ABSTRACT

Characterization of Cell Wall Peptidoglycan in *Enterococcus faecalis* Biofilm
Using Stable Isotope Labeling by Amino Acids in Cell Culture

Ashley Wallace

Director: Dr. Sung Joon Kim, Ph.D.

The formation of bacterial biofilms is a significant concern in healthcare settings. In an effort to expand the current body of knowledge about biofilm and provide information useful for clinical treatment, this study aims to characterize the chemical composition of the cell wall peptidoglycan of *Enterococcus faecalis* by liquid chromatography-mass spectrometry (LC-MS). Using Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC), the relative abundance of common cell wall modifications was compared for biofilm and planktonic bacteria. The biofilm phase was found to exhibit decreased O-acetylation, increased N-deacetylation, increased crosslinking, increased carboxypeptidase activity, and increased crosslinker biosynthesis. These findings are consistent with the lower metabolic activity of bacteria in the biofilm phase and the adhesion of biofilm cells to one another and to the hydrophilic components of the biofilm matrix.

APPROVED BY DIRECTOR OF HONORS THESIS:

_____

Dr. Sung Joon Kim, Department of Chemistry and Biochemistry

APPROVED BY THE HONORS PROGRAM:

_____

Dr. Elizabeth Corey, Director

DATE: _____

CHARACTERIZATION OF CELL WALL PEPTIDOGLYCAN IN *ENTEROCOCCUS FAECALIS* BIOFILM USING STABLE ISOTOPE LABELING BY AMINO ACIDS IN CELL CULTURE

A Thesis Submitted to the Faculty of

Baylor University

In Partial Fulfillment of the Requirements for the

Honors Program

By

Ashley G. Wallace

Waco, Texas

May 2016

TABLE OF CONTENTS

ACKNOWLEDGMENTS

Working on this thesis has taught me that the research process extends beyond the timeline of a single project or the work of a single person. I am privileged to have had the opportunity to work with people who invite me into this bigger process and inspire me with the way they approach their scholarship. Thank you to my thesis director Dr. Sung Joon Kim, who graciously welcomed me into his lab, provided a vision for what the research process should look like, and took the time to instruct me along the way. I would also like to thank my graduate student mentor James Chang for guiding me through each aspect of the project; from training to planning to analyzing data to writing to editing, his eye for the big picture as well as the details has been invaluable. I am also grateful to Erin Foster, who first introduced me to the lab, collaborated with me during her summer before medical school, and continues to be an excellent example of dedication and diligence. For these three people, their calling to faith and scholarship is evident in the way they approach their work and interact with those around them, and I am so blessed to have had the opportunity to be mentored by them.

I am also blessed by all of my friends, family, and professors who have supported me along the way. Thank you to Dr. Cheolho Sim and Dr. Charles Garner, who served on my thesis defense committee, and to Dr. Alejandro Ramirez, who helped ensure the optimal performance of the mass spectrometer. Thank you to everyone who has supported me in other ways, whether by kindling my love for science or providing advice or simply smiling as I described some detailed aspect of my project. Thank you for encouraging me to be a part of this research process, which has taught me much more than I could have imagined.

CHAPTER ONE

Introduction

Enterococcus faecalis

Enterococcus faecalis is a species of Gram-positive bacteria that resides in the gastrointestinal tract of healthy individuals. However, this opportunistic pathogen is also responsible for 65-80% of enterococcal nosocomial infections, such as endocarditis, urinary tract infections, and bacteremia (Guiton et al., 2009). The CDC's 2013 report classifies enterococcus as a "serious" level of threat and reports 66,000 serious enterococcal infections each year, of which a significant portion are caused by *E. faecalis*. 20,000 of these infections are resistant to antibiotics, and 1,300 result in death (US Department of Health and Human Services, 2013). Its ability to form biofilms that can better withstand host defenses and antibiotics makes *E. faecalis* particularly difficult to treat but also useful for studying bacterial biofilms. The study of biofilm can be approached from a variety of perspectives, and numerous factors have been found to contribute to biofilm formation. However, the structure of biofilm's cell wall, an important therapeutic target, has not been adequately explored.

Biofilm

The National Institutes of Health estimates that 80% of microbial infections involve biofilm. Figure 1 identifies several common sites of biofilm infection, and other routes of primary infection include burn wounds and the lungs of patients with cystic fibrosis (Yeagley et al., 2012). Because biofilm development occurs in a cycle of attachment, accumulation, and dispersal (Figure 2), the primary infection that survives antibiotic treatment can spread

to secondary sites, such as the liver, kidney, and brain, releasing harmful toxins and causing severe inflammatory responses that can damage organs (Costeron and Stewart, 1999).

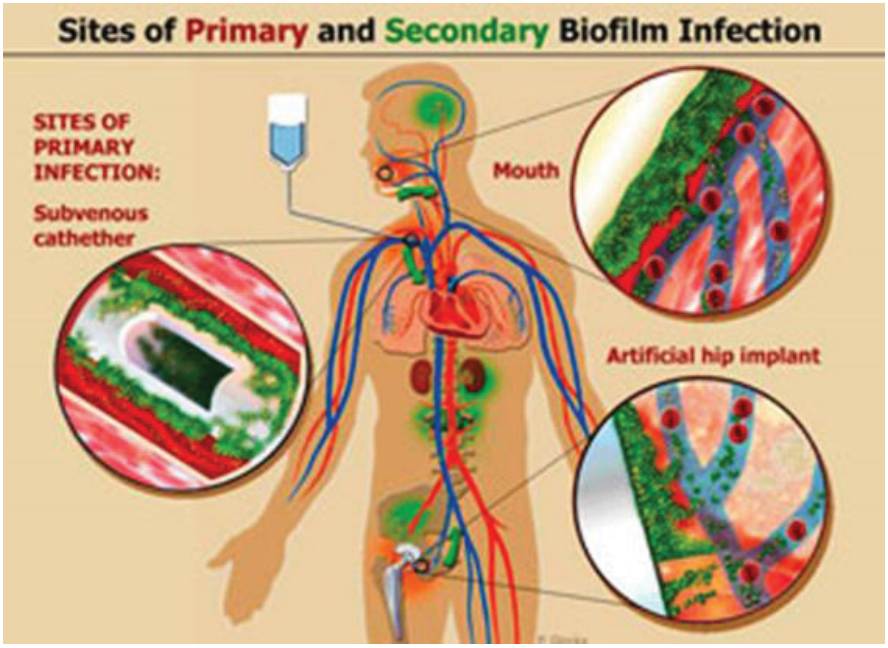


Figure 1: Sites of Primary and Secondary Biofilm Infection (Stoodley et al., 2003).

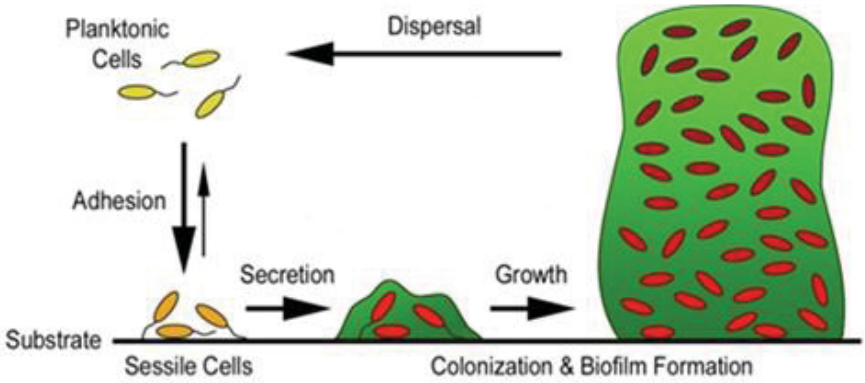


Figure 2: Cycle of Biofilm Development (Archer, 2011).

On a clinical level, biofilms are responsible for chronic infections that are more difficult to eliminate (Bjarnsholt, 2013) because of their decreased susceptibility to host

defense mechanisms, antimicrobial activity, and adverse conditions (Tu Quoc et al., 2007). It has been estimated that bacteria in the biofilm phase are more than 1000-fold more resistant to antibiotics and the host immune response than free-floating bacteria (Yeagley et al., 2012). Perhaps because of the limited amount of glucose and oxygen that can reach the bacteria within the biofilm, the bacteria enter a stationary phase of growth (Zimmerli et al., 2004). This metabolic dormancy also shields bacteria in a biofilm from the effects of antibiotics and host defenses (Costeron and Stewart, 1999).

In order to access the shielded bacteria in the middle of the biofilm, one of the current treatments involves mechanical debridement of the bacterial biofilm and flushing with antibiotics (Zimmerli et al., 2004). This treatment method is time-consuming, messy, and provides opportunity for further infection, and so prevention of biofilm formation has been the focus of most biofilm research so far. In an effort to better understand biofilm's function, researchers seek to characterize the structure of biofilm on the molecular level by examining essential elements such as the extracellular matrix, cell wall proteins, and peptidoglycan structure.

Biofilms consist of clustered communities of bacterial cells embedded in a matrix of extracellular polymeric substance (EPS) (Figure 3). EPS includes polysaccharides, proteins, nucleic acids, and lipids, which make up 90% of the dry mass of biofilms. The matrix serves a variety of functions, particularly allowing for adhesion and cohesion, but also serving as a diffusion barrier to prevent the infiltration of antimicrobials (Archer, 2011) and a means of trapping nutrients for the bacteria within the matrix (Flemming and Wingender, 2010).

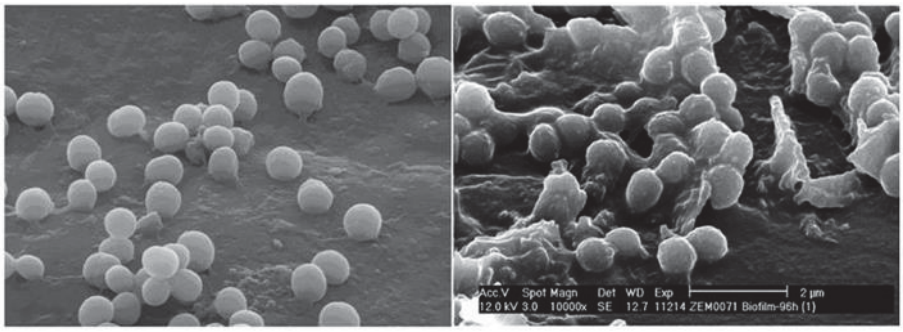


Figure 3: Electron Microscope Comparison of Free-Floating Bacteria and Biofilm of *Staphylococcus aureus* (Geipel, 2009).

The extracellular matrix is one of the defining features of biofilms, and the factors contributing to this phenotypic presentation are numerous (Archer, 2011). Factors implicated in biofilm formation include sortases, autolysins, extracellular DNA, teichoic acid charge, cell aggregation-associated proteins, and biofilm-associated proteins (Guiton et al., 2009)(Tu Quoc et al., 2007). Alterations in the peptidoglycan composition of the cell wall have also been found to dramatically influence biofilm formation. Exploratory experiments show that biofilm formation is dependent upon proper addition of the peptidoglycan crosslinking bridge, as demonstrated by the absence of crystal violet staining for FemA and FemAB deletion mutant strains of *Staphylococcus aureus*, which have genetic defects in the enzymes that add the crosslinking bridge (Figure 4).

