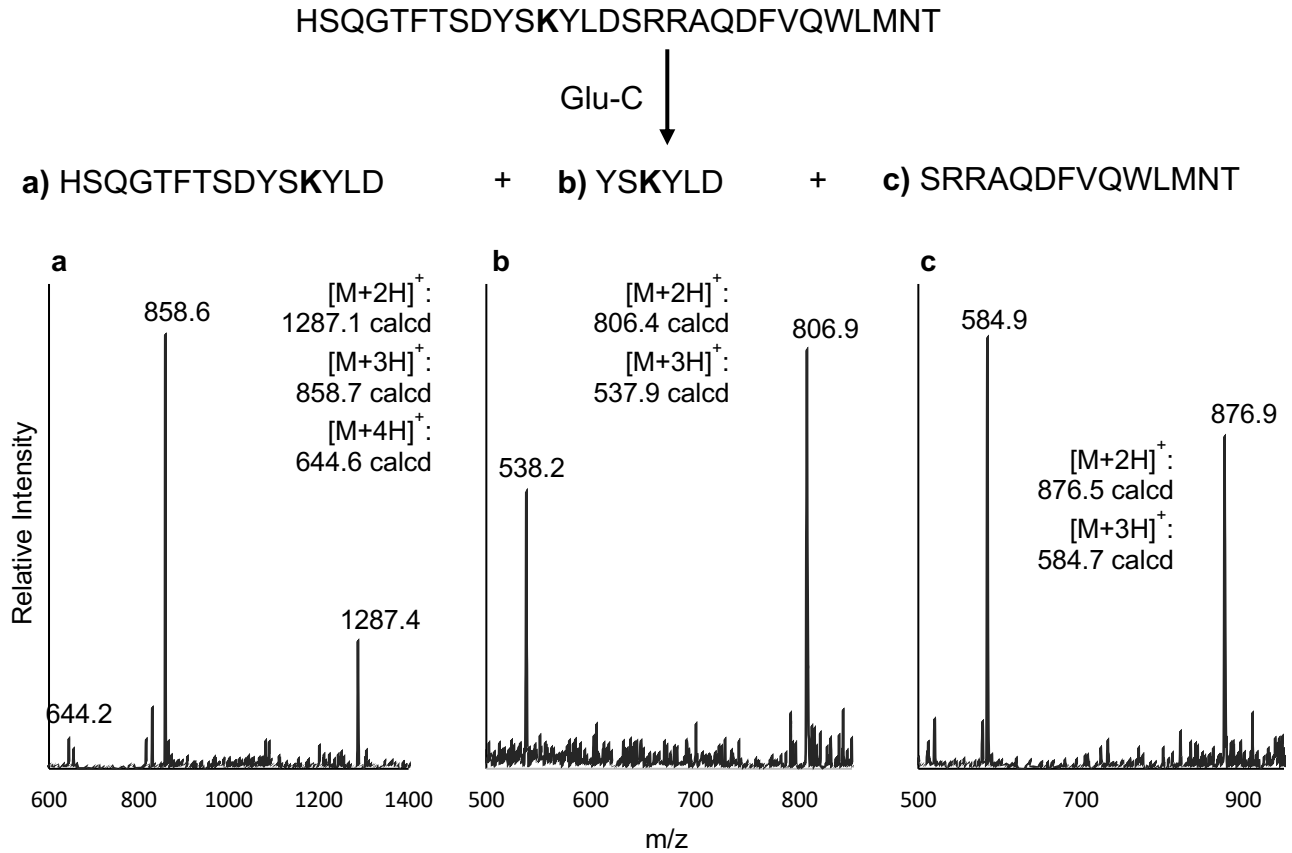


Figure 32. Mass spectra for fragments (**a**-**c**) produced during Glu-C digest of modified glucagon.

Modification of Exendin-4. Moving onto the modification of exendin-4, which also contains an N-terminal histidine similar to glucagon, it was predicted that the N-terminus would not be modified and instead modification would occur at its two internal lysine residues. In reaction with 5 molar equivalents of DEAC-Ahx-LPRTGGH-NH_2, Ni^{2+}, and Δ79 $SrtA_{suis}$, 90% of exendin-4 was modified with the generation of approximately 3:1 singly-modified species / double-modified species after 3.25 h at room temperature. The estimation of reaction conversion was determined in the same way as previous targets, using peaks derived from the 280 nm chromatogram and correcting the modified species for the absorbance of coumarin at 280 nm. Seen in the representative RP-HPLC (**Figure 33**) and confirmed by ESI-MS (**Appendix VIII**), the reaction generated a singly-modified species and a doubly-modified species along with the expected substrate hydrolysis by-product (DEAC-Ahx-LPRT-OH), and unreacted DEAC-LPRTGGH-NH_2 and exendin-4. Based on results obtained from glucagon, we speculated that the modified species represented isopeptide ligations at the lysine residues. However, given that only one major singly-modified species was observed, it suggested preferential modification of one lysine over the other. It was also possible that distinct singly-modified species were co-eluting under our RP-HPLC conditions. Further characterization of modified exendin-4 species was not completed, however it should be noted that enzymatic digest and MS or MS/MS analyses would be required for confirmation of modification at the lysine residues, and to assess whether specific lysines were preferred over others as modification sites.

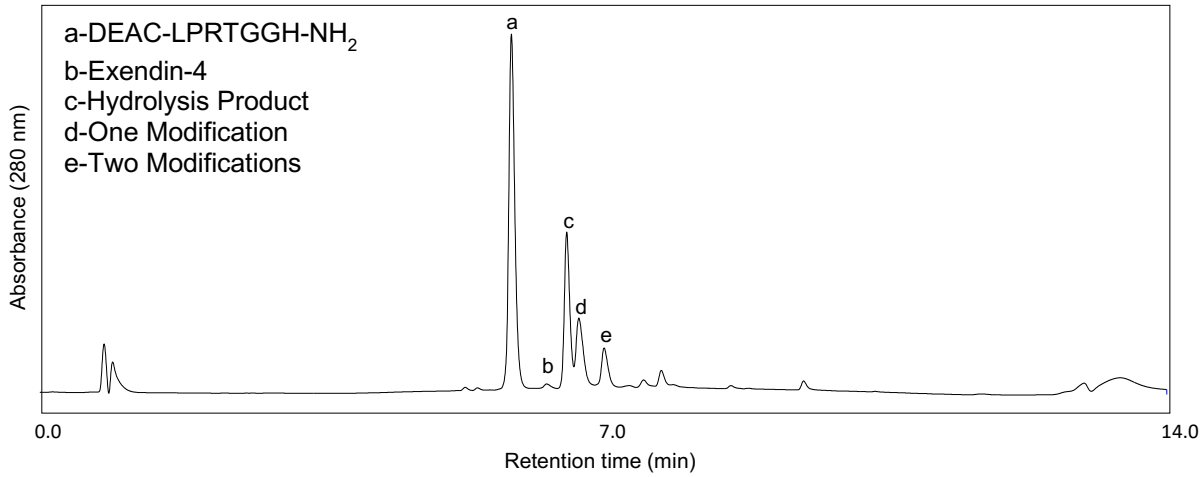


Figure 33. Isopeptide ligation of exendin-4. Peak Labels: **a**: DEAC-LPRTGGH-NH2, **b**: Exendin-4, **c**: hydrolysis product, **d**: one modification $[M+4H]^+$: 1253.63 m/z calcd, 1253.6 m/z obs, **e**: two modifications $[M+5H]^+$: 1168.4 m/z calcd, 1168 m/z obs. Reaction conditions: 100 μM DEAC-LPRTGGH-NH_2, 20 μM Exendin-4, 5 μM Δ79 $SrtA_{suis}$, 100 mM Na_3PO_4 (pH 7.5), 1500 mM NaCl, 200 μM $NiSO_4$, 1.04% (v/v) DMSO, 3.25 h at RT. Analyzed by RP-HPLC and ESI-MS (**Method E** (**Appendix III**), 280 nm).

Modification of β*-Endorphin.* As a final target we explored β-endorphin, which contains five internal lysine residues and an N-terminal tryptophan. We found that the use of 10 molar equivalents of DEAC-Ahx-LPRTGGH-NH_2 in the presence of Ni^{2+} and Δ79 $SrtA_{suis}$ converted 90% of the starting material in 10 h at room temperature as estimated by RP-HPLC. We were able to calculate this conversion by first observing that there is a linear decrease in peak area (280 nm) as the concentration of free β-endorphin decreases. An additional estimation was determined by peak areas of the two most abundant ions represented in LC-MS spectra, which indicated a 96% conversion of the starting material.

A range of peaks were observed in the RP-HPLC chromatogram (**Figure 34**) that were identified by LC-MS (**Appendix VIII**). These included the expected substrate hydrolysis by-product (DEAC-Ahx-LPRT-OH), unreacted DEAC-Ahx-LPRTGGH-NH_2 and β-endorphin, as well as many

ligation products characteristic of 1-6 modifications. Six modifications suggested that all five internal lysine residues as well as the N-terminus had been modified. Since an N-terminal tyrosine was previously demonstrated to be unreactive in SML using $SrtA_{staph,}$ our observations suggest that tyrosine may be able to serve as a nucleophile for the homolog $SrtA_{suis}$. However, of particular relevance to this work, the observation of multiple modifications provided strong evidence that the modification of lysine side chains was occurring at high levels with the β-endorphin target. No additional characterization of modified β-endorphin was performed.

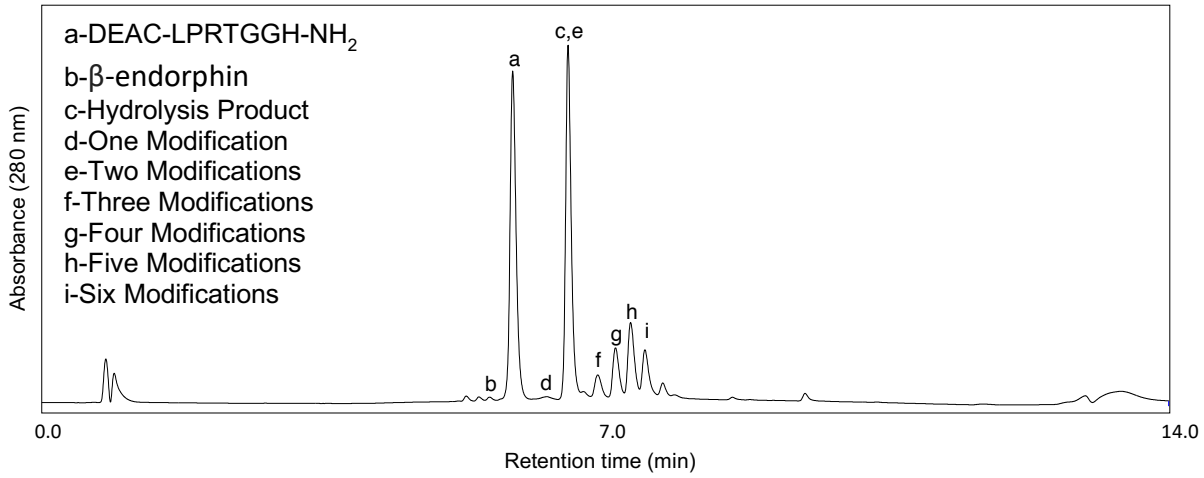


Figure 34. Isopeptide ligation involving β-endorphin. Peak Labels: **a**: DEAC-LPRTGGH-NH_2, **b**: β-endorphin, **c**: hydrolysis product, **d**: one modification $[M+4H]^+$: 1073.3 m/z calcd, 1073.5 m/z obs, **e**: two modifications $[M+5H]^+$: 1023.6 m/z calcd, 1023.6 m/z obs, **f**: three modifications $[M+4H]^+$: 1485.3 m/z calcd, 1485.0 m/z obs, **g**: four modifications $[M+5H]^+$: 1353.2 m/z calcd, 1353.2 m/z obs, **h**: five modifications $[M+5H]^+$: 1518.0 m/z calcd, 1518.0 m/z obs, **i**: six modifications $[M+6H]^+$: 1402.5 m/z calcd, 1402.4 m/z obs. Reaction conditions: 200 μM DEAC-LPRTGGH-NH_2, 20 μM β-endorphin, 5 μM Δ79 $SrtA_{suis}$, 100 mM Na_3PO_4 (pH 7.5), 150 mM NaCl, 400 μM $NiSO_4$, 2.08% (v/v) DMSO, 8 h at RT. Analyzed by RP-HPLC and ESI-MS (**Method E** (**Appendix III**), 280 nm).

4. Conclusions and Future Directions

The data in this thesis illustrate efficient isopeptide ligation catalyzed by sortase A from *Streptococcus suis* ($SrtA_{suis}$). Previous efforts have been met with modest success and were only achievable through the use of excess reagents or continue to be limited to N- or C-terminal locations. This work, in addition to previous studies performed by the Antos group, contributes to the development of a sortase based method for modification of lysine residues while potentially adding it to already existing modification strategies.

We have shown that $SrtA_{suis}$ is an efficient enzyme in catalyzing isopeptide ligation *in vitro* that is far superior than sortases typically used for protein modification. First we confirmed that two SrtA clones (Δ79 $SrtA_{suis}$ and Δ23 $SrtA_{suis}$) were comparable to $SrtA_{staph}$ at cleaving synthesized LPXTG peptides. Then we were able to specifically represent work using Δ79 $SrtA_{suis}$ after comparison with Δ23 $SrtA_{suis}$ in ligation reactions between peptides modeled after in vivo substrates and nucleophiles. Δ79 $SrtA_{suis}$ exhibited maximum reaction conversions of 18-21% after 2 h in contrast to 15-16% with Δ23 $SrtA_{suis}$ after a 16 h reaction. Therefore, we were able to determine that Δ79 $SrtA_{suis}$ gave the highest levels of product formation in significantly shorter reaction times.