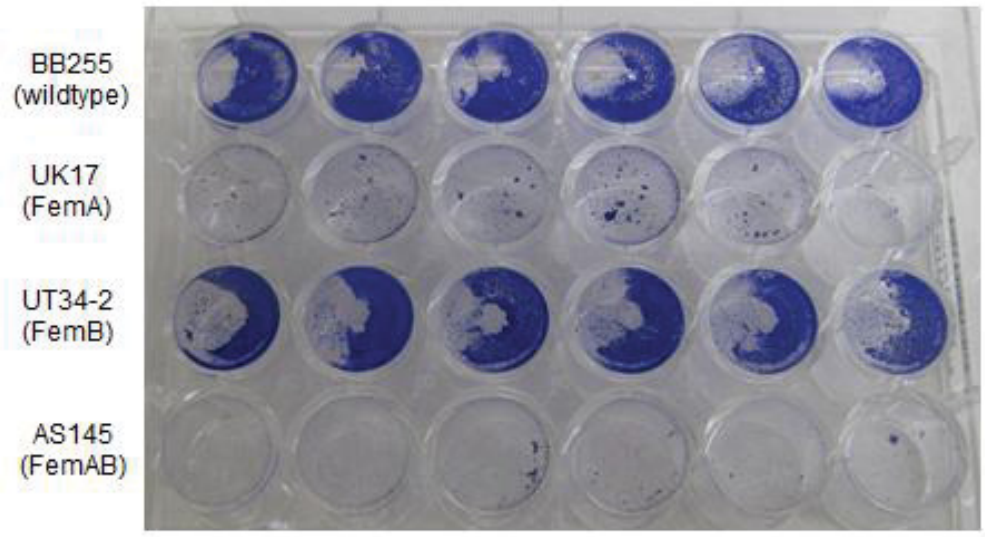


Figure 4: Crystal Violet Staining of *S. aureus* Biofilm Plate: wildtype, FemA, FemB, FemAB deletion mutants.

Peptidoglycan Structure

For gram positive bacteria such as *E. faecalis*, the cell membrane is surrounded by a thick layer of peptidoglycan (30-100 nm) and a variety of cell wall proteins and teichoic acids (Figures 5-6). This physical and defensive boundary is essential for the survival of bacteria, and any changes in the structure can provide valuable clues about the nature of biofilm.

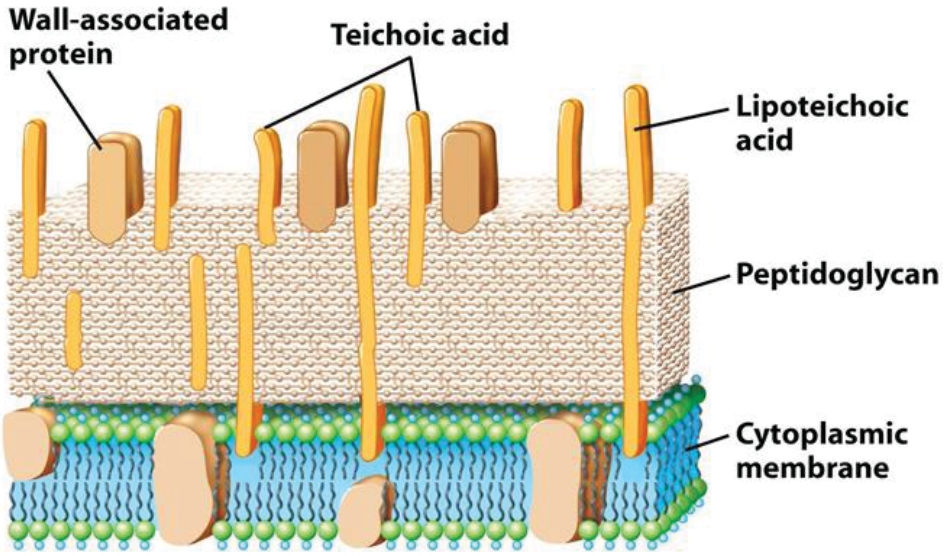


Figure 5: Representation of Cell Envelope in Gram Positive Bacteria (Brock et al., 1994).

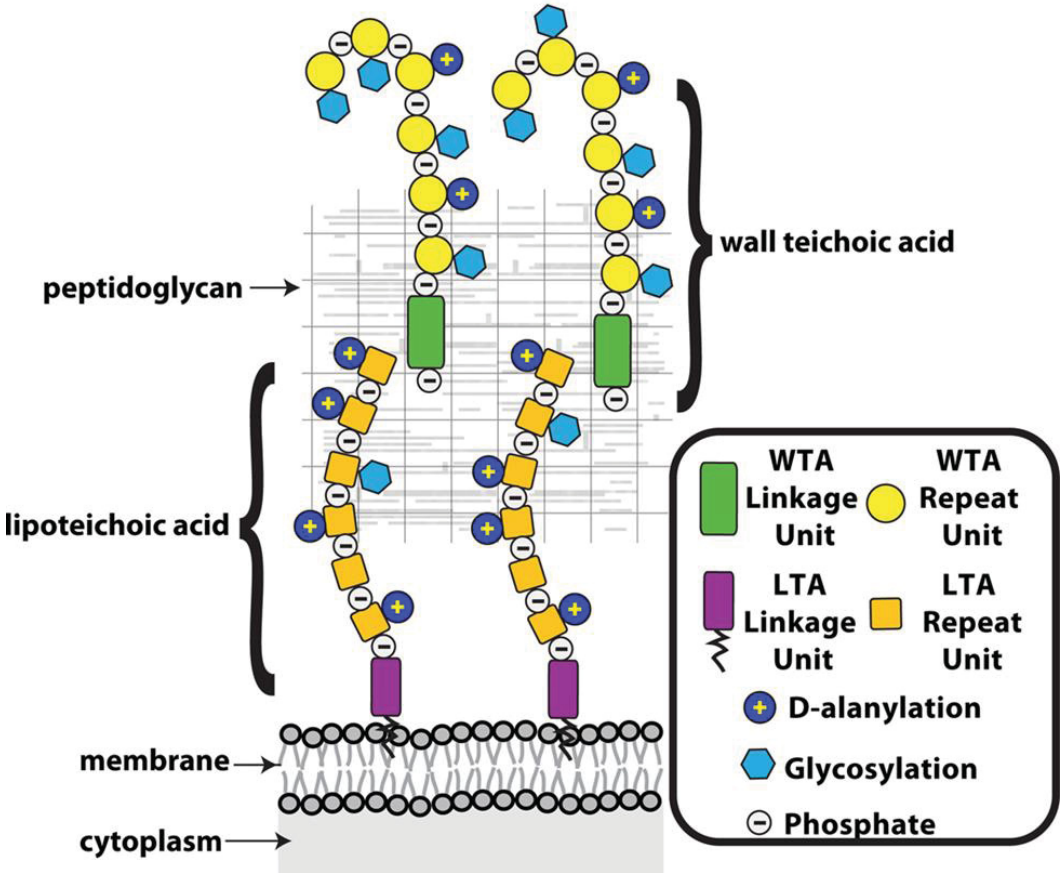


Figure 6: Teichoic Acids Found in Gram Positive Bacteria (Brown, 2013).

Each peptidoglycan unit consists of two β-1,4 linked sugars, N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM), and a seven amino acid branched peptide stem extending from the NAM sugar. These peptidoglycan units can be linked to each other by crosslinking in a branch of the peptide stem (Figure 7) and by linking of the sugars.

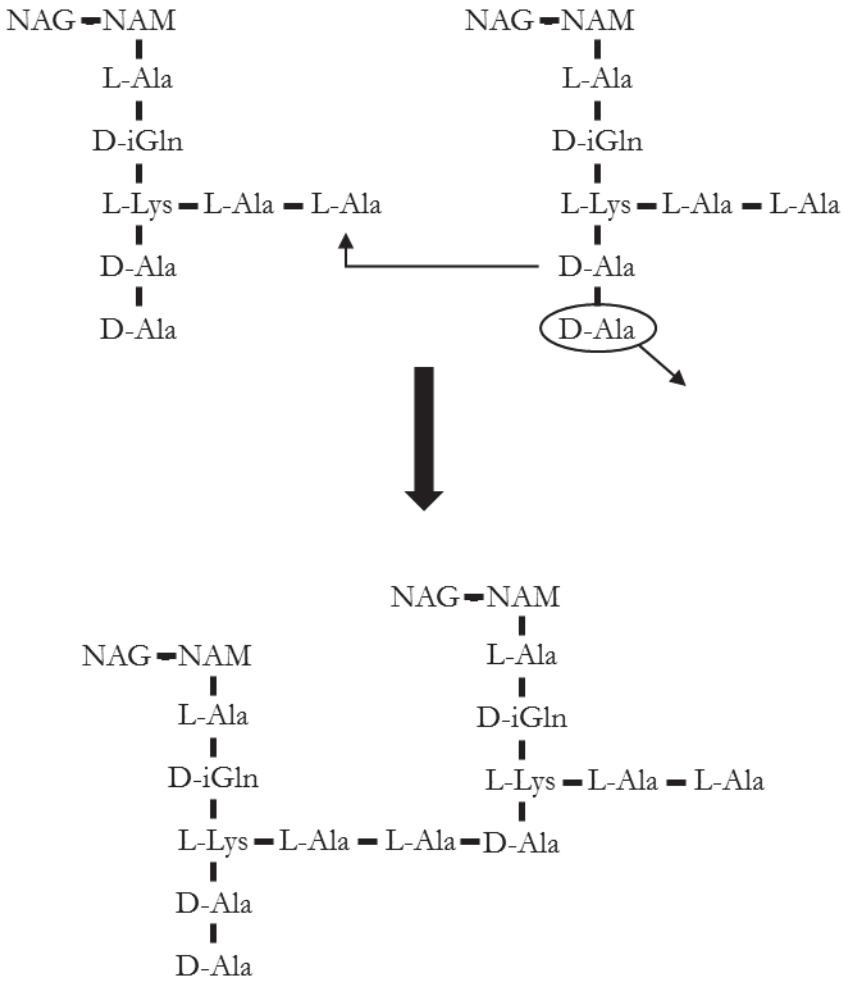


Figure 7: Crosslinking by transpeptidase for *E. Faecalis* Peptidoglycan Unit. Crosslink is a peptide bond between the L-Ala of one peptidoglycan unit to the penultimate D-Ala of the peptidoglycan stem of an adjacent glycan strand. During the formation of crosslinking the terminal D-Ala is cleaved from the acceptor stem.

Several modifications on the basic peptidoglycan structure are commonly observed, including O-acetylation and N-deacetylation of the sugars and removal of terminal alanine from the peptide stem. These modifications are known to help bacteria evade the host immune system and direct the remodeling of the cell wall (Figure 8).

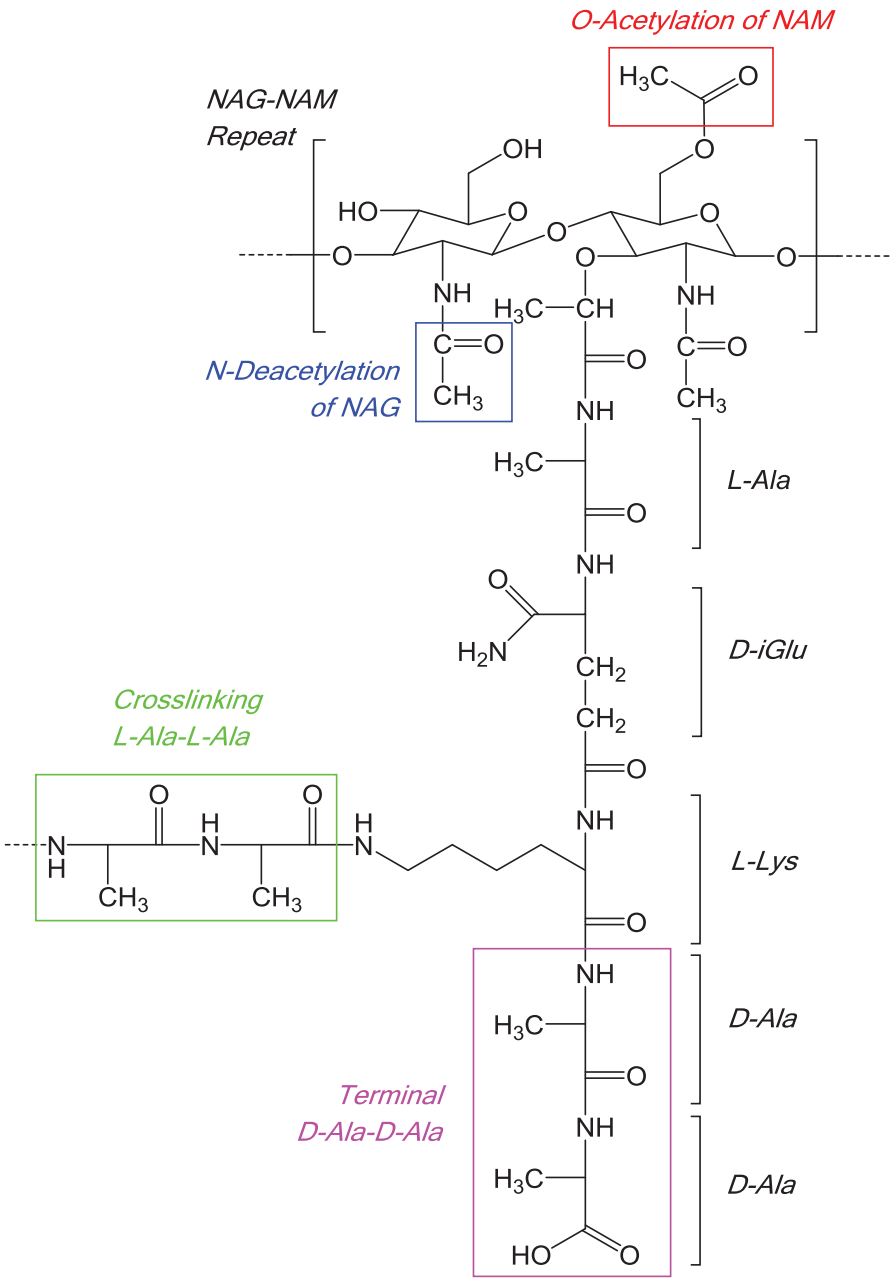


Figure 8: Chemical Structure of *E. faecalis* Peptidoglycan with Post-Synthetic Modifications.

O-acetylation occurs at the C6 hydroxyl group of NAM and serves as a defense mechanism for bacteria in several ways. O-acetylation precludes normal hydrogen binding of the C-6 hydroxyl group to the lysozyme active site. Any modification at the C-6 position also inhibits the activity of N-acetylmuramidases, an important class of autolysins involved in cell wall maintenance and renewal of gram-positive bacteria (Moynihan et al., 2014). O-acetylated species have been found to be more abundant in the stationary phase, in which bacteria are thought to be responding to an increased demand for resistance against antibiotics (Vollmer, 2008).

However, a decrease in O-acetylation alone is not sufficient to make bacteria susceptible to the body's defenses. N-deacetylation, carried out by the enzyme N-acetyl-D-glucosamine aminohydrolyase, which is influenced by several outside environmental influences, is another important factor in protecting bacteria from the host immune system and antimicrobial substances (Benachour et al., 2012). Like O-acetylated peptidoglycan, N-deacetylated peptidoglycan is a poor substrate for lysozyme. Deacetylation also increases the positive charge on the cell wall, protecting the bacteria against antimicrobial peptides of the host organism (Vollmer, 2008).

Another way that bacteria protect themselves is by cleaving the terminal alanine from the peptide stem. Removal of these alanine prevents glycopeptide antibiotics from binding to the terminal alanine and forming complexes which prevent crosslinking and ultimately inhibit cell wall synthesis (McKessar et al., 2000). Cleavage of the terminal D-Ala, though upregulated in strains with vancomycin resistance, is also observed in strains without vancomycin resistance (Foster, 2015).

Since peptidoglycan serves as an important structural and defensive component of bacteria, any upregulation or downregulation of these peptidoglycan modifications in biofilm is noteworthy.

SILAC

Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC) is a quantitative proteomic method that utilizes mass spectrometry to highlight the differences between two samples, one with isotopic label and one without (Figure 9).

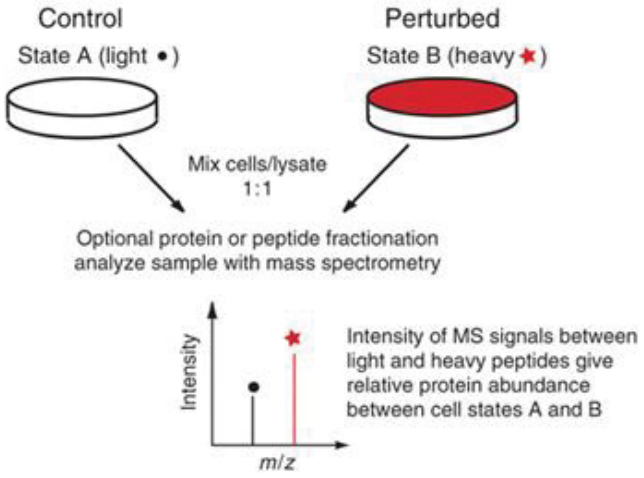


Figure 9: An Overview of the SILAC Technique (Ong and Mann, 2007).

One of the methods for labeling is metabolic incorporation of stable isotopes into living cells, which can be completed within five doublings and does not change the chemistry of the labeled proteins. The advantage of this technique is the ability to compare relative quantities of two or more samples on the same mass spectrum; it also allows multiple samples to be digested and prepared in identical conditions. While this technique was originally developed for studying mammalian cell culture, it has also been adapted to

quantify proteins in fungi, bacteria, single-celled eukaryotes, and plants (Ong and Mann, 2007).

One previously unexplored application of SILAC is the use of SILAC for the study of peptidoglycan in bacterial cell walls. Though this method is typically used to quantify smaller peptides, the peptidoglycan in the stem of *E. faecalis* contains lysine, an amino acid commonly used in SILAC experiments (Ong and Mann, 2007). Isotopic labeling of this lysine allows for the application of the SILAC technique to peptidoglycan samples.

Experimental Goals and Hypotheses

The purpose of this experiment was to apply SILAC techniques to the mass spectrometric study of *E. faecalis* in order to quantify the differences in peptidoglycan composition between biofilm and planktonic bacteria.

By generating a library of all potential variations of peptidoglycan fragments, the abundance and type of fragments in each sample can be identified by matching their masses to LC-MS data (Foster 2015). A ratio of peak intensities for heavy and light isotopologues of each fragment can be obtained, as is common in the analysis of SILAC data. In the case where biofilm is labeled with a heavy isotopomers of lysine and planktonic bacteria remain unlabeled, this ratio provides information about which modifications are more common in biofilm.

An advantage of SILAC is the ability to compare biofilm and planktonic samples on the same LC-MS run with identical sample digestion and preparation for accurate comparison. However, this novel application of SILAC to a larger molecule such as peptidoglycan introduces several challenges. The raw intensities of the peaks cannot be directly compared due to less than 100% isotopic enrichment of labeled lysine and

incomplete isotopic incorporation of labeled lysine into peptidoglycan. However, with the aid of an in-house MATLAB program, the actual enrichment percentage can be calculated by comparing the observed isotopic distribution to a theoretical distribution, and incomplete incorporation can be corrected using the intensities of fully incorporated and partially incorporated peaks. With these considerations taken into account, a more accurate ratio can be calculated for each fragment, yielding more accurate information about the relative abundance of each modification in biofilm and planktonic bacteria.

The peptidoglycan modifications of interest in this experiment are crosslinking, O-acetylation, N-deacetylation, cleavage of the two terminal D-Ala, and failure to add the two crosslinking L-Ala. Because of the sessile, cohesive character of biofilm, it was expected that biofilm would demonstrate a greater propensity toward crosslinking than planktonic bacteria. Biofilm's enhanced survival in the face of immune defenses and antibiotics suggests that biofilm would display increased O-acetylation and N-deacetylation as well. An increased removal of terminal alanine by L,D and D,D-carboxypeptidases was also anticipated in biofilm, since this modification is believed to decrease biofilm's sensitivity to antibiotics and host defenses. Biofilm was expected to exhibit a greater fidelity in the addition of the L-Ala of the crosslinking bridge by BppA1 and BppA2 due to its lower turnover rate of peptidoglycan synthesis.

CHAPTER TWO

Materials and Methods

Sample Preparation

Bacterial Growth and Harvesting

A clinical strain of vancomycin-susceptible *Enterococcus faecalis* (ATCC 29212, strain designation Portland) was grown in Enterococcus Defined Media (EDM) with labeled lysine substituted as needed. Planktonic bacteria and biofilm were harvested and stored for later mixing.

EDM was prepared using the compounds listed in Table 1 (Sigma-Aldrich, Fisher, Alfa Aesar, Amresco, EMD). Salt and buffer stock solution, vitamins stock solutions, and trace metals stock solution were prepared in advance. Nucleosides and tyrosine were added to approximately half of the final volume of autoclaved deionized water. This solution was first heated to dissolve insoluble solids, then allowed to cool to room temperature. Salt and buffer stock solution, vitamins stock solutions, and D-glucose were added to this solution. L-amino acids were added to this mixture, replacing natural abundance lysine ("light") with labeled lysine ("heavy") as appropriate. The labeled lysine used in this experiment was L-Lysine:2HCl (^{13}C6, D9, ^{15}N2)(Cambridge Isotope Laboratories). Trace metal stock solution was added drop by drop to avoid precipitation. Iron sulfate in acidic stock solution (pH 2.0) was added separately. The mixture was brought to the full volume using autoclaved DI water and adjusted to pH 7 using potassium hydroxide. The mixture was vacuum filtered through a 0.2 μm pore size filter (VWR) to remove any contamination. EDM was stored at 4°C.