Work was done that built on previous studies in our lab, involving the use of a Ni^{2+} coordination strategy as a way to improve isopeptide ligation efficiency by reducing reaction reversibility. We demonstrated that in contrast to reactions lacking Ni^{2+}, the systems employing metal additive resulted in a 2-3 fold increase in product formation and reached a maximum of 50% conversion after 5 h (Ac-K(Dnp)-GG**K**GG) and 65% conversion after 2 h (Ac-K(Dnp)-AisoQ**K**aa) at room temperature. These results were significantly greater than reactions using $SrtA_{staph}$,

which was only effective at producing ~5-7% of product after ~16 h at room temperature. This added supportive evidence that Δ79 SrtA$_{suis}$ is unique in its ability to generate isopeptide bonds *in vitro*.

After producing adequate amounts of ligation products we were able to confirm the site of conjugation by utilizing MS/MS. MS/MS analysis provided sufficient evidence by producing characteristic fragment ions that were consistent with the desired isopeptide ligation. The observed b_3/y_3 and b_4/y_4 ions were produced from fragmentation at sites adjacent to the modified lysine residue in both GG**K**GG and AisoQ**K**aa products. The difference between the b_4 and b_3 ions is ~629.5 m/z which corresponds to the calculated mass of a lysine residue ligated to the modification, Abz-LPAT-**K**, by an isopeptide bond (629.4 m/z calcd).

In the interest of reaction optimization, we showed that the production of trace amounts of side products involving Tris and glycerol could be eliminated by replacing Tris with PBS and removing glycerol from Δ79 SrtA$_{suis}$ stock solutions. We also found that this had no detrimental impacts on the production of isopeptide products and could be a way to eliminate any complications from these side products in future SMIL studies. We also identified trace amounts of alternative cleavage products stemming from the ability of Δ79 SrtA$_{suis}$ to cleave the peptide bond between the two glycine residues of the Abz-LPATGGH-K(Dnp) peptide producing an excised fragment of GH-K(Dnp) and hydrolysis product of Abz-LPATG-OH. We did not observe these species to participate in subsequent isopeptide production, so no additional steps were taken to minimize the formation of these side products.

As a final goal for this thesis, we successfully extended our approach to the modification of exposed lysine residues in the larger peptide targets: indolicidin, glucagon, exendin-4 and β-

endorphin. First, we modified Indolicidin by 90% after only 1.5 h at room temperature. Characterization by chymotrypsin digestion and ESI-MS analysis confirmed isopeptide ligation with the production of a fragment lacking the N-terminal segment of indolicidin, but included the internal lysine residue in modified form. In addition to these findings we propose that the presumed identity of the second minor singly-modified species produced corresponds to modification of the indolicidin N-terminal isoleucine. However, due to low quantities of this material, attempts at characterization were unsuccessful, but our observations suggest that isoleucine may be able to serve as a nucleophile for the homolog SrtA$_{suis}$.

Next, we modified glucagon with a reaction conversion of 98% after 5.75 h at room temperature where one major singly-modified compound was produced. After Glu-C digestion of the compound and ESI-MS analysis of a fragment that was lacking the first nine and last fourteen residues of glucagon, yet retaining an internal segment containing the lysine in modified form, it was characterized as the desired isopeptide product. With these results we were also able to show that having N-terminal histidine decreased undesired modification to the N-terminus. Also possessing an N-terminal histidine, 90% of Exendin-4 was modified with the generation of approximately 3:1 singly-modified species to doubly-modified species after 3.25 h at room temperature. Based on the results obtained from glucagon we speculated that the modified species represented isopeptide ligations at the two lysine residues exendin-4 contained. Given that only one major singly-modified species was observed we suggest that there is preferential modification of one lysine over the other or the possibility that a distinct singly-modified species was co-eluting under the RP-HPLC conditions. Enzymatic digest and MS or

MS/MS analyses would be required for confirmation of modification at the lysine residues as well as to determine whether a specific lysine was preferred.

Finally, β-endorphin was estimated at 96% conversion in ~10 h at room temperature and produced a range of products that were identified as ligation products characteristic of 1-6 modifications. With this target we presume that all five internal lysine residues have been modified, as well as the N-terminal tryptophan that may be able to serve as a nucleophile for the $SrtA_{suis}$ homolog. The observation of multiple modifications provide strong evidence of high levels of lysine side chain modification with β-endorphin.

The data presented in this thesis demonstrates effective improvements in isopeptide ligation methods. We showed that $SrtA_{suis}$ is compatible for modification of lysine residues present in small synthesized peptides as well as larger biological targets. With an interest in components that might influence reaction outcomes, a broad exploration of residues adjacent to the lysine nucleophile as well as a systematic study of selectivity between lysine modification versus N-terminal modification is a promising direction for future studies. It would also be beneficial to identify a lysine containing peptide motif that would give a desired product ratio in order to develop a method with precise control over number of ligated molecules. Long-term, we hope to apply this SMIL method to the modification of full-sized protein targets. Examples include, attaching ubiquitin or building ubiquitin oligomers linked together by isopeptide bond, and are known to regulate a number of biological processes such as degradation by the proteasome.[75] We would also like to apply SMIL to build potential therapeutics such as liraglutide which is a hormone treatment in diabetes management whose activity is dependent on a unique fatty acid modification linked together by isopeptide bond.[76] Our method may be valuable for

forming similar derivatives or used to make other protein therapeutics modified at lysine residues. Overall, this research has provided a basis for SMIL catalyzed by $SrtA_{suis}$ for future applications geared towards lysine modification.

5. Experimental

5.1 Instrumentation

Reversed-phase HPLC purifications and analyses were performed on a Dionex Ultimate 3000 HPLC system. All relevant chromatography columns and methods are listed in **Table 3**. These will hereafter be referred to by their corresponding method name (**A**-**H**). All methods employed the same mobile phase (MeCN, 0.1% (v/v) formic acid) / (95:5 H_2O/MeCN, 0.1% (v/v) formic acid) and used the following liquid chromatography (LC) columns: Luna 5u, C18 100Å column (250 x 10mm, C18 semi-prep, Phenomenex®); Kinetex 2.6 μm, C18 100Å LC column (100 x 2.1 mm, C18 analytical, Phenomenex®); Aeris 3.6 μm, XB-C8 200Å LC column (150 x 2.1 mm, C8 analytical, Phenomenex®); and Kinetex 2.6 μm, C4 100Å LC column (100 x 2.1 mm, C4 analytical, Phenomenex®).

For LC-MS analyses, the Dionex Ultimate 3000 HPLC was interfaced with an Advion CMS expression^L mass spectrometer. LC-MS data were analyzed using Advion Data Express software, and protein charge ladders were deconvoluted with a maximum-entropy algorithm provided by Analyst 1.4.2 software.

Desalting of protein solutions and size exclusion chromatography (SEC) were carried out on an NGC Quest 10 Plus FPLC system (Bio-Rad) equipped with a 10 mL Bio-Scale™ Mini Bio-Gel® P-6 Desalting Cartridge or an Enrich SEC 70 column (Bio-Rad). Additional details on mobile phase conditions and flow rates are provided in the appropriate sections below.

UV-Vis measurements for estimating the concentration of protein and peptide solutions were obtained using a Nanodrop ND-1000 spectrophotometer (ThermoFisher).

Imaging of SDS-PAGE gels stained with Coomassie Brilliant Blue R-250 was performed using a Gel Doc™ EZ Image (Bio-Rad).

Water used in all experimental procedures (npH_2O) was purified by a Milli-Q Advantage A10 system (Millipore).

Fluorescence measurements for assessing enzyme activity were obtained using a PTI QM-40 fluorimeter.

Table 3. Chromatography Methods for Purification and Analysis by HPLC or LC-ESI-MS

Method	Column	Gradient	Gradient	Flow Rate
Method (A)	C18 Semi Prep	10-90	10% (v/v) MeCN (0.0-2.0 min), 10% (v/v) MeCN → 90% (v/v) MeCN (2.0-15.0 min), hold 90% (v/v) MeCN (15.0-17.0 min), 90% (v/v) MeCN → 10% (v/v) MeCN (17.0-17.01 min), re-equilibrate at 10% (v/v) MeCN (17.01-19.0 min).	4.0 mL/min
Method (B)	C18 Semi Prep	20-90	20% (v/v) MeCN (0.0-2.0 min), 20% (v/v) MeCN → 90% (v/v) MeCN (2.0-15.0 min), hold 90% (v/v) MeCN (15.0-17.0 min), 90% (v/v) MeCN → 20% (v/v) MeCN (17.0-17.01 min), re-equilibrate at 20% (v/v) MeCN (17.01-19.0 min).	4.0 mL/min
Method (C)	C18 Analytical	10-90	10% (v/v) MeCN (0.0-0.5 min), 10% (v/v) MeCN → 90% (v/v) MeCN (0.5-7.0 min), hold 90% (v/v) MeCN (7.0-9.0 min), 90% (v/v) MeCN → 10% (v/v) MeCN (9.0-9.1 min), re-equilibrate at 10% (v/v) MeCN (9.1-12.0 min).	0.3 mL/min
Method (D)	C18 Analytical	5-90	5% (v/v) MeCN (0.0-2.0 min), 5% (v/v) MeCN → 90% (v/v) MeCN (2.0-10.0 min), hold 90% (v/v) MeCN (10.0-11.0 min), 90% (v/v) MeCN → 5% (v/v) MeCN (11.0-11.1 min), re-equilibrate at 5% (v/v) MeCN (11.1-13.25 min).	0.3 mL/min
Method (E)	C8 Analytical	5-90	5% (v/v) MeCN (0.0-2.0 min), 5% (v/v) MeCN → 90% (v/v) MeCN (2.0-10.0 min), hold 90% (v/v) MeCN (10.0-11.0 min), 90% (v/v) MeCN → 5% (v/v) MeCN (11.0-11.01 min), re-equilibrate at 5% (v/v) MeCN (11.01-14.3 min).	0.4 mL/min
Method (F)	C8 Analytical	5-40-90	5% (v/v) MeCN (0.0-2.0 min), 5% (v/v) MeCN → 40% (v/v) MeCN (2.0-9.0 min), 40% (v/v) MeCN → 90% (v/v) MeCN (9.0-10.0 min), hold 90% (v/v) MeCN (10.0-11.0 min), 90% (v/v) MeCN → 5% (v/v) MeCN (11.0-11.01 min), re-equilibrate at 10% (v/v) MeCN (11.01-14.3 min).	0.4 mL/min
Method (G)	C8 Analytical	10-4-90	10% (v/v) MeCN (0.0-2.0 min), 10% (v/v) MeCN → 40% (v/v) MeCN (2.0-9.0 min), 40% (v/v) MeCN → 90% (v/v) MeCN (9.0-10.0 min), hold 90% (v/v) MeCN (10.0-11.0 min), 90% (v/v) MeCN → 10% (v/v) MeCN (11.0-11.01 min), re-equilibrate at 10% (v/v) MeCN (11.01-14.3 min).	0.4 mL/min
Method (F)	C4 Analytical	10-90	10% (v/v) MeCN (0.0-0.5 min), 10% (v/v) MeCN → 90% (v/v) MeCN (0.5-7.0 min), hold 90% (v/v) MeCN (7.0-9.0 min), 90% (v/v) MeCN → 10% (v/v) MeCN (9.0-9.1 min), re-equilibrate at 10% (v/v) MeCN (9.1-12.0 min).	0.3 mL/min

5.2 Protein Expression and Purification, and Analysis of Larger Peptide Targets

Sequences of proteins used in this study:

Δ79 SrtA$_{suis}$

177 aa, MW: 19662.21 Da

MHHHHHHAILAAQWDAQRLPVIGGIAVPELGINLPIFKGVFNTSLMYGAGTMKENQEMGKGNYALASHHI

FGVTGAADVLFSPLDRAKNGMKIYITDKTNVYTYVIDSVEIVSPESVYVIDDVEGRTEVTLVTCTDYYATQRIVV

KGVLESTTPYNETAKDILDSFNKSYNQYDYGQ

Δ23 SrtA$_{suis}$

219 aa, MW: 24556.51 Da

MHHHHHHNFIIGWNTNKYQISNVTTEDIEKNKQAETTFDFEQVQSISTEAILAAQWDAQRLPVIGGIAVPEL

GINLPIFKGVFNTSLMYGAGTMKENQEMGKGNYALASHHIFGVTGAADVLFSPLDRAKNGMKIYITDKTNV

YTYVIDSVEIVSPESVYVIDDVEGRTEVTLVTCTDYYATQRIVVKGVLESTTPYNETAKDILDSFNKSYNQYDYG

Q

Indolicidin

13 aa, MW: 1906.32 Da

ILPWKWPWWPWRR - NH_2

Glucagon

29 aa, MW: 3482.79 Da

HSQGTFTSDYSKYLDSRRAQDFVQWLMNT

Exendin-4

38 aa, MW: 4186.62 Da

HGEGTFTSDLSKQMEEEAVRLFIEWLKNGGPSSGAPPPS - NH_2

β-endorphin
31 aa, MW: 3465.02 Da

YGGFMTSEKSQTPLVTLFKNAIIKNAYKKGE

Plasmid preparation. Expression vectors for Δ79 $SrtA_{suis}$ and Δ23 $SrtA_{suis}$ were obtained via commercial gene synthesis from ATUM. Both constructs were synthesized in the pD441-SR plasmid backbone, and included N-terminal His_6 tags for purification.

Expression and purification of sortase proteins. Both $SrtA_{suis}$ derivatives were generated using the same expression and purification protocol. Kanamycin resistant plasmids coding for the desired protein were transformed into BL21 (DE3) competent *E. coli* cells via heat shock. The cells were then plated onto lysogeny broth (LB) agar plates containing kanamycin at a final concentration of 50 μg/mL. The plates were incubated at 37 °C overnight and a single colony was chosen to inoculate 50 mL of sterile LB broth containing 50 μg/mL of kanamycin. These seed cultures were then incubated with shaking at 37 °C for about 18 hours. The 50 mL seed cultures were then added to 1 L of sterile LB broth with 50 μg/mL kanamycin and incubated with shaking at 37 °C. The cultures were allowed to grow until an OD_{600} of 0.8-0.9 was reached. At that point, IPTG was added to a final concentration of 1 mM to induce protein expression and the cultures were left to incubate with shaking at 37 °C for 3 hours. Cells were harvested by centrifugation at 4,629 x g for 10 min. Cell pellets were then frozen at -80 °C prior to further purification.

For purification, cell pellets were thawed on ice and resuspended in 30 mL of lysis buffer (20 mM Tris (pH 7.5), 150 mM NaCl, 0.5 mM EDTA). Once thawed, lysozyme was added (1 mg/mL final concentration) to aid in cell lysis and suspensions were incubated at 4 ^{o}C for 1 hour on a rocking platform. Additionally, cells were lysed via sonification for two 30 second intervals at 50% output on ice. The cell lysate was clarified by centrifugation at 20000 x g for 30 min, and the protein containing supernatant was collected. The supernatant was then applied to a nickel Ni-NTA column containing 5 mL of Ni-NTA agarose slurry (Thermo Scientific) that had been equilibrated in wash buffer (20 mM Tris (pH 7.5), 150 mM NaCl, 20 mM imidazole). After binding $SrtA_{\textit{suis}}$ to the Ni-NTA column, it was rinsed with 10 column volumes of wash buffer (50 mL of 20 mM Tris pH 7.5, 150 mM NaCl, and 20 mM imidazole), and the desired protein was then eluted with 10 mL of elution buffer (20 mM Tris pH 7.5, 150 mM NaCl and 300 mM imidazole).

The eluted fractions were desalted using a 10 mL Bio-Scale Mini Bio-Gel desalting cartridge (Bio-Rad) on an NGC Quest 10 Plus FPLC system (Bio-Rad) to remove imidazole. Desalting was achieved following the manufacturer's directions using a mobile phase buffer consisting of 50 mM Tris pH 7.5 and 150 mM NaCl at a flow rate of 2 mL/min. Glycerol was then added to the desalted protein solutions to a concentration of 10% (v/v), and aliquots were stored at -80 ^{o}C. Concentrations for all protein stock solutions were determined by absorbance at 280 nm using a Nanodrop^{TM} ND-1000 spectrophotometer (Thermo Scientific) and appropriate molar extinction coefficients listed in **Appendix I**. The molar extinction coefficients were estimated from the individual protein sequences using ExPASy ProtParam.

The purity of both $SrtA_{\textit{suis}}$ constructs was verified by SDS-PAGE, LC-MS analysis, and SEC. For LC-MS, analytical scale RP-HPLC separations upstream of the mass spectrometer were

achieved using a Kinetex 2.6 μm, C4 100 Å LC column (100 x 2.1 mm) from Phenomenex with chromatography **Method F** (see **Table 3**)**.** For SEC, analysis was performed using an Enrich SEC 70 column (Bio-Rad) with a mobile phase consisting of 50 mM Tris pH 7.5 and 150 mM NaCl at a flow rate of 2 mL/min.

For SMIL reactions utilizing PBS buffer, Tris-glycerol stocks of Δ79 $SrtA_{suis}$ were desalted using a 10 mL Bio-Scale Mini Bio-Gel desalting cartridge (Bio-Rad) on an NGC Quest 10 Plus FPLC system (Bio-Rad). Desalting was achieved following the manufacturer's directions using a mobile phase buffer consisting of PBS (100 mM sodium phosphate, 150 mM NaCl, pH 7.2) at a flow rate of 2 mL/min. Glycerol was not added to these stock solutions to eliminate by-product formation resulting from reactions with glycerol.

Acquisition and Characterization of Larger Peptide Targets. The peptide targets used for SMIL in this study (indolicidin, glucagon (1-29), exendin-4, and human β-endorphin) were all obtained from Anaspec with no further purification needed. Prior to use, these targets were reconstituted in npH_2O to a final concentration of 1 mM. The purity and identity of each target was confirmed by RP-HPLC and LC-MS (**Method E**, **Table 3**).

5.3 Peptide Synthesis

General. All model peptides were synthesized using manual Fmoc solid phase peptide synthesis (SPPS). Syntheses were performed in either glass or polypropylene synthesis vessels fitted with appropriate frits and inlet/outlet caps. All manipulations (washing, coupling, deprotection) were conducted at room temperature and included gentle agitation on a bench-top rocking platform

unless otherwise noted. A colorimetric ninhydrin test kit (Anaspec) was utilized to assess the success of SPPS coupling reactions. All materials, including standard Fmoc amino acids, Fmoc Rink Amide 4-Methylbenzhydrylamine hydrochloride (MBHA) resin, and reagents for coupling, deprotection, and resin cleavage were obtained from commercial sources and used without further purification. Incorporation of the 2,4-dinitrophenyl (Dnp) chromophore was achieved using a commercially available lysine building block (Fmoc-Lys(Dnp)-OH) purchased from ApexBio. Boc-2-aminobenzoic acid, D-Fmoc-Ala-OH, and Fmoc-6-Ahx-OH were obtained from Chem-Impex International. D-Fmoc-isoQ-OH was obtained from aapptec. DEAC-N-hydroxysuccinimide (NHS) for installation of the coumarin fluorophore was purchased from Sigma.

Synthesis of Abz-LPATGG-K(Dnp) and Abz-LPATGGH-K(Dnp). Solid phase synthesis vessels were loaded with Fmoc-protected Rink amide MBHA resin (0.2 mmol scale). The resin was washed/swollen with ~10 mL of N-Methyl-2-pyrrolidone (NMP) (3x, 10 min per wash). The removal of Fmoc was achieved by treatment with 10 mL of 80:20 NMP/piperidine at room temperature (2x, 10 min per treatment), followed by washing with ~10 mL of NMP (3x, 5-10 min per wash). Suitably protected amino acid building blocks were then coupled as follows: Fmoc-protected amino acids (0.60 mmol, 3.0 equivalents relative to resin loading), hexafluorophosphate benzotriazole tetramethyl uronium (HBTU) (0.60 mmol) and diisopropylethylamine (DIPEA) (1.0 mmol) were dissolved in 3.0 mL of NMP. The solution was mixed and then added to the deprotected resin. Couplings were incubated for 1-24 h at room temperature on a bench top rocking platform. The resins were then washed with ~10 mL of NMP (3x, 10 min per wash). The success of each coupling reaction was confirmed using a colorimetric

ninhydrin test and if coupling was incomplete, the amino acid coupling step was repeated. Repeated Fmoc deprotection and washing was then carried to assemble the desired sequence. Finally, an Abz fluorophore was installed at the N-terminus of each peptide through coupling of Boc-2-aminobenzoic acid following the same single amino acid coupling procedure described above.

Synthesis of Ac-K(Dnp)-GGKGG-NH_2 and Ac-K(Dnp)-AisoQKaa-NH_2. Synthesis of lysine-containing model nucleophiles was carried out as stated above. Following removal of the final Fmoc protecting group, the peptide N-termini were capped using acetic anhydride (Ac). Capping was achieved by treating the resins with a solution of acetic anhydride (472 μL, 5.0 mmol) and DIPEA (860 μL, 4.9 mmol) in 3.0 mL of NMP for at least 3 h at room temperature.

Cleavage of peptides from the resin. After assembly of the desired peptide sequences, the resins were washed with ~10 mL of NMP (3x, 10 min per wash) followed by 10 mL of dichloromethane (DCM) (3x, 10 min per wash). Peptides were cleaved from the resin by treatment with 5 mL of 95:2.5:2.5 trifluoroacetic acid (TFA)/ triisopropylsilane (TIPS)/H_2O (2x, ~30 min per treatment). The drained cleavage solutions were then concentrated on a rotary evaporator. The remaining peptide solutions were then precipitated by adding it drop-wise to 40 mL of dry-ice chilled diethyl ether. The precipitated peptides were collected by centrifugation at 4,696 x g for 5 min, and after discarding the supernatant, the crude peptide pellets were vacuum dried overnight. Crude peptides were next resolubilized in 1:1 DMSO/H_2O, and purification of the desired peptide was done using a Luna 5u, C18 100A column (250 x 10 mm) (semi-prep, Phenomenex) and RP-HPLC

(**Methods A-B, Table 3**). Pure fractions were pooled then cleared of formic acid and MeCN using a rotary evaporator. The remaining peptide solutions were then lyophilized. Lyophilized peptides were resolubilized in npH_2O / DMSO to give the compositions indicated in **Appendix I**. Concentrations were estimated using the absorbance of the Dnp chromophore at 365 nm (extinction coefficient: 17,300 $M^{-1}cm^{-1}$) on a Nanodrop™ ND-1000 spectrophotometer (Thermo Scientific). Concentrations were then adjusted to yield final stock solutions containing 1 mM total peptide. The purity and identity of all final peptide stock solutions were confirmed by RP-HPLC and LC-MS (**Methods C-E**, **Table 3**).

Synthesis of DEAC-Ahx-LPRTGGH-NH_2. SPPS was carried out as described above to assemble the core Ahx-LPRTGGH-NH_2 peptide. The peptide was cleaved from the resin as described above, and the crude peptide was then dissolved in NMP to a final concentration of 200 mM. Attachment of the coumarin fluorophore was then achieved by combining Ahx-LPRTGGH-NH_2 (50 mM final concentration), DEAC-NHS (50 mM final concentration), and DIPEA (100 mM final concentration) in NMP. The coupling reaction was incubated for ~1 hour at room temperature. DEAC-Ahx-LPRTGGH-NH_2 was then directly purified from the reaction mixture by RP-HPLC (**Method A, Table 3**). Pure fractions were pooled, rotary evaporated, and lyophilized. The lyophilized DEAC-Ahx-LPRTGGH-NH_2 was dissolved in 80 μL npH_2O and 60 μL DMSO. The concentration was estimated using the absorbance of the DEAC chromophore at 429 nm (extinction coefficient: 46,800 $M^{-1}cm^{-1}$) on a Nanodrop™ ND-1000 spectrophotometer (Thermo Scientific). Prior to use in SMIL reactions, the concentration was adjusted to yield a final stock solution of 1 mM total

peptide with the final composition indicated in **Appendix I**. The purity and identity of DEAC-Ahx-LPRTGGH-NH_2 was confirmed by RP-HPLC and LC-MS (**Method E, Table 3**).

5.4 General procedures for SMIL

Analysis of $SrtA_{suis}$ in vitro activity. The *in vitro* activity of both $SrtA_{suis}$ constructs (Δ79 and Δ23) was first assessed by fluorescence analysis and compared to the activity of $SrtA_{staph}$. Reactions were carried out on a 100 μL scale. Stock solutions of SrtA, Abz-LPATGG-K(Dnp) peptide substrate, and H_2NOH (stock solution in npH_2O) were combined in appropriate ratios to the desired reagent concentrations. Final reagents concentrations of 50 μM Abz-LPATGG-K(Dnp), 5 mM NH_2OH, and 20 μM SrtA were utilized. Reactions with $SrtA_{suis}$ also contained 10% (v/v) 10x reaction buffer lacking Ca^{2+} (500 mM Tris (pH 7.5) and 1.5 M NaCl), while reactions with $SrtA_{staph}$ contained 10% (v/v) 10x reaction buffer supplemented with Ca^{2+} (500 mM Tris (pH 7.5), 1.5 M NaCl, 10 mM Ca^{2+}). Reactions also contained residual glycerol (< 0.6% v/v) from the SrtA stock solutions, and residual DMSO (0.6% v/v) from the peptide substrate stock solution. All reactions were incubated for 2 hours at room temperature, and then quenched by the addition of 900 μL of 1:1 DMSO:H_2O. Fluorescence was measured with a PTI QM-40 fluorimeter with an excitation wavelength of 320 nm and an emission wavelength of 420 nm. The fluorescence for each reaction was reported relative to the background level of fluorescence for a control reaction in the absence of SrtA.

Complimentary to the fluorescence studies, the same reactions involving Δ79 $SrtA_{suis}$ and Δ23 $SrtA_{suis}$ were set up and monitored at 30 minute time intervals by RP-HPLC (**Method E, Table 3**). The extent of reaction conversion at each time point was estimated by comparing peak areas

for the relevant reaction species observed in the 360 nm chromatogram. The identity of all reaction components was also confirmed by LC-MS (**Method E, Table 3**).

Model Peptide Studies for SMIL. All model SMIL reactions with LPXTG substrates, model nucleophiles, $NiSO_4$, and SrtA followed the general reaction compositions stated in **Table 4**. Reactions were carried out on either a 100 μL or 200 μL scale, and stock solutions of relevant reaction components were combined to give the desired reagent concentrations. For all of these studies, $SrtA_{suis}$ stock solutions in 90:10 50 mM Tris pH 7.5 and 150 mM NaCl / glycerol were utilized. Stock solutions of $NiSO_4$ were prepared in npH_2O. Reactions with $SrtA_{suis}$ also contained 10% (v/v) 10x reaction buffer lacking Ca^{2+} (500 mM Tris (pH 7.5) and 1.5 M NaCl), while reactions with $SrtA_{staph}$ contained 10% (v/v) 10x reaction buffer supplemented with Ca^{2+} (500 mM Tris (pH 7.5), 1.5 M NaCl, 10 mM Ca^{2+}). Reactions also contained residual glycerol (< 0.6% v/v) from the SrtA stock solutions, and residual DMSO (< 3% v/v) from the peptide substrate stock solution. All reactions were incubated at room temperature.