Table 1: Formulation for EDM. Amounts are shown for 1 L. An asterisk indicates that the compound was added from a separate stock solution.

Salt and Buffer Stock	K_2HPO_4	4.3 g
	KH_2PO_4	4.0 g
	$(NH_4)_2SO_4$	1 g
	EDTA	1 mg
	H_3BO_3*	0.1 mg
Vitamins Stock 1	Biotin*	0.1 mg
	Folic Acid*	0.2 mg
	Thiamine HCl	2 mg
	Calcium pantothenate	2 mg
	Pyridoxine HCl	20 mg
	Niacin	2 mg
	Riboflavin	2 mg
	Inositol	1 mg
Vitamins Stock 2	P-aminobenzoic acid	0.05 mg
Vitamins Stock 3	Cyanocobalamin	0.01 mg
Trace Metal Stock	$NiAc_2 \cdot 4H_2O$	0.1 mg
	$(NH_4)_6Mo_7O_{24} \cdot 4H_2O$	0.2 mg
	$MgSO_4 \cdot 7H_2O$	30 mg
	$MnSO_4 \cdot H_2O$	10 mg
	$CuSO_4 \cdot 5H_2O$	1 mg
	$ZnSO_4 \cdot 7H_2O$	1 mg
	NaCl	10 mg
	$CaCl_2$	1 mg
Iron Stock	$FeSO_4 \cdot 7H_2O$	10 mg
Nucleosides	Adenine	5 mg
	Uracil	
	Cytosine	
	Guanine	
	Xanthine	
Tyrosine	L-Tyrosine	100 mg
Glucose	D-Glucose	10 g
L- Amino Acids	L-Aspartic Acid	100 mg
	L-Cysteine	
	L-Tryptophan	

	L-Proline	
	L-Isoleucine	
	L-Leucine	
	L-Valine	
	L-Phenylalanine	
	L-Arginine	
	L-Threonine	
	L-Histidine	
	L-Serine	
	L-Methionine	
	L-Asparagine	
	L-Alanine	
	Glycine	
	L-Lysine	
	L-Glutamic Acid	
	L-Glutamine	

An overnight culture in tryptic soy broth (TSB) media was inoculated with 1% inoculum of frozen stock. The culture was grown for 24 hours at 37°C at 80 rpm.

The wells of six-well tissue culture plates (Corning) were filled with 4 mL of EDM with either light or heavy lysine according to experimental design. Each well was inoculated with 1% inoculum of overnight culture, and the plates were incubated at 37°C and 80 rpm until the bacteria reached the beginning of the stationary phase ($OD_{600} \sim 1.0$).

24 hours after the beginning of the stationary phase, planktonic bacteria and nascent biofilm were harvested from appropriate wells. OD_{600} measurements were taken for planktonic bacteria (Spectronic). Grown culture from wells without labeled lysine was transferred by pipetting into centrifuge tubes (VWR), ensuring that no biofilm was removed from the surfaces of the well. These samples containing planktonic bacteria were pelleted by centrifuging with Allegra X-15R tabletop centrifuge with rotor SX4750 (Beckman Coulter) at 4750 rpm for 8 minutes with maximum acceleration and deceleration. Supernatant was

removed from each centrifuge tube, and each pellet was resuspended in 4 mL phosphate buffered saline (PBS). Samples were frozen at -80°C.

For wells with labeled lysine, grown culture was removed and the wells were carefully rinsed twice with 4 mL PBS. 1 mL PBS was added to each well, and wells were scraped to loosen biofilm. Suspended biofilm was transferred from the plates to centrifuge tubes. Samples were pelleted by centrifuging at 4750 rpm for 8 minutes. Supernatant was removed from each centrifuge tube, and each pellet was resuspended in 4 mL PBS. Samples were stored at -80°C.

Sterilization and Mixing of Samples

Samples were boiled in a hot water bath for 30 minutes to sterilize and deactivate bacterial enzymes involved in cell wall growth and modification. Then OD_{600} measurements were used to combine labeled biofilm samples and unlabeled planktonic samples in 1:1 ratios. Initial OD_{600} measurements were taken for each sample, and PBS was used to adjust the optical densities to 0.6 for samples to be combined. Once the bacterial suspensions were adjusted to equal densities, equal volumes of the samples were combined.

Cell Wall Isolation

Bacteria in the prepared mixtures were disrupted mechanically, and then cell walls were isolated chemically in preparation for peptidoglycan analysis by LC-MS. The prepared mixtures were pelleted by centrifuging for 8 minutes at 4750 rpm. The supernatant was removed, and the pellet was resuspended in 950 μL PBS.

Suspended bacteria samples were divided evenly into microcentrifuge tubes, which each contained approximately 300 μL volume equivalent of 0.5 mm diameter glass

disruption beads (VWR). The samples were then agitated (Disruptor Genie, Scientific Industries) with the beads for 8 one-minute intervals, with one minute of rest in between. The beads were removed from the samples by filtration with Steriflip 20 μm nylon net vacuum filtration system (EMD Millipore).

The volume of each sample was reduced by spinning down in microcentrifuge tubes (VWR 1207 microcentrifuge), discarding the supernatant, and resuspending the pellet in a total of 1.5 mL PBS. Approximately 3.5 mL of 2% sodium dodecyl sulfate (SDS) in deionized water was combined with each sample, and this mixture was placed in a boiling water bath for 30 minutes to remove lipids and denature proteins prior to enzymatic digestion. Samples were washed five times by spinning down in microcentrifuge tubes for 5 minutes, discarding the supernatant containing SDS, and resuspending the pellet in 1 mL deionized water. The washed pellets were resuspended in 2 mL 50 mM Tris pH 8.0 buffer made with HPLC-grade water (Fisher Scientific).

Enzymatic Digestion

Isolated cell wall samples were digested using DNase, trypsin, and mutanolysin to obtain fragments of peptidoglycan for analysis by LC-MS.

200 μg DNase (Sigma-Aldrich) was added to each sample, and samples were incubated for 24 hours at 37°C and 80 rpm to digest nucleic acids released during cell disruption. Afterwards, 200 μg trypsin (Sigma-Aldrich) was added, and samples were incubated for 24 more hours at 37°C and 80 rpm to allow for digestion of proteins covalently attached to the cell wall.

Samples were centrifuged using Sorvall Legend Micro 21 Centrifuge (Thermo Scientific) for 8 minutes at 14,800 rpm to separate the cell wall from unwanted peptide

fragments. Cell wall pellets were resuspended in 1 mL Tris buffer. Two 660-unit doses of mutanolysin (Sigma-Aldrich) were added to each sample to hydrolyze the β-(1,4) glycosidic links of peptidoglycan, and digestion occurred at room temperature for 24 hours after each dose for a total of 48 hours of digestion time to allow for thorough digestion.

In order to remove undigested components that may clog the chromatography column, samples containing digested cell wall were centrifuge filtered with filters of two different pore sizes. Samples were first centrifuge filtered through a 0.45 μm filter (EMD Millipore) by spinning for 5 minutes. Samples were subsequently centrifuge filtered through a 30 kDa molecular weight cutoff filter (VWR) by spinning for 5 minutes.

Lyophilization and Reduction

Samples were lyophilized two times, with sodium borohydride reduction after the first lyophilization to reduce NAM and facilitate better chromatographic separation.

Samples were frozen at -80°C in preparation for lyophilizing. Frozen samples were lyophilized at -55°C using FreeZone 1 (Labconco). In preparation for optimal sodium borohydride reduction, each dried sample was dissolved in 1 mL of 0.375 M sodium borate pH 9.0 buffer made with HPLC-grade water. 10 mg of sodium borohydride (Fisher Scientific) in 960 μL borate buffer was added to each sample, and the centrifuge tubes were inverted to ensure thorough mixing. Samples were left to reduce at room temperature for 30 minutes before adding 125 μL of 85% phosphoric acid (Acros) to quench the reaction. Samples were frozen at -80°C and lyophilized for the second time.

Preparation of Mass Spectrometry Sample

Lyophilized samples were centrifuge filtered, and 100 μL Pierce C18 tips (Thermo Scientific) were used to clean up the peptidoglycan sample for LC-MS.

Lyophilized samples were dissolved in 500 μL of Tris buffer. Samples were centrifuge filtered through 0.45 μm and 30 kDa filters for 8 minutes each in order to remove any undissolved solids that could block the ultra-performance liquid chromatography (UPLC) column.

The sample was brought to 1% trifluoroacetic acid (TFA) concentration using 10% TFA stock solution. The C18 tip was prepared first by wetting the membrane twice with 100 μL of 50% acetonitrile (ACN) in water and discarding the solvent. The tip was then equilibrated by aspirating 100 μL of 0.1% TFA in HPLC-grade water twice and discarding the solvent. 100 μL of the prepared sample was slowly aspirated and dispensed ten times to bind peptidoglycan to the membrane. The unbound sample and salts were rinsed away by twice aspirating 100 μL of 0.1% TFA/5% ACN and discarding the solvent. Then, 100 μL of 0.1% formic acid in 50% ACN was aspirated slowly to elute the bound peptidoglycan into the autosampler vial.

Liquid Chromatography-Mass Spectrometry

Peptidoglycan fragments were separated using liquid chromatography, and a coupled mass spectrometer was used to elucidate the composition of analyte.

C18 NanoACQUITY Ultra Performance Liquid Chromatography System (Waters) was used to separate prepared peptidoglycan fragments based on hydrophobicity. The reverse phase BEH C18 column had a length of 100 mm, a diameter of 75 μm, a bead size of 1.7 μm, and a pore size of 130 Å. The flow rate throughout the chromatographic

separation was 0.6 μL/min. The elution gradient changed in a linear fashion between the time points listed below for a total run time of 90 minutes (Table 2).

Table 2: Elution Gradient with ACN as the Organic Phase. The first minute represents isocratic flow during the loading of the column.

time (min)	% water	% ACN
0.00	98.0	2.0
1.00	98.0	2.0
60.00	50.0	50.0
61.00	15.0	85.0
66.00	15.0	85.0
67.00	98.0	2.0
90.00	98.0	2.0

The sample was ionized by nanoflow electrospray ionization (ESI) with a spray voltage of 35 V, a capillary voltage of 3.5 kV, and a flow rate of 0.6 μL/min. Synapt G2 High Definition Mass Spectrometer (HDMS) with Time-of-Flight (TOF) mass analyzer (Waters) was run in positive ion mode. The resolving power of the calibrated mass analyzer was over 40,000 FWHM with the mass error within 1 ppm RMS. Fibrinopeptide B (Glu-Fib) was used as an internal standard to correct for the drift of the instrument.