Table 4. Reaction composition for model peptide SMIL reactions. **S**= Abz-LPATGG-K(Dnp) or Abz-LPATGGH-K(Dnp), **N**= Ac-K(Dnp)-GGKGG-NH_2 or Ac-K(Dnp)-AisoQKaa-NH_2.

Reagent	Final Concentration
Substrate (**S**)	100 μM
Nucleophile (**N**)	50 μM
Δ79 $SrtA_{suis}$, Δ23 $SrtA_{suis}$, Or $SrtA_{staph}$	5 μM
$NiSO_4$	0 / 200 μM

In general, reactions were monitored every 30 min over the course of 16 hours via RP-HPLC (**Method E, Table 3**), with the exception of reactions involving the ligation of Abz-LPATGGH-K(Dnp) with Ac-K(Dnp)-AisoQKaa-NH_2 which utilized **Method D, Table 3**. Product conversion was estimated from peak areas derived from the 360 nm chromatogram. Upon completion of the reaction time course, the reaction components were analyzed via LC-MS, using the previously stated method and a mass range of 300-1800 Da.

MS/MS characterization. To confirm the site of ligation on the side chain of the lysine residue in model peptide substrates, MS/MS analysis was performed. Putative isopeptide products from model peptide studies were collected by RP-HPLC using **Method E** for the Ac-K(Dnp)-GG**K**GG product and **Method D** for the Ac-K(Dnp)-AisoQ**K**aa product. The collected products were lyophilized, then reconstituted in 1:1 H_2O/MeOH containing 1% (v/v) formic acid to a final concentration of ~20 μM. In collaboration with Prof. Megan Gessel, samples were subjected to MS/MS analysis with a Waters Xevo G2-XS QTOF mass spectrometer, and fragment ions were observed with a mass range of 50-1500 Da. Chosen parent masses and collision energies for each isopeptide product are depicted in **Table 5.** Predicted fragments for both products, specifically b/y ions, were determined using ChemDraw and compared to the observed fragments in order to confirm isopeptide ligation.

Table 5. Parent mass and collision energy for MS/MS analysis of isopeptide products. **K** = isopeptide ligation to Abz-LPAT.

Product	Parent Mass	Collision Energy
Ac-K(Dnp)-GG**K**GG-NH_2	1211.56	37
Ac-K(Dnp)-AisoQ**K**aa-NH_2	1324.25	35

SMIL with Larger Peptide Targets. For these studies, the DEAC-Ahx-LPRTGGH-NH_2 substrate was utilized in combination with the targets indolicidin, glucagon, exendin-4, and β-endorphin. Reactions were performed on a 200 μL scale, and stock solutions of relevant reaction components were combined to give the desired reagent concentrations represented in **Table 6**.

Table 6. Reaction compositions for SMIL reactions with larger peptide targets.

Reagent	indolicidin, glucagon, exendin-4	β-endorphin
Substrate (DEAC-Ahx-LPRTGGH-NH_2)	100 μM	200 μM
Nucleophile (Larger Peptide Target)	20 μM	20 μM
$NiSO_4$ (Ni^{2+})	200 μM	400 μM
Δ79 $SrtA_{suis}$	5 μM	5 μM

For all of these studies, Δ79 $SrtA_{suis}$ stock solutions in PBS were utilized. Stock solutions of $NiSO_4$ were prepared in npH_2O. All reactions also contained 10% (v/v) 10x PBS (100 mM Na_3PO_4 (pH 7.5) and 150 mM NaCl) and residual DMSO (1.04% for 100 μM and 2.08% for 200 μM of DEAC-Ahx-LPRTGGH-NH_2 stock solution). Reactions were incubated at room temperature and monitored at 30 min time intervals over the course of 12 hours by RP-HPLC (**Methods C-G, Table**

3). Upon completion, the reactions were also analyzed by LC-MS in order to identify reaction components.

For SMIL reaction involving indolicidin, glucagon, and exendin-4, the extent of reaction progress was estimated by comparing peak areas derived from the 280 nm RP-HPLC chromatogram. To account for the contribution from the DEAC fluorophore, a correction factor was determined by analyzing varying concentrations of free DEAC-LPRTGGH-NH_2 (100 μM, 80 μM, 60 μM, 40 μM, 20 μM, and 10 μM) by RP-HPLC. The peak area for DEAC-LPRTGGH-NH_2 at its lambda max (425 nm) was divided by the corresponding peak area at 280 nm at each concentration to estimate the value of the correction factor (**cf**), which was consistent at all concentrations tested (**Eq. 1**). With the correction factor, the peak area (**PA**) of each ligation product at 425 nm was divided by the **cf** to give another value termed **x** (**Eq. 2**). Next, **x** was subtracted from the peak area of the product at 280 nm, resulting in a corrected peak area value ($cPA_{product}280nm$) for the ligation products (**Eq. 3**). To finally calculate overall reaction conversion, the adjusted peak area for the ligation product ($cPA_{product}280nm$) was divided by the sum of the peak areas for the starting material and all ligation products to give the corrected % conversion (**Eq. 4**).

$$\frac{PA_{substrate}\ 425nm}{PA_{substrate}\ 280nm} = \text{correction factor} = cf \qquad \textbf{Eq. 1}$$

$$\frac{PA_{product}\ 425nm}{cf} = x \qquad \textbf{Eq. 2}$$

$$PA_{product}\ 280nm - x = corected\ PA_{product} = cPA_{product}280nm \qquad \textbf{Eq. 3}$$

$$\left(\frac{cPA_{product}\ 280nm}{cPA_{product}280nm + PA_{nucleophile}\ 280nm}\right) \times 100 = corrected\ \%\ conversion \qquad \textbf{Eq. 4}$$

In the case of the modification of β-endorphin, the extent of product formation was estimated by the peak areas (PA) of the two most abundant ions observed for all relevant reaction components in the LC-MS spectra. The relevant reaction components were unreacted β-endorphin, one-modified β-endorphin (1mod), two-modified β-endorphin (2mod), three-modified β-endorphin (3mod), four-modified β-endorphin (4mod), five-modified β-endorphin (5mod), and 6-modified β-endorphin (6mod). Calculation of reaction conversion was done by taking the sum of all PA divided by the sum of all PA plus the PA of unreacted β-endorphin (**Eq 5**).

$$\frac{\sum_{n=1mod}^{6mod} PA(n)}{PA_{unreacted\ \beta-\mathrm{endorphin}} + \sum_{n=1mod}^{6mod} PA(n)} \qquad \textbf{Eq. 5}$$

As an additional check, free β-endorphin was diluted in npH_2O to final concentrations of 20 μM, 10 μM, 5 μM, 2.5 μM, and 1.25 μM, and was then monitored on RP-HPLC at 280 nm (**Method E, Table 3**). A linear decrease in or PA was observed confirming that a decrease in β-endorphin concentration is directly proportional to a decrease in PA. After determining this, the overall reaction conversion was estimated by comparing the PA for unreacted β-endorphin at 0 h and 10 h using **Eq. 6**.

$$\frac{PA^{0h}_{unreacted\ \beta-\mathrm{endorphin}}}{\left(PA^{0h}_{unreacted\ \beta-\mathrm{endorphin}} - PA^{10h}_{unreacted\ \beta-\mathrm{endorphin}}\right)} \qquad \textbf{Eq. 6}$$

Characterization of Reaction Products. In the case of modified indolicidin, the presumed isopeptide product was separated and collected by RP-HPLC (**Method G**, **Table 3**). The collected product was lyophilized and then reconstituted in 44.5 μL H_2O, 5 μL Ca^{2+} reaction buffer (500 mM Tris (pH 7.5), 1.5 M NaCl, and 10 mM Ca^{2+}), and 0.5 μL of chymotrypsin (1 mM HCl, 2 mM $CaCl_2$,

0.5 μg/μL chymotrypsin) to undergo chymotrypsin digestion. For this digestion, sequencing grade chymotrypsin was obtained from VWR, and reconstituted in 1 mM HCl and 2 mM $CaCl_2$ at a final concentration of 0.5 μg/μL. After addition of chymotrypsin, the reaction was incubated at 37 ^oC, and LC-MS (300-1800 Da) was used to analyze the production of fragments approximately every ~60 minutes. The observed proteolytic fragments were compared to predicted fragments generated using ChemDraw.

For modified glucagon, the collected isopeptide product was lyophilized and then reconstituted in 44 μL H_2O, 5 μL PBS (100 mM Na_3PO_4 (pH 7.5) and 150 mM NaCl), and 1 μL of Glu-C (0.5 μg/μL Glu-C, 100 mM Na_3PO_4 (pH 7.5) and 150 mM NaCl). For this analysis, Glu-C was obtained from VWR and reconstituted in 20 μL of PBS (100 mM Na_3PO_4 (pH 7.5) and 150 mM NaCl) to a final concentration of 0.5 μg/mL prior to use. Following addition of Glu-C, the reaction was incubated at 37 ^oC and the proteolysis reaction was monitored by LC-MS at time intervals of ~60 minutes. As before, the observed proteolytic fragments were compared to predicted fragments generated using Chemdraw.

6. Literature Cited

(1) Carter, P. Site-Directed Mutagenesis. *Biochem. J.* **1986**, *237* (1), 1–7. https://doi.org/10.1042/bj2370001.

(2) Edelheit, O.; Hanukoglu, A.; Hanukoglu, I. Simple and Efficient Site-Directed Mutagenesis Using Two Single-Primer Reactions in Parallel to Generate Mutants for Protein Structure-Function Studies. *BMC Biotechnol.* **2009**, *9* (1), 61. https://doi.org/10.1186/1472-6750-9-61.

(3) Koniev, O.; Wagner, A. Developments and Recent Advancements in the Field of Endogenous Amino Acid Selective Bond Forming Reactions for Bioconjugation. *Chem Soc Rev* **2015**, *44* (15), 5495–5551. https://doi.org/10.1039/C5CS00048C.

(4) Behrens, C. R.; Liu, B. Methods for Site-Specific Drug Conjugation to Antibodies. *mAbs* **2014**, *6* (1), 46–53. https://doi.org/10.4161/mabs.26632.

(5) Popp, M. W.; Dougan, S. K.; Chuang, T.-Y.; Spooner, E.; Ploegh, H. L. Sortase-Catalyzed Transformations That Improve the Properties of Cytokines. *Proc. Natl. Acad. Sci.* **2011**, *108* (8), 3169–3174. https://doi.org/10.1073/pnas.1016863108.

(6) James, E. I.; Jenkins, L. D.; Murphy, A. R. Peptide-Thiophene Hybrids as Self-Assembling Conductive Hydrogels. *Macromol. Mater. Eng.* **2019**, 1900285. https://doi.org/10.1002/mame.201900285.

(7) Steen Redeker, E.; Ta, D. T.; Cortens, D.; Billen, B.; Guedens, W.; Adriaensens, P. Protein Engineering For Directed Immobilization. *Bioconjug. Chem.* **2013**, *24* (11), 1761–1777. https://doi.org/10.1021/bc4002823.

(8) Arora, L.; Narula, A. Gene Editing and Crop Improvement Using CRISPR-Cas9 System. *Front. Plant Sci.* **2017**, *8*, 1932. https://doi.org/10.3389/fpls.2017.01932.

(9) Long, J. D.; Lee, J.-M.; Aylward, E. H.; Gillis, T.; Mysore, J. S.; Abu Elneel, K.; Chao, M. J.; Paulsen, J. S.; MacDonald, M. E.; Gusella, J. F. Genetic Modification of Huntington Disease Acts Early in the Prediagnosis Phase. *Am. J. Hum. Genet.* **2018**, *103* (3), 349–357. https://doi.org/10.1016/j.ajhg.2018.07.017.

(10) Yamamoto, Y.; Nagasato, M.; Yoshida, T.; Aoki, K. Recent Advances in Genetic Modification of Adenovirus Vectors for Cancer Treatment. *Cancer Sci.* **2017**, *108* (5), 831–837. https://doi.org/10.1111/cas.13228.

(11) Gropp, M.; Itsykson, P.; Singer, O.; Ben-Hur, T.; Reinhartz, E.; Galun, E.; Reubinoff, B. E. Stable Genetic Modification of Human Embryonic Stem Cells by Lentiviral Vectors. *Mol. Ther.* **2003**, *7* (2), 281–287. https://doi.org/10.1016/S1525-0016(02)00047-3.

(12) Gantz, V. M.; Jasinskiene, N.; Tatarenkova, O.; Fazekas, A.; Macias, V. M.; Bier, E.; James, A. A. Highly Efficient Cas9-Mediated Gene Drive for Population Modification of the Malaria Vector Mosquito *Anopheles Stephensi*. *Proc. Natl. Acad. Sci.* **2015**, *112* (49), E6736–E6743. https://doi.org/10.1073/pnas.1521077112.

(13) Parmar, N.; Singh, K. H.; Sharma, D.; Singh, L.; Kumar, P.; Nanjundan, J.; Khan, Y. J.; Chauhan, D. K.; Thakur, A. K. Genetic Engineering Strategies for Biotic and Abiotic Stress Tolerance and Quality Enhancement in Horticultural Crops: A Comprehensive Review. *3 Biotech* **2017**, *7* (4), 239. https://doi.org/10.1007/s13205-017-0870-y.

(14) Wang, L.; Xie, J.; Schultz, P. G. EXPANDING THE GENETIC CODE. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35* (1), 225–249. https://doi.org/10.1146/annurev.biophys.35.101105.121507.
(15) Mukai, T.; Lajoie, M. J.; Englert, M.; Söll, D. Rewriting the Genetic Code. *Annu. Rev. Microbiol.* **2017**, *71* (1), 557–577. https://doi.org/10.1146/annurev-micro-090816-093247.
(16) Dozier, J.; Distefano, M. Site-Specific PEGylation of Therapeutic Proteins. *Int. J. Mol. Sci.* **2015**, *16* (10), 25831–25864. https://doi.org/10.3390/ijms161025831.
(17) Boutureira, O.; Bernardes, G. J. L. Advances in Chemical Protein Modification. *Chem. Rev.* **2015**, *115* (5), 2174–2195. https://doi.org/10.1021/cr500399p.
(18) Ducry, L.; Stump, B. Antibody−Drug Conjugates: Linking Cytotoxic Payloads to Monoclonal Antibodies. *Bioconjug. Chem.* **2010**, *21* (1), 5–13. https://doi.org/10.1021/bc9002019.
(19) Rashidian, M.; Dozier, J. K.; Distefano, M. D. Enzymatic Labeling of Proteins: Techniques and Approaches. *Bioconjug. Chem.* **2013**, *24* (8), 1277–1294. https://doi.org/10.1021/bc400102w.
(20) Antos, J. M.; Truttmann, M. C.; Ploegh, H. L. Recent Advances in Sortase-Catalyzed Ligation Methodology. *Curr. Opin. Struct. Biol.* **2016**, *38*, 111–118. https://doi.org/10.1016/j.sbi.2016.05.021.
(21) Rabuka, D. Chemoenzymatic Methods for Site-Specific Protein Modification. *Curr. Opin. Chem. Biol.* **2010**, *14* (6), 790–796. https://doi.org/10.1016/j.cbpa.2010.09.020.
(22) Zhang, Y.; Park, K.-Y.; Suazo, K. F.; Distefano, M. D. Recent Progress in Enzymatic Protein Labelling Techniques and Their Applications. *Chem. Soc. Rev.* **2018**, *47* (24), 9106–9136. https://doi.org/10.1039/C8CS00537K.
(23) Rush, J. S.; Bertozzi, C. R. New Aldehyde Tag Sequences Identified by Screening Formylglycine Generating Enzymes *in Vitro* and *in Vivo*. *J. Am. Chem. Soc.* **2008**, *130* (37), 12240–12241. https://doi.org/10.1021/ja804530w.
(24) Tanaka, T.; Kamiya, N.; Nagamune, T. N-Terminal Glycine-Specific Protein Conjugation Catalyzed by Microbial Transglutaminase. *FEBS Lett.* **2005**, *579* (10), 2092–2096. https://doi.org/10.1016/j.febslet.2005.02.064.
(25) Mazmanian, S. K. Staphylococcus Aureus Sortase, an Enzyme That Anchors Surface Proteins to the Cell Wall. *Science* **1999**, *285* (5428), 760–763. https://doi.org/10.1126/science.285.5428.760.
(26) Spirig, T.; Weiner, E. M.; Clubb, R. T. Sortase Enzymes in Gram-Positive Bacteria: Sortase Enzymes in Gram-Positive Bacteria. *Mol. Microbiol.* **2011**, *82* (5), 1044–1059. https://doi.org/10.1111/j.1365-2958.2011.07887.x.
(27) Nobbs, A. H.; Lamont, R. J.; Jenkinson, H. F. Streptococcus Adherence and Colonization. *Microbiol. Mol. Biol. Rev.* **2009**, *73* (3), 407–450. https://doi.org/10.1128/MMBR.00014-09.
(28) Marraffini, L. A.; Schneewind, O. Anchor Structure of Staphylococcal Surface Proteins: V. ANCHOR STRUCTURE OF THE SORTASE B SUBSTRATE IsdC. *J. Biol. Chem.* **2005**, *280* (16), 16263–16271. https://doi.org/10.1074/jbc.M500071200.

(29) Dramsi, S.; Trieu-Cuot, P.; Bierne, H. Sorting Sortases: A Nomenclature Proposal for the Various Sortases of Gram-Positive Bacteria. *Res. Microbiol.* **2005**, *156* (3), 289–297. https://doi.org/10.1016/j.resmic.2004.10.011.
(30) Comfort, D.; Clubb, R. T. A Comparative Genome Analysis Identifies Distinct Sorting Pathways in Gram-Positive Bacteria. *Infect. Immun.* **2004**, *72* (5), 2710–2722. https://doi.org/10.1128/IAI.72.5.2710-2722.2004.
(31) Dasgupta, S.; Samantaray, S.; Sahal, D.; Roy, R. P. Isopeptide Ligation Catalyzed by Quintessential Sortase A: MECHANISTIC CUES FROM CYCLIC AND BRANCHED OLIGOMERS OF INDOLICIDIN. *J. Biol. Chem.* **2011**, *286* (27), 23996–24006. https://doi.org/10.1074/jbc.M111.247650.
(32) Maresso, A. W.; Chapa, T. J.; Schneewind, O. Surface Protein IsdC and Sortase B Are Required for Heme-Iron Scavenging of Bacillus Anthracis. *J. Bacteriol.* **2006**, *188* (23), 8145–8152. https://doi.org/10.1128/JB.01011-06.
(33) Kang, H. J.; Coulibaly, F.; Proft, T.; Baker, E. N. Crystal Structure of Spy0129, a Streptococcus Pyogenes Class B Sortase Involved in Pilus Assembly. *PLoS ONE* **2011**, *6* (1), e15969. https://doi.org/10.1371/journal.pone.0015969.
(34) Dramsi, S.; Caliot, E.; Bonne, I.; Guadagnini, S.; Prevost, M.-C.; Kojadinovic, M.; Lalioui, L.; Poyart, C.; Trieu-Cuot, P. Assembly and Role of Pili in Group B Streptococci. *Mol. Microbiol.* **2006**, *60* (6), 1401–1413. https://doi.org/10.1111/j.1365-2958.2006.05190.x.
(35) Hendrickx, A. P. A.; Budzik, J. M.; Oh, S.-Y.; Schneewind, O. Architects at the Bacterial Surface — Sortases and the Assembly of Pili with Isopeptide Bonds. *Nat. Rev. Microbiol.* **2011**, *9* (3), 166–176. https://doi.org/10.1038/nrmicro2520.
(36) Marraffini, L. A.; Schneewind, O. Targeting Proteins to the Cell Wall of Sporulating Bacillus Anthracis. *Mol. Microbiol.* **2006**, *62* (5), 1402–1417. https://doi.org/10.1111/j.1365-2958.2006.05469.x.
(37) Swaminathan, A.; Mandlik, A.; Swierczynski, A.; Gaspar, A.; Das, A.; Ton-That, H. Housekeeping Sortase Facilitates the Cell Wall Anchoring of Pilus Polymers in Corynebacterium Diphtheriae. *Mol. Microbiol.* **2007**, *66* (4), 961–974. https://doi.org/10.1111/j.1365-2958.2007.05968.x.
(38) *Bergey's Manual of Systematic Bacteriology*, 2nd ed.; Boone, D. R., Castenholz, R. W., Garrity, G. M., Eds.; Springer: New York, 2001.
(39) Guimaraes, C. P.; Witte, M. D.; Theile, C. S.; Bozkurt, G.; Kundrat, L.; Blom, A. E. M.; Ploegh, H. L. Site-Specific C-Terminal and Internal Loop Labeling of Proteins Using Sortase-Mediated Reactions. *Nat. Protoc.* **2013**, *8* (9), 1787–1799. https://doi.org/10.1038/nprot.2013.101.
(40) Williamson, D. J.; Fascione, M. A.; Webb, M. E.; Turnbull, W. B. Efficient N-Terminal Labeling of Proteins by Use of Sortase. *Angew. Chem. Int. Ed.* **2012**, *51* (37), 9377–9380. https://doi.org/10.1002/anie.201204538.
(41) Pritz, S.; Wolf, Y.; Kraetke, O.; Klose, J.; Bienert, M.; Beyermann, M. Synthesis of Biologically Active Peptide Nucleic Acid−Peptide Conjugates by Sortase-Mediated Ligation. *J. Org. Chem.* **2007**, *72* (10), 3909–3912. https://doi.org/10.1021/jo062331l.
(42) Shi, J.; Kundrat, L.; Pishesha, N.; Bilate, A.; Theile, C.; Maruyama, T.; Dougan, S. K.; Ploegh, H. L.; Lodish, H. F. Engineered Red Blood Cells as Carriers for Systemic Delivery of a Wide

Array of Functional Probes. *Proc. Natl. Acad. Sci.* **2014**, *111* (28), 10131–10136. https://doi.org/10.1073/pnas.1409861111.
(43) Pasqual, G.; Chudnovskiy, A.; Tas, J. M. J.; Agudelo, M.; Schweitzer, L. D.; Cui, A.; Hacohen, N.; Victora, G. D. Monitoring T Cell–Dendritic Cell Interactions in Vivo by Intercellular Enzymatic Labelling. *Nature* **2018**, *553* (7689), 496–500. https://doi.org/10.1038/nature25442.
(44) Park, K.; Jung, J.; Son, J.; Kim, S. H.; Chung, B. H. Anchoring Foreign Substances on Live Cell Surfaces Using Sortase A Specific Binding Peptide. *Chem. Commun.* **2013**, *49* (83), 9585. https://doi.org/10.1039/c3cc44753g.
(45) Leung, M. K. M.; Hagemeyer, C. E.; Johnston, A. P. R.; Gonzales, C.; Kamphuis, M. M. J.; Ardipradja, K.; Such, G. K.; Peter, K.; Caruso, F. Bio-Click Chemistry: Enzymatic Functionalization of PEGylated Capsules for Targeting Applications. *Angew. Chem. Int. Ed.* **2012**, *51* (29), 7132–7136. https://doi.org/10.1002/anie.201203612.
(46) Li, Z.; Theile, C. S.; Chen, G.-Y.; Bilate, A. M.; Duarte, J. N.; Avalos, A. M.; Fang, T.; Barberena, R.; Sato, S.; Ploegh, H. L. Fluorophore-Conjugated Holliday Junctions for Generating Super-Bright Antibodies and Antibody Fragments. *Angew. Chem. Int. Ed.* **2015**, *54* (40), 11706–11710. https://doi.org/10.1002/anie.201505277.
(47) Galimidi, R. P.; Klein, J. S.; Politzer, M. S.; Bai, S.; Seaman, M. S.; Nussenzweig, M. C.; West, A. P.; Bjorkman, P. J. Intra-Spike Crosslinking Overcomes Antibody Evasion by HIV-1. *Cell* **2015**, *160* (3), 433–446. https://doi.org/10.1016/j.cell.2015.01.016.
(48) Krueger, A. T.; Kroll, C.; Sanchez, E.; Griffith, L. G.; Imperiali, B. Tailoring Chimeric Ligands for Studying and Biasing ErbB Receptor Family Interactions. *Angew. Chem. Int. Ed.* **2014**, *53* (10), 2662–2666. https://doi.org/10.1002/anie.201307869.
(49) Beerli, R. R.; Hell, T.; Merkel, A. S.; Grawunder, U. Sortase Enzyme-Mediated Generation of Site-Specifically Conjugated Antibody Drug Conjugates with High In Vitro and In Vivo Potency. *PLOS ONE* **2015**, *10* (7), e0131177. https://doi.org/10.1371/journal.pone.0131177.
(50) Pierce, N. W.; Lee, J. E.; Liu, X.; Sweredoski, M. J.; Graham, R. L. J.; Larimore, E. A.; Rome, M.; Zheng, N.; Clurman, B. E.; Hess, S.; et al. Cand1 Promotes Assembly of New SCF Complexes through Dynamic Exchange of F Box Proteins. *Cell* **2013**, *153* (1), 206–215. https://doi.org/10.1016/j.cell.2013.02.024.
(51) Zhang, J.; Wang, M.; Tang, R.; Liu, Y.; Lei, C.; Huang, Y.; Nie, Z.; Yao, S. Transpeptidation-Mediated Assembly of Tripartite Split Green Fluorescent Protein for Label-Free Assay of Sortase Activity. *Anal. Chem.* **2018**, *90* (5), 3245–3252. https://doi.org/10.1021/acs.analchem.7b04756.
(52) Schoonen, L.; Pille, J.; Borrmann, A.; Nolte, R. J. M.; van Hest, J. C. M. Sortase A-Mediated N-Terminal Modification of Cowpea Chlorotic Mottle Virus for Highly Efficient Cargo Loading. *Bioconjug. Chem.* **2015**, *26* (12), 2429–2434. https://doi.org/10.1021/acs.bioconjchem.5b00485.
(53) Koudelka, K. J.; Pitek, A. S.; Manchester, M.; Steinmetz, N. F. Virus-Based Nanoparticles as Versatile Nanomachines. *Annu. Rev. Virol.* **2015**, *2* (1), 379–401. https://doi.org/10.1146/annurev-virology-100114-055141.
(54) Barwal, I.; Kumar, R.; Kateriya, S.; Dinda, A. K.; Yadav, S. C. Targeted Delivery System for Cancer Cells Consist of Multiple Ligands Conjugated Genetically Modified CCMV Capsid