Data Analysis: SILAC

Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC) was employed to compare the relative abundance of peptidoglycan species in biofilm and planktonic bacteria. LC-MS data was obtained for prepared samples. Raw mass spectrometry data was read using MassLynx (Waters) and analyzed using in-house MATLAB program (MathWorks).

In Silico Muropeptide Library Generation

In a manner similar to bottom-up proteomics, a library of all possible species was generated, considering the change in mass arising from peptidoglycan modifications combinatorically with the following parameters: crosslinking, isotopic label, missing NAG, alanylation, lactate substitution, O-acetylation, and N-deacetylation.

The most readily observed species range from dimers to pentamers, corresponding to a crosslinking (c) of 1 to 4.

The library was generated for peptidoglycan grown labeled and unlabeled lysine (label). Species with light lysine had isotopic distributions according to natural abundance, and species labeled with heavy L-[^{13}C6, D9, ^{15}N2]Lys had modified isotopic distributions due to the 17 additional neutrons.

One possible modification is the removal of NAG from each peptidoglycan subunit (sugar_missing). Therefore, the maximum number of sugars missing is equal to the total number of subunits in a species (c+1).

On the most recently added peptidoglycan subunit in each species, the two terminal D-alanine of the peptidoglycan stem are subject to editing by carboxypeptidases, and the two crosslinking L-alanine must be added by BppA enzymes (Bouhss et al., 2002). Thus, the number of total alanine remaining after editing can vary from -2 to 2 (ala). For vancomycin resistant strains, the terminal D-alanine is known to be replaced by D-lactate. Therefore, for species with two terminal D-alanine, an additional variable is used to represent this substitution (lac).

O-acetylation of NAM is a common peptidoglycan modification (o_ac). The maximum number of O-acetylated NAMs in a species is equal to the number of subunits (c+1). N-deacetylation of NAG is another common modification (n_dac), and the

maximum possible occurrence is equal to the total number of NAGs in a species (c+1-sugar_missing).

Since each of these variables corresponds to a unique change in mass, the net change of mass from the initial species can be calculated based on the number of each modification and the change in mass that each modification contributes (Formula 1).

Formula 1: Contribution of Modifications to Final Exact Mass

$$final\ exact\ mass = \ initial\ exact\ mass + \sum(modification\ mass)$$

The initial charge of a species corresponds to the total number of protonated amines. Thus, for most modifications, the amine in the crosslinking bridge gives the species an initial charge of +1. Each instance of N-deacetylation introduces an amine on NAG, increasing the initial charge by +1.

Each species can be ionized multiple times during electrospray ionization. The mass of a species is increased according to the number of protons added, and its final charge is equal to the initial charge plus the number of times it is protonated. The mass-to-charge ratio (m/z value) for each species is equal to the total mass divided by the final charge (Formula 2).

Formula 2: Calculating the Mass-to-Charge Ratio

$$m/z \ = \frac{final\ exact\ mass + (mass_{proton} * number\ of\ ionizations)}{initial\ charge + number\ of\ ionizations}$$

Isotopic Distribution

From the library, the isotopic distribution for each library entry was determined based on the isotopic enrichment of labeled lysine.

The chemical composition is changed by a fixed number each time a species is modified, so the chemical formula for each library entry was determined by counting the number of modifications. This chemical formula was used to generate the isotopic distribution arising from natural abundance by MATLAB function *isotopicdist*, which utilizes the fast Fourier transform algorithm to calculate the probability distribution of isotopologues (Ipsen, 2014). The isotopic distribution for each species is represented by a series of isotopic masses and their corresponding probabilities, as pictured for a dimer pentapeptide in Figure 10. This distribution is representative of the light isotopic distribution and is used as the basis for calculating the isotopic distribution of labeled species.

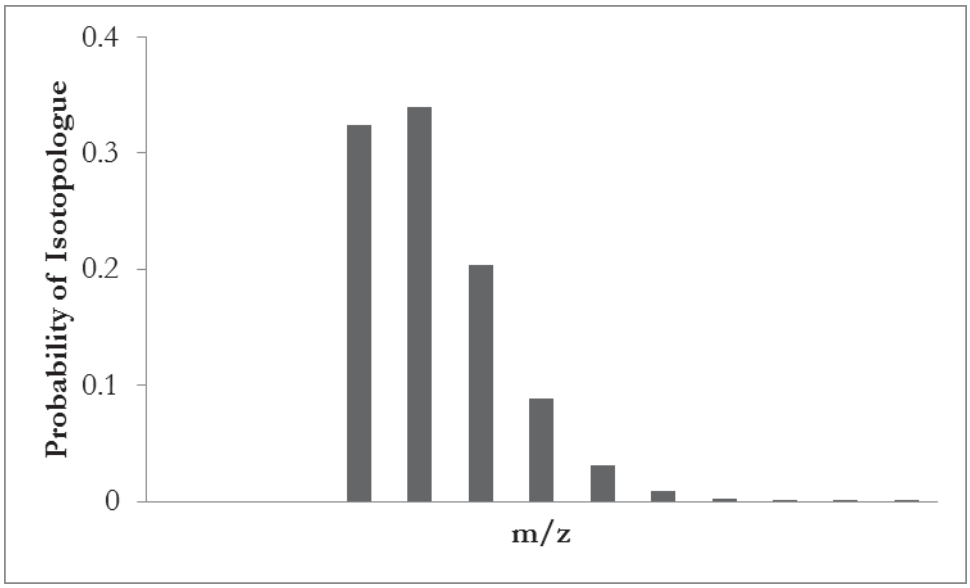


Figure 10: Isotopic Distribution for Natural Isotopic Abundance Dimer Pentapeptide with No Modifications.

Because the isotopic enrichment of labeled lysine is less than 100%, a second distribution arises from the partial enrichment for labeled species. Since the probability distribution of discrete events occurring with a fixed probability can be modeled using a binomial distribution, the distribution arising from isotopic enrichment can also be modeled in this way. The number of trials in this binomial distribution is the maximum number of isotopically enriched atoms, and the probability of success is the isotopic enrichment percentage. The binomial equation is shown below, where P represents the probability of a particular isotopologue within the distribution, N represents the theoretical maximum number of labels, k represents the number of labels incorporated, and E represents the probability of isotopic enrichment (Formula 3).

Formula 3: Binomial Equation for the Probability of Each Isotopologue

$$P = \frac{N!}{k!\,(N-k)!} E^k (1-E)^{N-k}$$

The isotopic and binomial probability distributions were convoluted to produce isotopic distributions for heavy species corrected for partial isotopic enrichment. The probability of each mass in the isotopic distribution was multiplied by the matrix representing the binomial distribution. The resulting probability matrices for every mass from the isotopic distribution were binned and summed according to their new masses (Figure 11). The convoluted probability matrix predicts the observed isotopic distributions for labeled species (combined array). The predicted isotopic distribution for labeled species is shown in Figure 12.

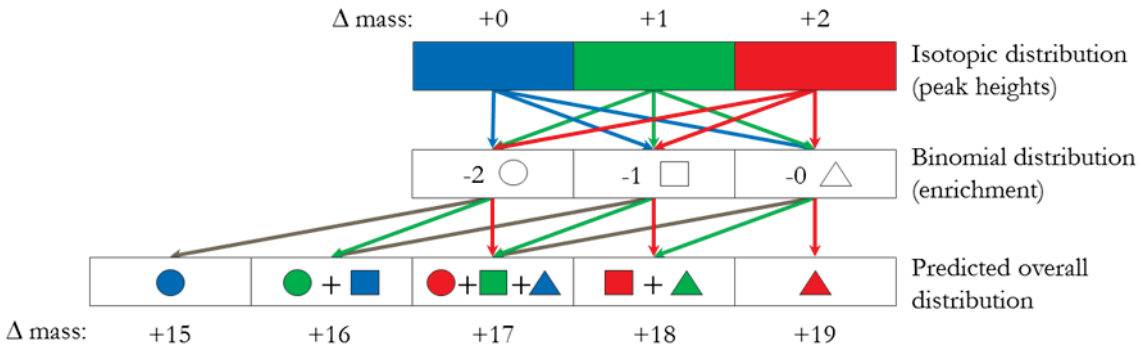


Figure 11: Convolution of Isotopic Distribution and Binomial (Enrichment) Distribution.

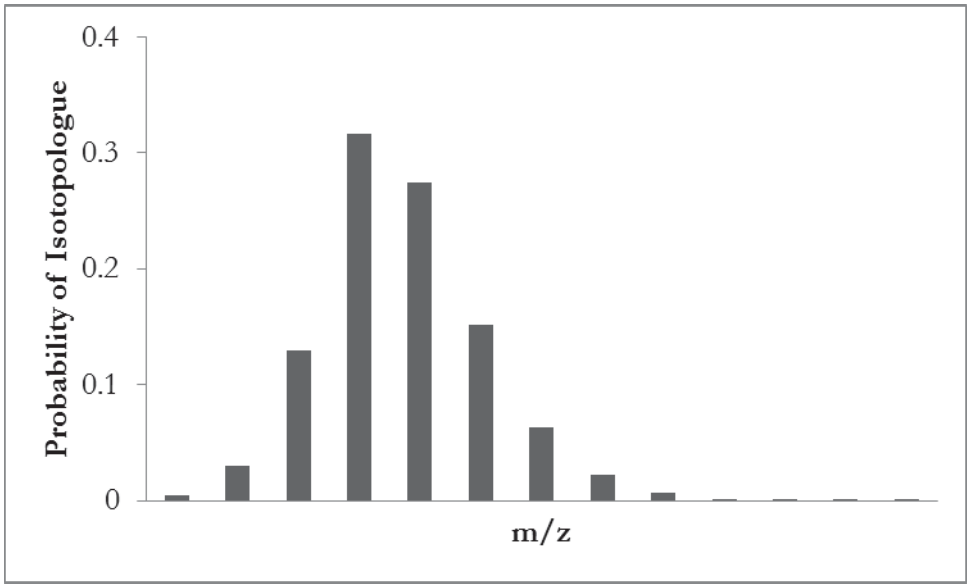


Figure 12: Predicted Isotopic Distribution for Heavy Lysine Labeled Dimer Pentapeptide with no Modifications.

Matching and Isotopic Correction

The *in silico* generated library and distributions were used to identify corresponding peaks on the mass spectrum, and the associated intensities were corrected for the isotopic distribution.

The observed m/z values for each species were recorded and corrected for Glu-Fib using the ratio between observed and theoretical Glu-Fib m/z values. The corrected m/z

values and their associated charge states were used to calculate the accurate mass of each ionized species. Then all exact mass entries within 50 ppm of each accurate mass were identified (match).

The observed intensity of the most abundant peak for each matched species was corrected for the isotopic distribution. Since the calculated isotopic distribution represents the theoretical intensity as a probability, the intensity for the corresponding monoisotopic m/z value was calculated based on the probability of finding the species at the m/z value of the most abundant peak (Formula 4).

Formula 4: Intensity Corrected for Isotopic Distribution

$$intensity_{monoisotopic} = \frac{intensity_{most\ abundant\ peak}}{P(most\ abundant\ peak\ m/z)}$$

Correction for Partial Incorporation of Peptidoglycan with Isotopic Label

Before comparing the corrected intensities for peptidoglycan composition of bacteria grown and labeled under different conditions, the intensities of the peaks were corrected based on partial lysine incorporation.

Calculating Labeling Efficiency

In order to find the lysine incorporation ratio, the intensities of complete and partial lysine incorporated species were compared after correcting for the isotopic distribution. Every permutation of complete or partial lysine incorporation can be represented using the binomial coefficient, where N is the theoretical maximum number of lysine incorporated and k is the actual number of lysine incorporated. P is the probability of observing a particular

number of incorporated lysine, and $P_{enriched}$ is the probability of labeled lysine incorporation at a particular position (Formula 5).

Formula 5: Probability of Observing a Species with a Particular Number of Labels

$$P = \binom{N}{k} {P_{enriched}}^{k} (1 - P_{enriched})^{N-k}$$

Since intensity is directly proportional to probability, the ratio between the probabilities of any two permutations is equal to the observed ratio of corrected intensities corresponding to these permutations. This ratio, r, for a dimer with heavy lysine label is shown below, in terms of the probability of labeled lysine incorporation at a particular position, P(H) (Formula 6).

Formula 6: Ratio of Probabilities and Intensities

$$r = \frac{P(H)^2}{2 \cdot P(H) \cdot (1 - P(H))} = \frac{intensity_{HH}}{intensity_{HL}}$$

Solving for P(H) yields the following expression, and the probability of heavy lysine incorporation at a particular position can be calculated based on the ratio of observed intensities (Formula 7).

Formula 7: Probability of Label Incorporation at a Single Position for a Dimer

$$P(H) = \frac{2r}{1 + 2r}$$

Correcting Intensities

Using the calculated labeling efficiency, the intensities of completely labeled and completely unlabeled species were corrected.

Because of partial incorporation, the completely labeled peak is less intense than it would be for full lysine incorporation, and so its intensity must be corrected to account for the partially labeled and unlabeled species from that sample. For a dimer, the probability of full incorporation, P(HH), is equivalent to the probability of two incorporations, $P(H)^2$ (Formula 8).

Formula 8**:** Probability of Label Incorporation at Both Positions for a Dimer

$$P(HH) = [P(H)]^2 = \left(\frac{2r}{1+2r}\right)^2$$

For the completely labeled peak, the ratio of corrected intensity to observed intensity is proportional to the ratio of theoretical full incorporation to actual incorporation. Solving for corrected intensity yields the following expression (Formula 9).

Formula 9: Corrected Intensity of Fully Labeled Dimer with Partial Incorporation

$$I_{corrected} = \frac{I_{observed}}{\left(\frac{2r}{1+2r}\right)^2}$$

Since labeled and unlabeled samples are mixed in equal proportions for SILAC, the contribution of unlabeled species from the labeled sample must be accounted for. The probability of a completely unlabeled peak in a labeled sample, $P_H(LL)$, is equal to the

probability of two non-incorporations, $P_H(L)^2$. Given that $P_H(L)$ is equal to 1 – P(H), $P_H(LL)$ can be expressed in terms of r (Formula 10).

Formula 10: Probability of Dimer with No Label Incorporation in Labeled Sample

$$P_H(LL) = [1 - P(H)]^2 = \left(1 - \frac{2r}{1+2r}\right)^2$$

The observed intensity for the completely unlabeled species must be corrected by subtracting the contribution of completely unlabeled species in labeled sample (Formula 11).

Formula 11: True Intensity of Dimer from Unlabeled Sample, Corrected for Contribution from Labeled Sample

$$I_{corrected}^{unlabeled} = I_{observed}^{unlabeled} - I_{observed}^{labeled} \frac{\left(1 - \frac{2r}{1+2r}\right)^2}{\left(\frac{2r}{1+2r}\right)^2}$$

Applying Corrections

After the intensities were corrected for isotopic distribution and partial lysine incorporation, the corrected intensity for each labeled species was compared to the corrected intensity for the corresponding unlabeled species. This procedure was performed in order to observe the change in peptidoglycan composition as bacteria forms biofilm.

CHAPTER THREE

Results

Bacterial Growth and Collection

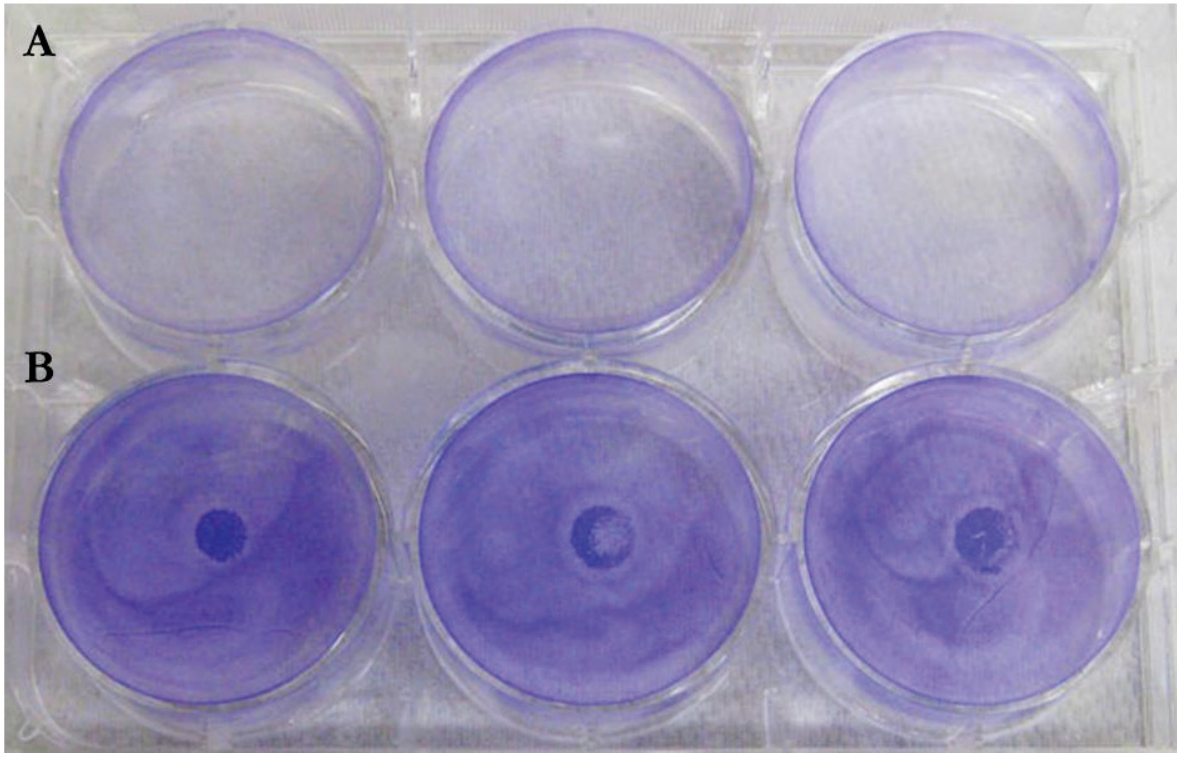


Figure 13: Growth and Harvest of *E. faecalis* Biofilm. Row A shows crystal violet stained wells after harvesting biofilm. Row B shows stained wells without harvesting.

After 24 hours of growth in favorable conditions, *E. faecalis* bacteria were observed in three forms. First, bacteria were found suspended in the media. Second, bacteria were observed in clusters that were either free-floating or loosely attached to the plate surface. The suspended and flocculent bacteria were collected with the media. Any remaining loose bacteria was removed by a gentle PBS rinse. Third, bacteria formed a thin, adhesive biofilm on the bottom and side surfaces of the wells, and this milky white film was not removed by

the PBS rinse. The layer of biofilm varied in density due to orbital shaking, with greater density of growth in the middle and side walls of the wells. Bacteria in the biofilm phase was loosened mechanically by scraping and removed with an additional PBS. Crystal violet staining was used to visualize the growth and collection of biofilm (Figure 13), which was collected in significant amounts at the time point described in the methods section.

PG Library and Matches

Using a bottom-up proteomics approach, an in-house MATLAB program was used to generate a library of peptidoglycan (PG) fragments with various combinations of acetylation, alanylation, and crosslinking. In order to include the most abundant and relevant species and to calculate the correct mass for these species, molecular structures and processes were considered during library generation.

First, although PG normally exhibits sugar-linking, enzymatic digestion by mutanolysin breaks these bonds. Since enzymatic digestion with excess mutanolysin was carried out for 24 hours, it can be assumed that no appreciable amount of sugar-linked species remain. Thus, sugar-linking was not considered in library generation.

Second, a typical crosslink involves the connection of two PG units by an L-Ala-L-Ala crosslinking bridge (Figures 7-8). During bridge formation, the terminal D-Ala of the crosslinked PG stem is removed, and so the mass of the predicted species is adjusted accordingly (Lupoli et al., 2011).

Third, O-acetylation and N-deacetylation occur on NAM and NAG, respectively. The position of these edits will not affect the results observed in the mass spectrum, but this may be an interesting area for future study (Moynihan et al., 2014).

Fourth, alanylation depends on the combined activities of BppA enzymes and carboxypeptidases. BppA1 and BppA2 are responsible for adding the first and second L-Ala of the crosslinking bridge (Bouhss et al., 2001)(Bouhuss et al., 2002). Carboxypeptidases can cleave either the first (D,D-carboxypeptidase) or the second (L,D-carboxypeptidase) terminal D-Ala from the peptide stem. Activity or inactivity of these two carboxypeptidases and two BppA enzymes gives rise to five different alanylation states in *E. faecalis* peptidoglycan.

With the parameters described in Table 3, a total of 192 possible PG fragments were generated for species containing labeled lysine and unlabeled lysine (natural abundance). For 111 of these species, matches were identified in the LC-MS data (Table 4 and Figure 14). Only peaks with intensities above 1000 units and ppm differences below 10 ppm were considered.

Table 3: Parameters for Library Generation. For crosslinking, the minimum value of 1 represents species with 1 crosslink (i.e. dimers), and the maximum value of 4 represents species with 4 crosslinks (i.e. pentamers). Acetylation values range from -1 to +1, with -1 indicating a net of one N-deacetylation and +1 indicating a net of one O-acetylation. An alanylation value of -2 indicates the absence of four alanine, with two missing from the peptide stem and two missing from the crosslinking bridge. An alanylation value of 2 represents a pentapeptide in which no alanine are missing from the peptide stem or crosslinking bridge.

	Minimum	Maximum
Crosslinking	1	4
Acetylation	-1	+1
Alanylation	-2	2

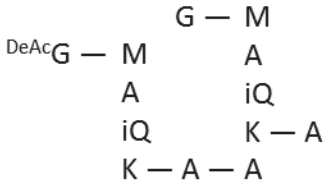
Table 4: List of Peptidoglycan Fragment Matches Observed. A heavy label indicates that a species contains labeled lysine, while a light label indicates species with lysine in natural abundance. The ID letters correspond to the structures in Figure 14.