on Doxorubicin GNPs Complex. *Sci. Rep.* **2016**, *6* (1), 37096. https://doi.org/10.1038/srep37096.
(55) Chen, I.; Dorr, B. M.; Liu, D. R. A General Strategy for the Evolution of Bond-Forming Enzymes Using Yeast Display. *Proc. Natl. Acad. Sci.* **2011**, *108* (28), 11399–11404. https://doi.org/10.1073/pnas.1101046108.
(56) Hirakawa, H.; Ishikawa, S.; Nagamune, T. Design of Ca $^{2+}$ -Independent *Staphylococcus Aureus* Sortase A Mutants. *Biotechnol. Bioeng.* **2012**, *109* (12), 2955–2961. https://doi.org/10.1002/bit.24585.
(57) Jeong, H.-J.; Abhiraman, G. C.; Story, C. M.; Ingram, J. R.; Dougan, S. K. Generation of Ca2+-Independent Sortase A Mutants with Enhanced Activity for Protein and Cell Surface Labeling. *PLOS ONE* **2017**, *12* (12), e0189068. https://doi.org/10.1371/journal.pone.0189068.
(58) Dorr, B. M.; Ham, H. O.; An, C.; Chaikof, E. L.; Liu, D. R. Reprogramming the Specificity of Sortase Enzymes. *Proc. Natl. Acad. Sci.* **2014**, *111* (37), 13343–13348. https://doi.org/10.1073/pnas.1411179111.
(59) Piotukh, K.; Geltinger, B.; Heinrich, N.; Gerth, F.; Beyermann, M.; Freund, C.; Schwarzer, D. Directed Evolution of Sortase A Mutants with Altered Substrate Selectivity Profiles. *J. Am. Chem. Soc.* **2011**, *133* (44), 17536–17539. https://doi.org/10.1021/ja205630g.
(60) David Row, R.; Roark, T. J.; Philip, M. C.; Perkins, L. L.; Antos, J. M. Enhancing the Efficiency of Sortase–Mediated Ligations through Nickel–Peptide Complex Formation. *Chem. Commun.* **2015**, *51* (63), 12548–12551. https://doi.org/10.1039/C5CC04657B.
(61) Yamamura, Y.; Hirakawa, H.; Yamaguchi, S.; Nagamune, T. Enhancement of Sortase A-Mediated Protein Ligation by Inducing a β-Hairpin Structure around the Ligation Site. *Chem. Commun.* **2011**, *47* (16), 4742. https://doi.org/10.1039/c0cc05334a.
(62) Liu, F.; Luo, E. Y.; Flora, D. B.; Mezo, A. R. Irreversible Sortase A-Mediated Ligation Driven by Diketopiperazine Formation. *J. Org. Chem.* **2014**, *79* (2), 487–492. https://doi.org/10.1021/jo4024914.
(63) Möhlmann, S.; Mahlert, C.; Greven, S.; Scholz, P.; Harrenga, A. In Vitro Sortagging of an Antibody Fab Fragment: Overcoming Unproductive Reactions of Sortase with Water and Lysine Side Chains. *ChemBioChem* **2011**, *12* (11), 1774–1780. https://doi.org/10.1002/cbic.201100002.
(64) Bellucci, J. J.; Bhattacharyya, J.; Chilkoti, A. A Noncanonical Function of Sortase Enables Site-Specific Conjugation of Small Molecules to Lysine Residues in Proteins. *Angew. Chem. Int. Ed.* **2014**, n/a-n/a. https://doi.org/10.1002/anie.201408126.
(65) McConnell, S. A.; Amer, B. R.; Muroski, J.; Fu, J.; Chang, C.; Ogorzalek Loo, R. R.; Loo, J. A.; Osipiuk, J.; Ton-That, H.; Clubb, R. T. Protein Labeling via a Specific Lysine-Isopeptide Bond Using the Pilin Polymerizing Sortase from *Corynebacterium Diphtheriae*. *J. Am. Chem. Soc.* **2018**, *140* (27), 8420–8423. https://doi.org/10.1021/jacs.8b05200.
(66) Kilpper-Balz, R.; Schleifer, K. H. Streptococcus Suis Sp. Nov., Nom. Rev. *Int. J. Syst. Bacteriol.* **1987**, *37* (2), 160–162. https://doi.org/10.1099/00207713-37-2-160.
(67) Nikghalb, K. D.; Horvath, N. M.; Prelesnik, J. L.; Banks, O. G. B.; Filipov, P. A.; Row, R. D.; Roark, T. J.; Antos, J. M. Expanding the Scope of Sortase-Mediated Ligations by Using Sortase Homologues. *ChemBioChem* **2018**, *19* (2), 185–195. https://doi.org/10.1002/cbic.201700517.

(68) Schleifer, K. H.; Kandler, O. Peptidoglycan Types of Bacterial Cell Walls and Their Taxonomic Implications. *Bacteriol. Rev.* **1972**, *36* (4), 407–477.
(69) Poreba, M.; Szalek, A.; Rut, W.; Kasperkiewicz, P.; Rutkowska-Wlodarczyk, I.; Snipas, S. J.; Itoh, Y.; Turk, D.; Turk, B.; Overall, C. M.; et al. Highly Sensitive and Adaptable Fluorescence-Quenched Pair Discloses the Substrate Specificity Profiles in Diverse Protease Families. *Sci. Rep.* **2017**, *7* (1), 43135. https://doi.org/10.1038/srep43135.
(70) Poreba, M.; Drag, M. Current Strategies for Probing Substrate Specificity of Proteases. *Curr. Med. Chem.* **2012**, *17* (33), 3968–3995. https://doi.org/10.2174/092986710793205381.
(71) Rozek, A.; Friedrich, C. L.; Hancock, R. E. Structure of the Bovine Antimicrobial Peptide Indolicidin Bound to Dodecylphosphocholine and Sodium Dodecyl Sulfate Micelles. *Biochemistry* **2000**, *39* (51), 15765–15774.
(72) Ding, X.; Saxena, N. K.; Lin, S.; Gupta, N.; Anania, F. A. Exendin-4, a Glucagon-like Protein-1 (GLP-1) Receptor Agonist, Reverses Hepatic Steatosis Inob/Ob Mice. *Hepatology* **2006**, *43* (1), 173–181. https://doi.org/10.1002/hep.21006.
(73) Chabenne, J. R.; DiMarchi, M. A.; Gelfanov, V. M.; DiMarchi, R. D. Optimization of the Native Glucagon Sequence for Medicinal Purposes. *J. Diabetes Sci. Technol.* **2010**, *4* (6), 1322–1331. https://doi.org/10.1177/193229681000400605.
(74) Chaudhry, S. R.; Kum, B. Biochemistry, Endorphin. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2019.
(75) Pham, G. H.; Rana, A. S. J. B.; Korkmaz, E. N.; Trang, V. H.; Cui, Q.; Strieter, E. R. Comparison of Native and Non-Native Ubiquitin Oligomers Reveals Analogous Structures and Reactivities: Comparison of Native and Non-Native Ubiquitin Oligomers. *Protein Sci.* **2016**, *25* (2), 456–471.
(76) Wajcberg, E.; Amatur Amarah. Liraglutide in the Management of Type 2 Diabetes. *Drug Des. Devel. Ther.* **2010**, 279.

7. Appendix

7.1 Appendix I. Quantification Parameters and Stock Solution Composition for Reagents

	Lambda max	Molar Extinction coefficient (M^{-1} cm^{-1})	HPLC Method for Confirming Purity	Final Stock Solution Composition	Final Stock Concentration
Abz-LPATGG-K(Dnp)	365 nm	17,300	E	13% (v/v) DMSO, npH_2O	807.2 μM
Abz-LPATGGH-K(Dnp)	365 nm	17,300	E	10% (v/v) DMSO, npH_2O	1837.2 μM
DEAC-Ahx-LPRTGGH-NH_2	429 nm	46,800	C	10% (v/v) DMSO, npH_2O	952.0 μM
Ac-K(Dnp)-GGKGG-NH_2	365 nm	17,300	E	10% (v/v) DMSO, npH_2O	804.6 μM
Ac-K(Dnp)-AisoQKaa-NH_2	365 nm	17,300	D	2.5% (v/v) DMSO, npH_2O	1180.0 μM
Indolicidin	280 nm	27,500	E	npH_2O	1 mM
Glucagon	280 nm	8,480	E	npH_2O	1 mM
Exendin-4	280 nm	5,500	E	npH_2O	1 mM
β-endorphin	280 nm	2,980	E	npH_2O	1 mM
Δ79 $SrtA_{suis}$	280 nm	23,380	F	50 mM Tris (pH 7.5), 150 mM NaCl, and 10% (v/v) glycerol	436.3 μM
Δ23 $SrtA_{suis}$	280 nm	30,370	F	50 mM Tris (pH 7.5), 150 mM NaCl, and 10% (v/v) glycerol	85.9 μM
Δ79 $SrtA_{suis}$ In PBS	280 nm	23,380	F	100 mM sodium phosphate (pH 7.2), 150 mM NaCl	79.12 μM

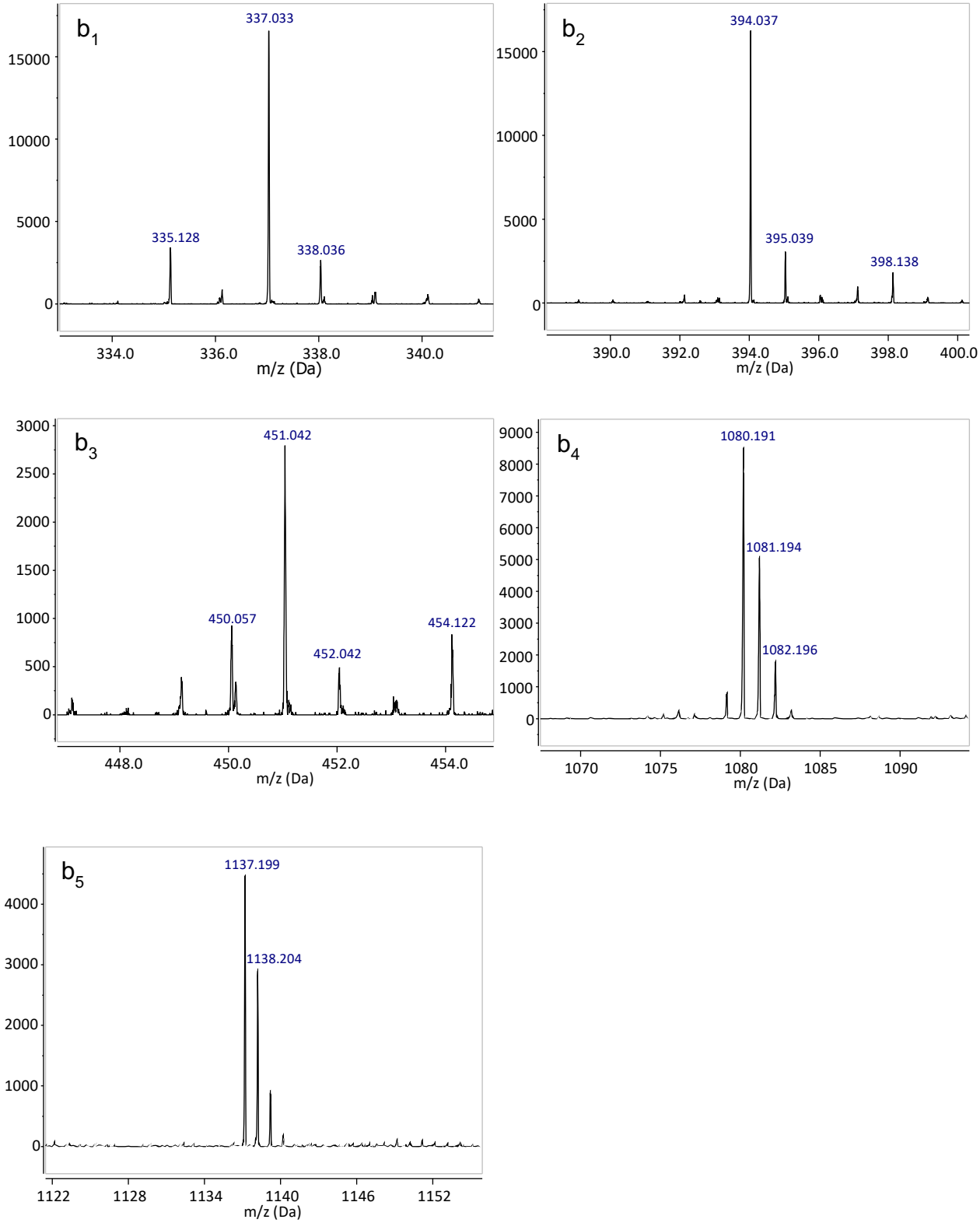
7.2 Appendix II. Table of Calculated and Observed Masses for All Model Peptide Studies

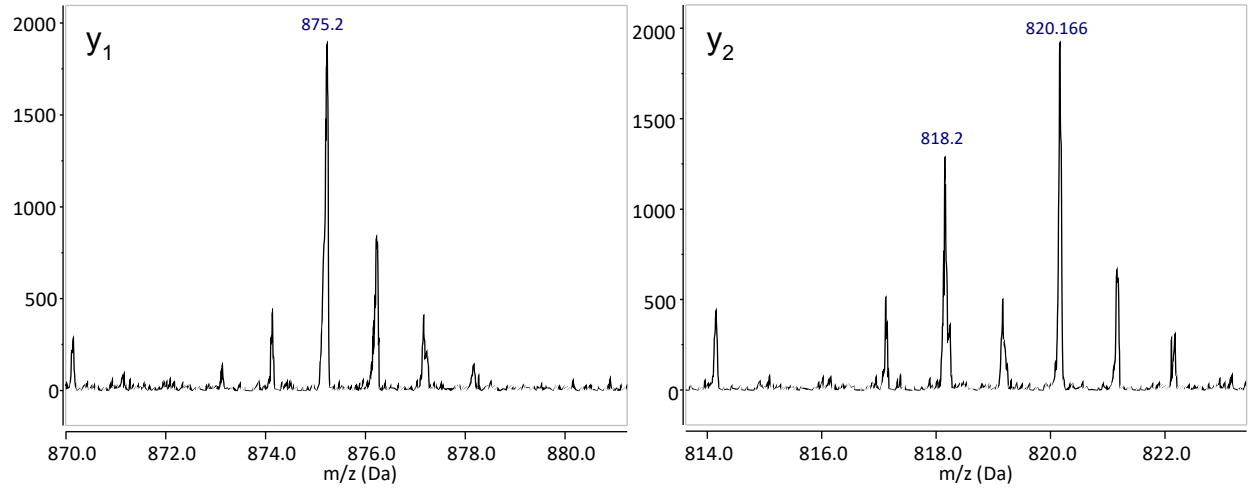
Reaction Components		Mass $[M+1H]^+$	
		calcd	obs
Substrates	Abz-LPATGG-K(Dnp)	927.4	927.4
	Abz-LPATGGH-K(Dnp)	1064.5	1064.6
	DEAC-Ahx-LPRTGGH-NH_2	1092.6	1092.6
Nucleophiles	Ac-K(Dnp)-GGKGG-NH_2	710.3	710.4
	Ac-K(Dnp)-AisoQKaa-NH_2	823.4	823.4
Excised Fragments	GG-K(Dnp)	426.2	426.1
	GGH-K(Dnp)	563.2	563.2
Isopeptide Products	Abz-LPAT-AisoQ**K**aa-NH_2	1324.65	1324.1
	Abz-LPAT-GG**K**GG-NH_2	1211.6	1211.7
Hydroxyl Amine Product	Abz-LPAT-NHOH	535.3	535.2
Hydrolysis Product	Abz-LPAT-OH	520.3	520.2
	DEAC-Ahx-LPRT-OH	842.5	842.5
Undesirable By-Products	Abz-LPAT-(Tris)	594.3	594.3
	Abz-LPAT-(glycerol)	623.3	623.3
	GH-K(Dnp)	506.2	506.2
	Abz-LPATG-OH	577.3	577.4

7.3 Appendix III. Chromatography Methods for Purification and Analysis by HPLC or LC-ESI-MS.

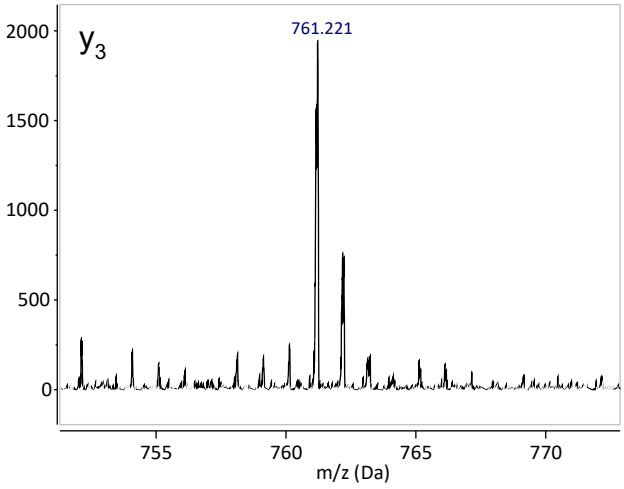
Method	Column	Gradient	Gradient	Flow Rate
Method (A)	C18 Semi Prep	10-90	10% (v/v) MeCN (0.0-2.0 min), 10% (v/v) MeCN → 90% (v/v) MeCN (2.0-15.0 min), hold 90% (v/v) MeCN (15.0-17.0 min), 90% (v/v) MeCN → 10% (v/v) MeCN (17.0-17.01 min), re-equilibrate at 10% (v/v) MeCN (17.01-19.0 min).	4.0 mL/min
Method (B)	C18 Semi Prep	20-90	20% (v/v) MeCN (0.0-2.0 min), 20% (v/v) MeCN → 90% (v/v) MeCN (2.0-15.0 min), hold 90% (v/v) MeCN (15.0-17.0 min), 90% (v/v) MeCN → 20% (v/v) MeCN (17.0-17.01 min), re-equilibrate at 20% (v/v) MeCN (17.01-19.0 min).	4.0 mL/min
Method (C)	C18 Analytical	10-90	10% (v/v) MeCN (0.0-0.5 min), 10% (v/v) MeCN → 90% (v/v) MeCN (0.5-7.0 min), hold 90% (v/v) MeCN (7.0-9.0 min), 90% (v/v) MeCN → 10% (v/v) MeCN (9.0-9.1 min), re-equilibrate at 10% (v/v) MeCN (9.1-12.0 min).	0.3 mL/min
Method (D)	C18 Analytical	5-90	5% (v/v) MeCN (0.0-2.0 min), 5% (v/v) MeCN → 90% (v/v) MeCN (2.0-10.0 min), hold 90% (v/v) MeCN (10.0-11.0 min), 90% (v/v) MeCN → 5% (v/v) MeCN (11.0-11.1 min), re-equilibrate at 5% (v/v) MeCN (11.1-13.25 min).	0.3 mL/min
Method (E)	C8 Analytical	5-90	5% (v/v) MeCN (0.0-2.0 min), 5% (v/v) MeCN → 90% (v/v) MeCN (2.0-10.0 min), hold 90% (v/v) MeCN (10.0-11.0 min), 90% (v/v) MeCN → 5% (v/v) MeCN (11.0-11.01 min), re-equilibrate at 5% (v/v) MeCN (11.01-14.3 min).	0.4 mL/min
Method (F)	C8 Analytical	5-40-90	5% (v/v) MeCN (0.0-2.0 min), 5% (v/v) MeCN → 40% (v/v) MeCN (2.0-9.0 min), 40% (v/v) MeCN → 90% (v/v) MeCN (9.0-10.0 min), hold 90% (v/v) MeCN (10.0-11.0 min), 90% (v/v) MeCN → 5% (v/v) MeCN (11.0-11.01 min), re-equilibrate at 10% (v/v) MeCN (11.01-14.3 min).	0.4 mL/min
Method (G)	C8 Analytical	10-4-90	10% (v/v) MeCN (0.0-2.0 min), 10% (v/v) MeCN → 40% (v/v) MeCN (2.0-9.0 min), 40% (v/v) MeCN → 90% (v/v) MeCN (9.0-10.0 min), hold 90% (v/v) MeCN (10.0-11.0 min), 90% (v/v) MeCN → 10% (v/v) MeCN (11.0-11.01 min), re-equilibrate at 10% (v/v) MeCN (11.01-14.3 min).	0.4 mL/min
Method (F)	C4 Analytical	10-90	10% (v/v) MeCN (0.0-0.5 min), 10% (v/v) MeCN → 90% (v/v) MeCN (0.5-7.0 min), hold 90% (v/v) MeCN (7.0-9.0 min), 90% (v/v) MeCN → 10% (v/v) MeCN (9.0-9.1 min), re-equilibrate at 10% (v/v) MeCN (9.1-12.0 min).	0.3 mL/min

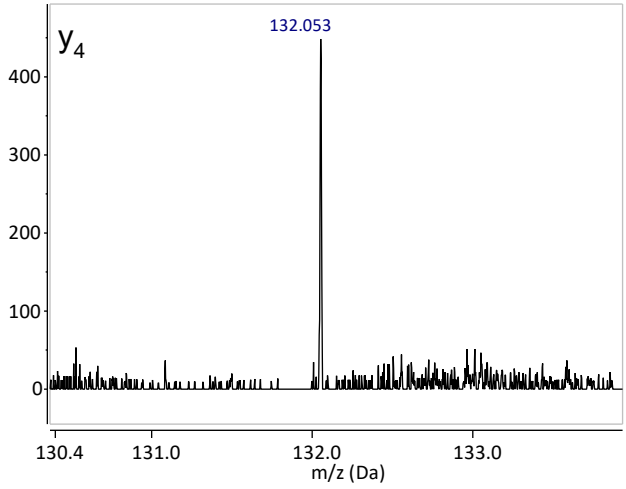
7.4 Appendix IV. Representative MS/MS Spectra for b and y Ions Produced from Main Peptide of Ac-K(Dnp)-GGKGG-NH_2 Isopeptide Product.

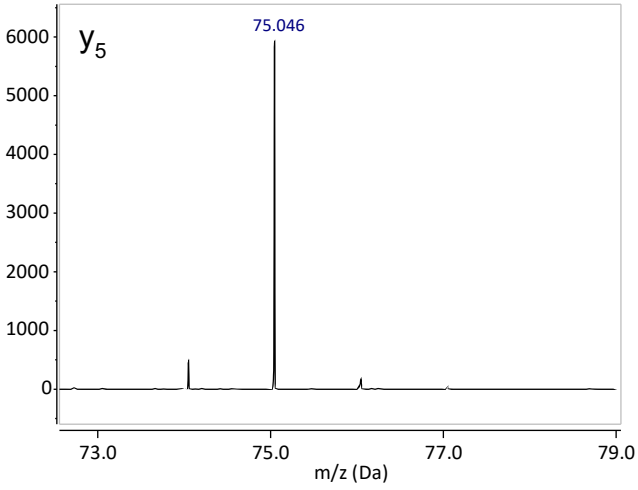



y1
875.2
2000
1500
1000
500
0
870.0
872.0
874.0
876.0
878.0
880.0
m/z (Da)

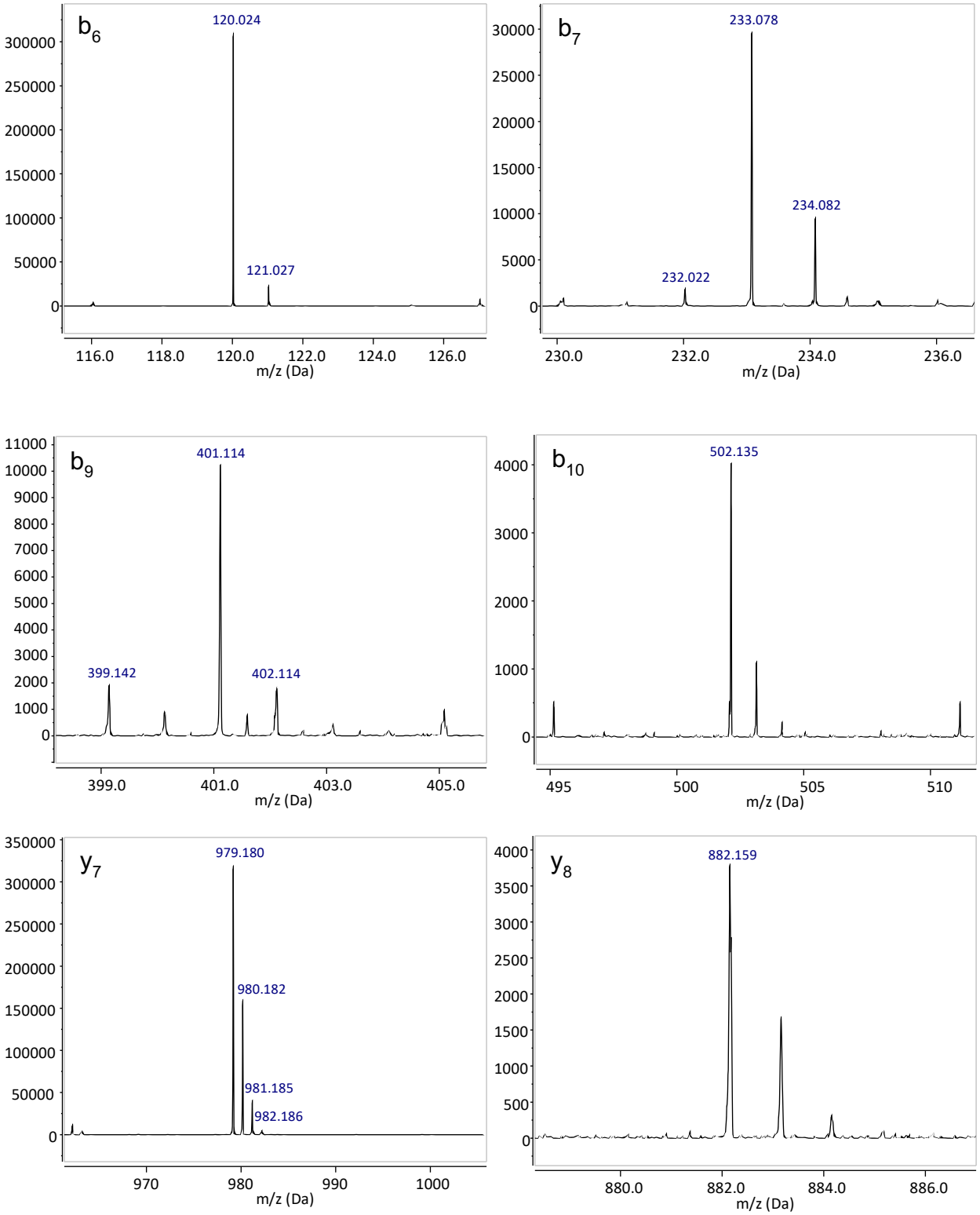
y2
820.166
818.2
2000
1500
1000
500
0
814.0
816.0
818.0
820.0
822.0
m/z (Da)