Crosslinks	Alanylation	Acetylation	Label	Charge	Exact Mass	ID
1	-2	0	light	2	1846.9015	E
1	-1	-1	light	2	1876.9354	A
1	-1	-1	light	3	1876.9354	A
1	-2	0	heavy	2	1881.0429	E
1	-2	1	light	2	1888.9121	J
1	-1	-1	heavy	2	1911.0767	A
1	-1	-1	heavy	3	1911.0767	A
1	-1	0	light	2	1917.9387	F
1	-1	0	light	3	1917.9387	F
1	-2	1	heavy	2	1923.0535	J
1	0	-1	light	2	1947.9725	B
1	0	-1	light	3	1947.9725	B
1	-1	0	heavy	2	1952.0800	F
1	-1	0	heavy	3	1952.0800	F
1	-1	1	light	2	1959.9492	K
1	-1	1	light	3	1959.9492	K
1	0	-1	heavy	2	1982.1139	B
1	0	-1	heavy	3	1982.1139	B
1	0	0	light	2	1988.9758	G
1	0	0	light	3	1988.9758	G
1	-1	1	heavy	2	1994.0906	K
1	-1	1	heavy	3	1994.0906	K
1	1	-1	light	2	2019.0096	C
1	1	-1	light	3	2019.0096	C
1	0	0	heavy	2	2023.1171	G
1	0	0	heavy	3	2023.1171	G
1	0	1	light	2	2030.9863	L
1	0	1	light	3	2030.9863	L
1	1	-1	heavy	2	2053.1510	C
1	1	-1	heavy	3	2053.1510	C
1	1	0	light	2	2060.0129	H
1	1	0	light	3	2060.0129	H
1	0	1	heavy	2	2065.1277	L
1	0	1	heavy	3	2065.1277	L

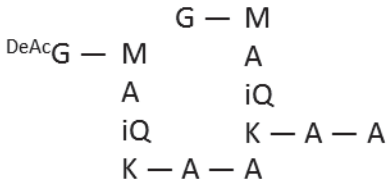
Crosslinks	Alanylation	Acetylation	Label	Charge	Exact Mass	ID
1	2	-1	light	2	2090.0467	D
1	2	-1	light	3	2090.0467	D
1	1	0	heavy	2	2094.1543	H
1	1	0	heavy	3	2094.1543	H
1	1	1	light	2	2102.0234	M
1	1	1	light	3	2102.0234	M
1	2	-1	heavy	2	2124.1881	D
1	2	-1	heavy	3	2124.1881	D
1	2	0	light	2	2131.0500	I
1	2	0	light	3	2131.0500	I
1	1	1	heavy	2	2136.1648	M
1	1	1	heavy	3	2136.1648	M
1	2	0	heavy	2	2165.1914	I
1	2	0	heavy	3	2165.1914	I
1	2	1	light	2	2173.0606	N
1	2	1	light	3	2173.0606	N
1	2	1	heavy	2	2207.2019	N
1	2	1	heavy	3	2207.2019	N
2	-1	-1	light	3	2897.4329	O
2	-1	0	light	3	2938.4362	P
2	-1	-1	heavy	3	2948.6450	O
2	-1	1	light	3	2980.4467	T
2	-1	0	heavy	3	2989.6482	P
2	0	0	light	2	3009.4733	Q
2	0	0	light	3	3009.4733	Q
2	-1	1	heavy	3	3031.6588	T
2	0	1	light	3	3051.4839	U
2	0	0	heavy	2	3060.6854	Q
2	0	0	heavy	3	3060.6854	Q
2	1	0	light	2	3080.5104	R
2	1	0	light	3	3080.5104	R
2	0	1	heavy	3	3102.6959	U
2	1	1	light	3	3122.5210	V
2	1	0	heavy	2	3131.7225	R

Crosslinks	Alanylation	Acetylation	Label	Charge	Exact Mass	ID
2	1	0	heavy	3	3131.7225	R
2	2	0	light	2	3151.5475	S
2	2	0	light	3	3151.5475	S
2	1	1	heavy	3	3173.7330	V
2	2	1	light	3	3193.5581	W
2	2	0	heavy	2	3202.7596	S
2	2	0	heavy	3	3202.7596	S
2	2	1	heavy	3	3244.7701	W
3	-1	0	light	4	3958.9337	X
3	-1	0	heavy	4	4027.2165	X
3	0	0	light	3	4029.9708	Y
3	0	0	light	4	4029.9708	Y
3	0	0	heavy	3	4098.2536	Y
3	0	0	heavy	4	4098.2536	Y
3	1	0	light	3	4101.0079	Z
3	1	0	light	4	4101.0079	Z
3	1	0	heavy	3	4169.2907	Z
3	1	0	heavy	4	4169.2907	Z
3	2	0	light	3	4172.0450	AA
3	2	0	light	4	4172.0450	AA
3	2	0	heavy	3	4240.3278	AA
3	2	0	heavy	4	4240.3278	AA
4	0	0	light	4	5050.4683	BB
4	1	0	light	4	5121.5054	CC
4	0	0	heavy	4	5135.8218	BB
4	2	0	light	4	5192.5426	DD
4	1	0	heavy	4	5206.8589	CC
4	2	0	heavy	4	5277.8960	DD

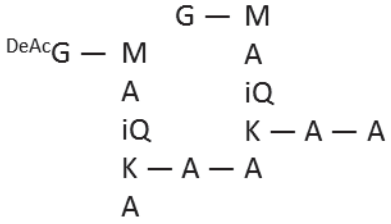
(A)



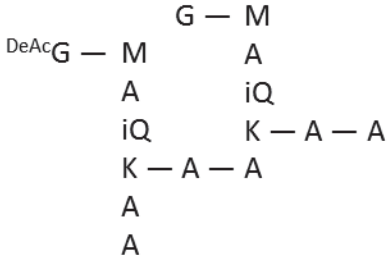
(B)



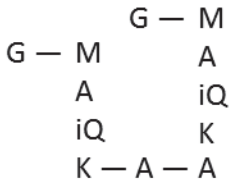
(C)



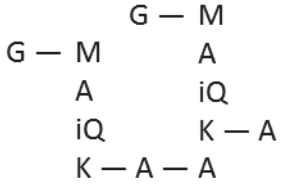
(D)



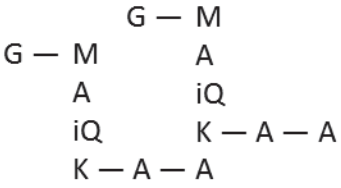
(E)



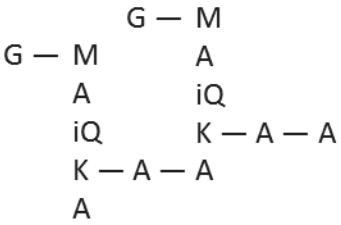
(F)



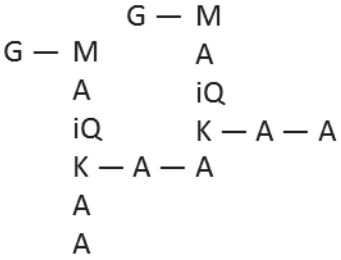
(G)



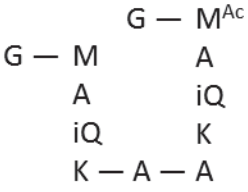
(H)



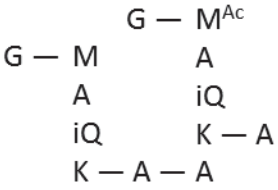
(I)



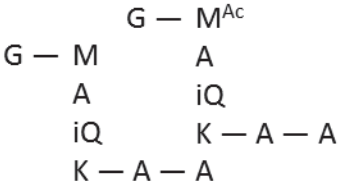
(J)



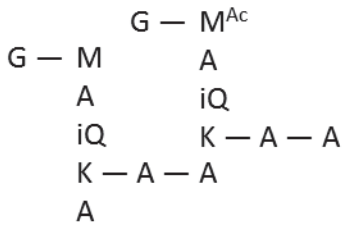
(K)



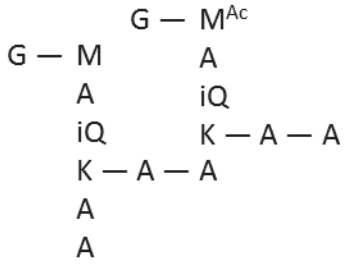
(L)



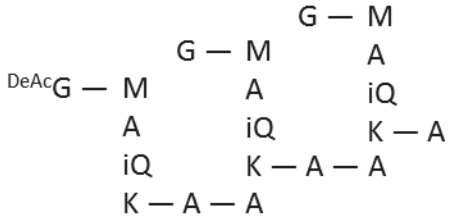
(M)



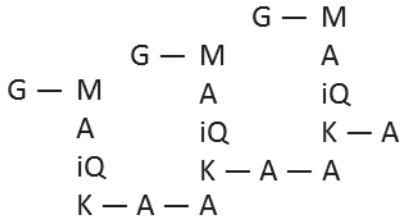
(N)



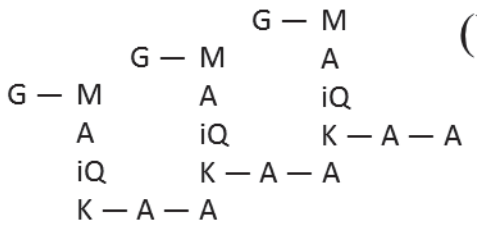
(O)



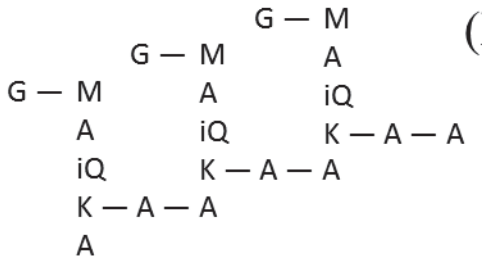
(P)



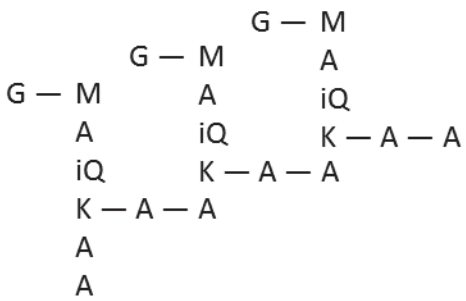
(Q)



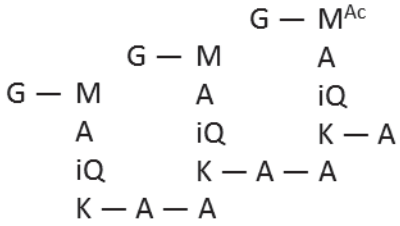
(R)