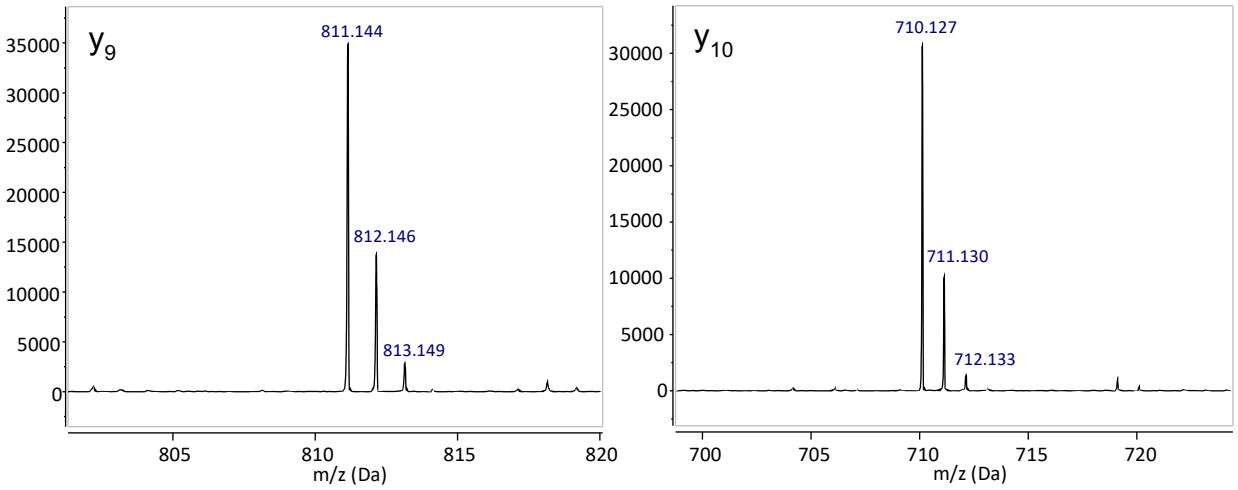

y3
761.221
2000
1500
1000
500
0
755
760
765
770
m/z (Da)


y4
132.053
400
300
200
100
0
130.4
131.0
132.0
133.0
m/z (Da)

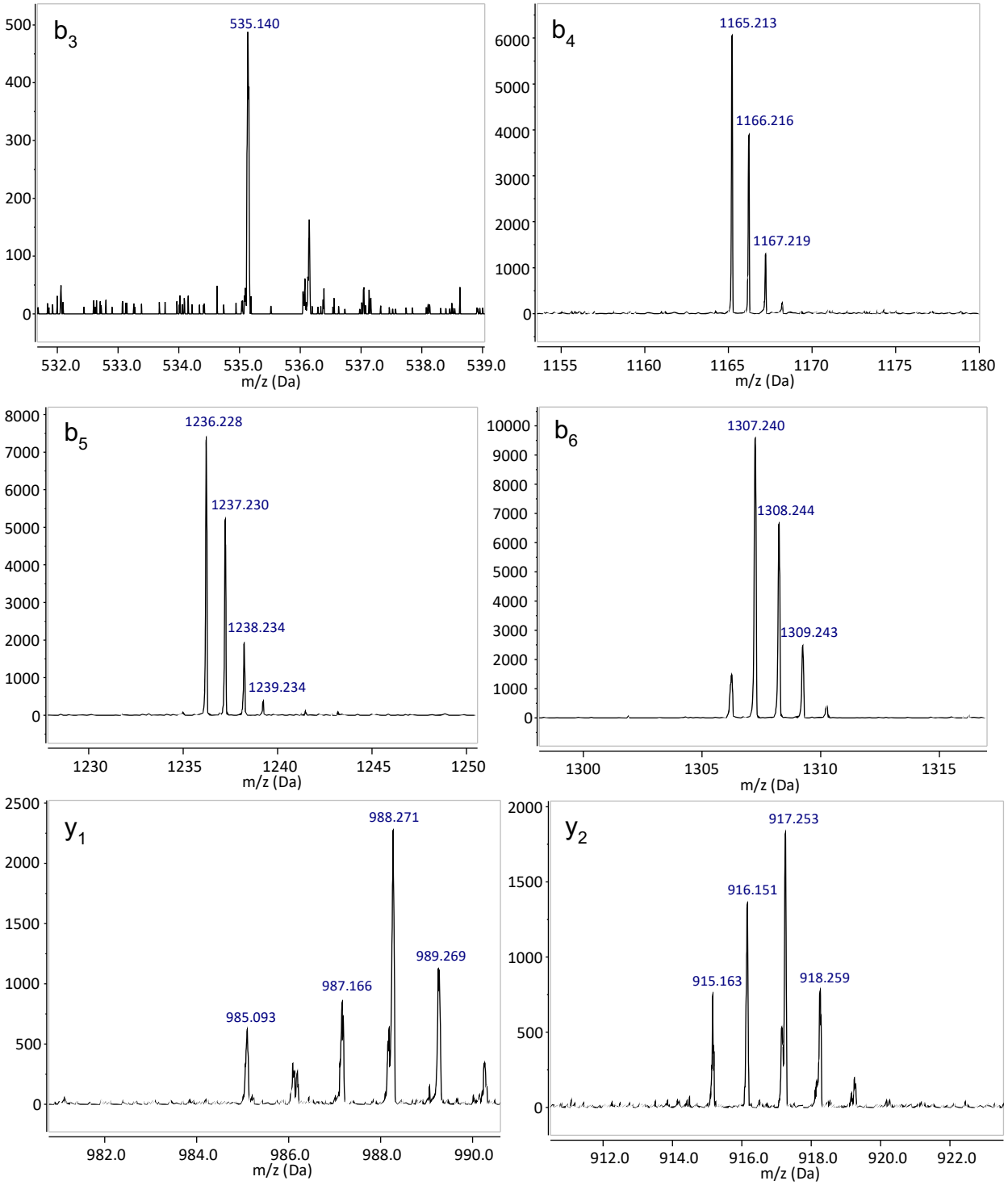

y5
75.046
6000
5000
4000
3000
2000
1000
0
73.0
75.0
77.0
79.0
m/z (Da)

7.5 Appendix V. Representative MS/MS Spectra for b and y Ions Produced from the Modification of Ac-K(Dnp)-GGKGG-NH_2 Isopeptide Product.

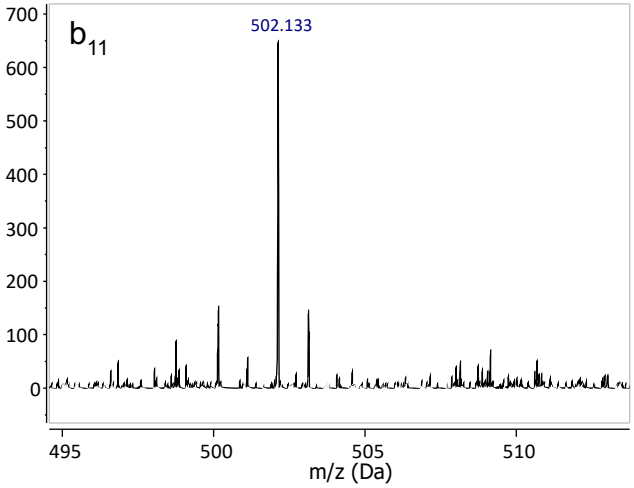


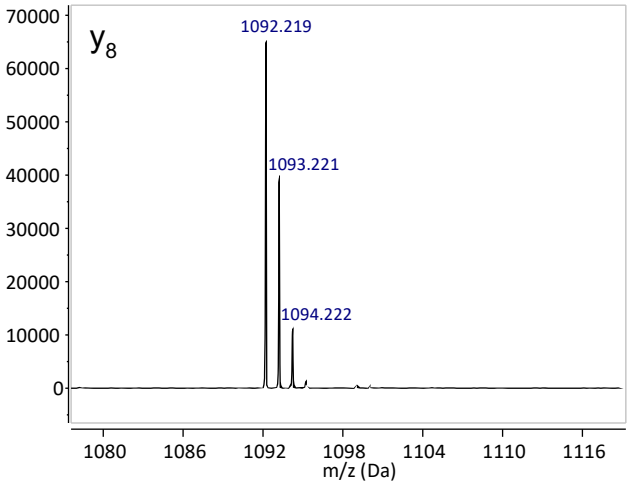

y9
811.144
812.146
813.149
35000
30000
25000
20000
15000
10000
5000
0
805
810
815
820
m/z (Da)
y10
710.127
711.130
712.133
30000
25000
20000
15000
10000
5000
0
700
705
710
715
720
m/z (Da)

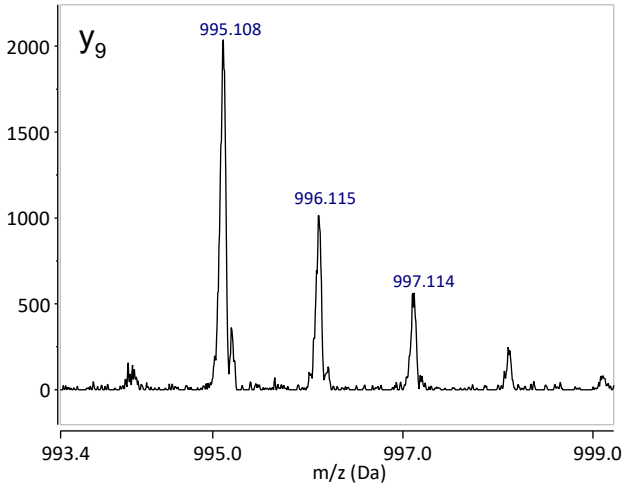
7.6 Appendix VI. Representative MS/MS Spectra for b and y Ions Produced from Main Peptide of Ac-K(Dnp)-AisoQKaa-NH_2 Isopeptide Product.

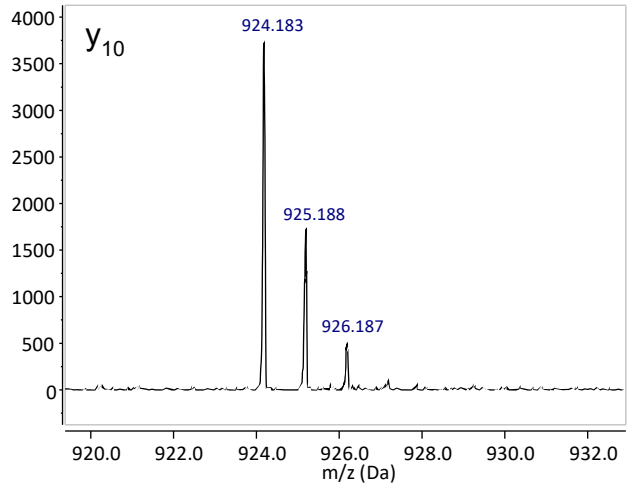


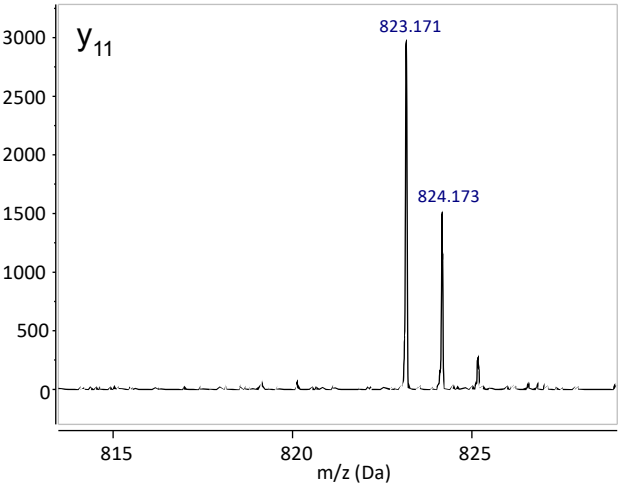
7.7 Appendix VII. Representative MS/MS Spectra for b and y Ions Produced from the Modification of Ac-K(Dnp)-AisoQKaa-NH_2 Isopeptide Product.











7.8 Appendix VIII. Table of Calculated and Observed Masses for Larger Peptide Studies

Reaction Components			m/z	
			calcd	obs
Substrate	DEAC-Ahx-LPRTGGH-NH_2	$[M+1H]^+$	1092.6	1092.6
Nucleophiles	Indolicidin	$[M+3H]^+$	636.5	636.4
	Glucagon	$[M+4H]^+$	821.2	871.6
	Exendin-4	$[M+4H]^+$	1047.3	1047.7
	β-endorphin	$[M+4H]^+$	867.3	867.3
Singly-Modified Products	DEAC-Ahx-LPRT-Indolicidin	$[M+1H]^+$, $[M+4H]^+$	1325.7, 683.6	1325.1, 683.5
	DEAC-Ahx-LPRT -Glucagon	$[M+3H]^+$	1436.7	1436.3
	DEAC-Ahx-LPRT -Exendin-4	$[M+4H]^+$	1253.6	1253.6
	DEAC-Ahx-LPRT - β-endorphin	$[M+4H]^+$	1073.3	1073.5
Second Singly-Modified Product	DEAC-Ahx-LPRT -Indolicidin	$[M+4H]^+$	683.6	683.4
Doubly-Modified Products	$(DEAC\text{-}Ahx\text{-}LPRT)_2$-Indolicidin	$[M+4H]^+$	889.8	889.5
	$(DEAC\text{-}Ahx\text{-}LPRT)_2$-Exendin-4	$[M+5H]^+$	1168.4	1168.0
	$(DEAC\text{-}Ahx\text{-}LPRT)_2$- β-endorphin	$[M+5H]^+$	1023.6	1023.6
Triply-Modified Product	$(DEAC\text{-}Ahx\text{-}LPRT)_3$- β-endorphin	$[M+4H]^+$	1485.3	1485.0
Four-Modified Product	$(DEAC\text{-}Ahx\text{-}LPRT)_4$- β-endorphin	$[M+5H]^+$	1353.2	1353.1
Five-Modified Product	$(DEAC\text{-}Ahx\text{-}LPRT)_5$- β-endorphin	$[M+5H]^+$	1518.0	1518.0
Six-Modified Product	$(DEAC\text{-}Ahx\text{-}LPRT)_6$- β-endorphin	$[M+6H]^+$	1402.5	1402.4
Hydrolysis Product	DEAC-Ahx-LPRT-OH	$[M+1H]^+$	842.5	842.5
Undesirable By-Products	GH-K(Dnp)	$[M+1H]^+$	506.2	506.2
	DEAC-Ahx-LPRT -OH	$[M+1H]^+$	577.3	577.3