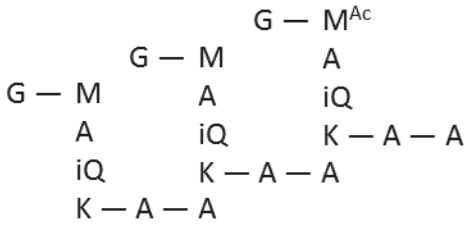
(S)



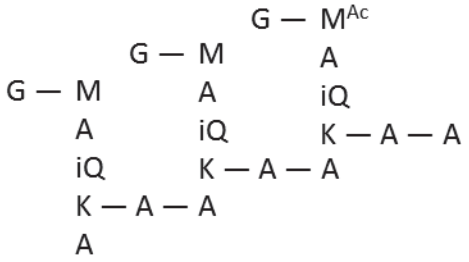
(T)



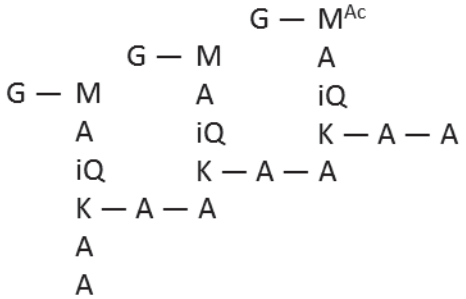
(U)



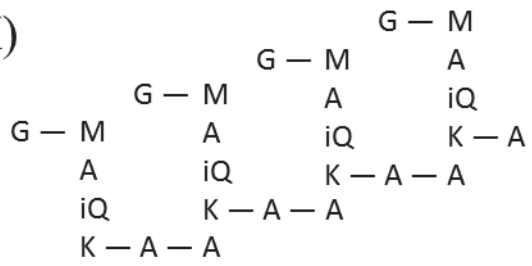
(V)



(W)



(X)



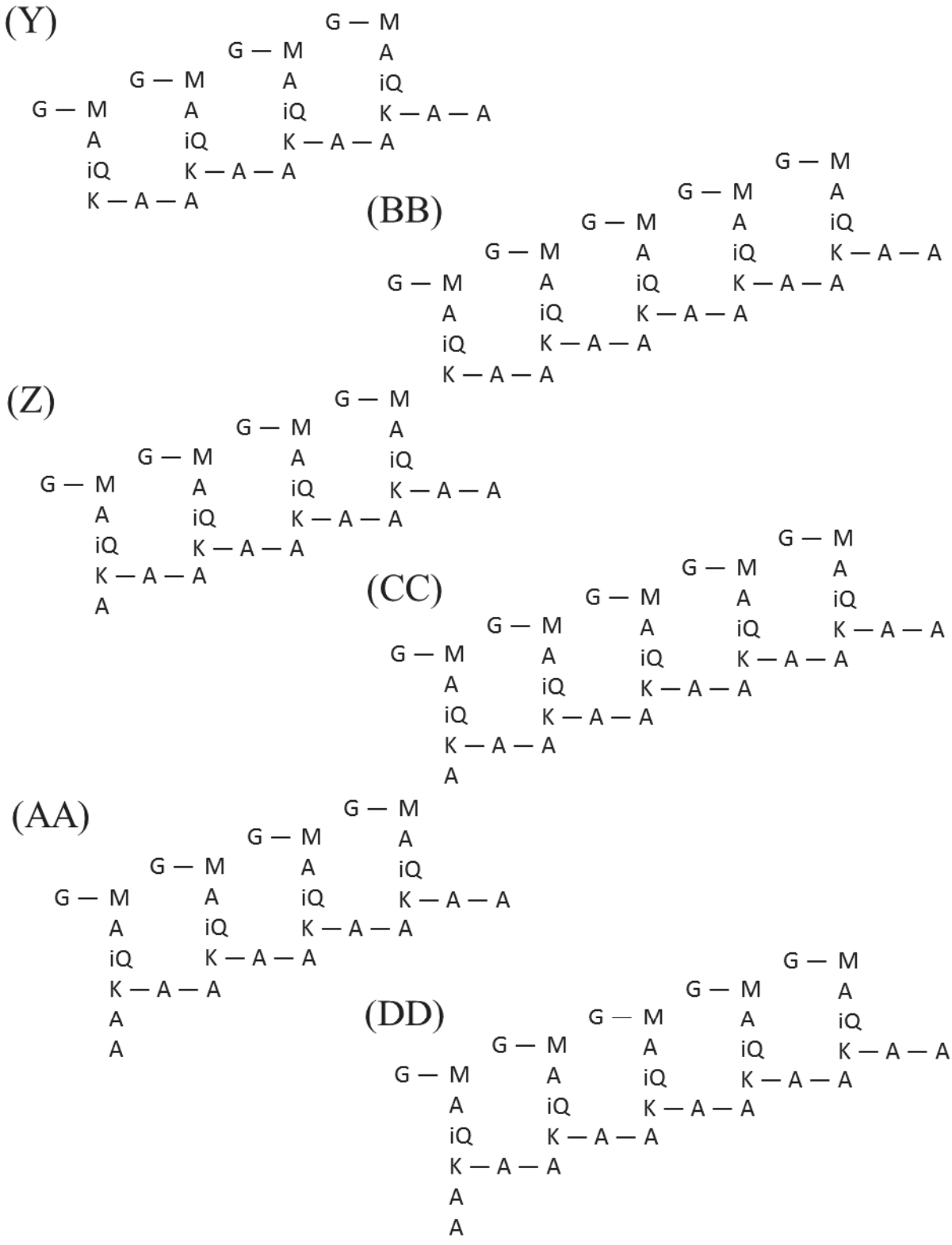


Figure 14: Structures of Matched Peptidoglycan Fragments. G and M represent the NAG and NAM. A, iQ, and K represent their respective amino acids: alanine, iso-glutamine, and lysine. Acetylation and deacetylation are indicated by the superscripts Ac and DeAc.

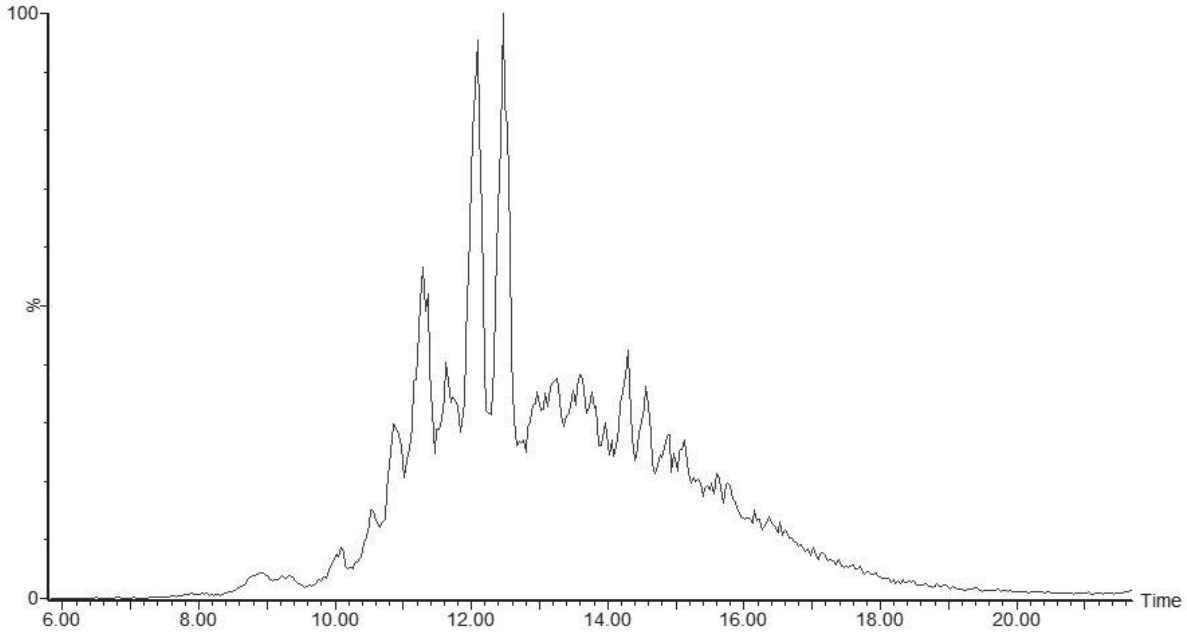


Figure 15: Select Ion Chromatogram (SIC) for NAG, Indicating Elution of Peptidoglycan Species.

Figure 15 shows the SIC for NAG, which is commonly broken from PG fragments during ionization. Thus, the elution profile of NAG is indicative of when all PG species will elute, and the SIC suggests that the majority of PG species elute between 11 and 15 minutes after the injection of sample. The sharp peaks between 12 and 13 minutes further indicate the time at which PG fragments are most abundantly observed.

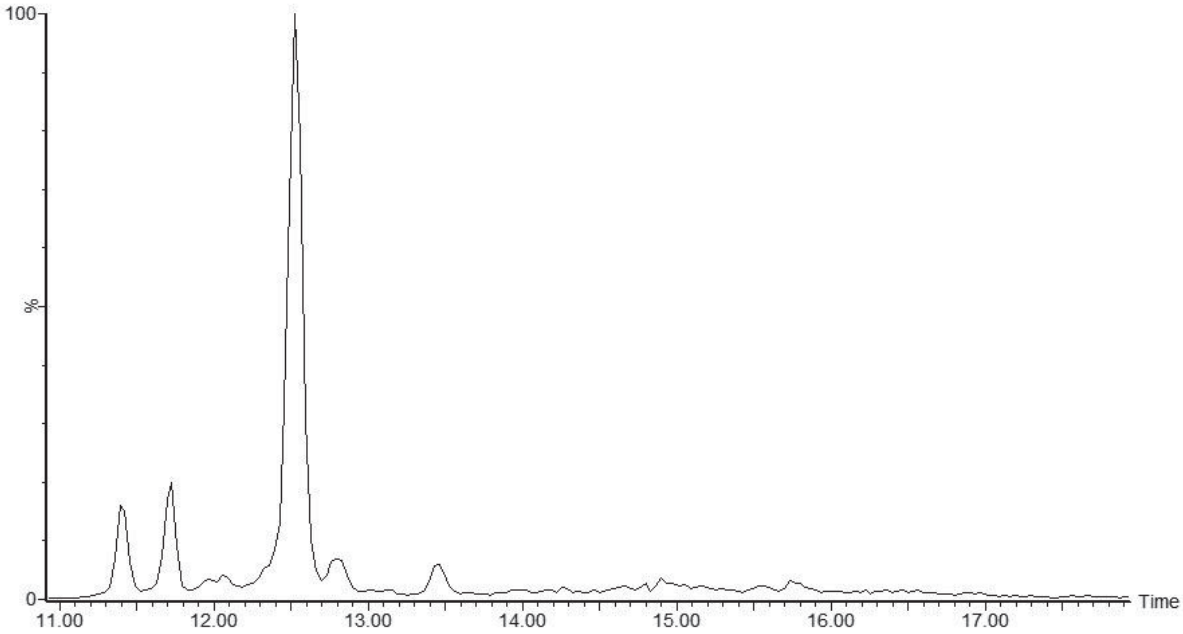


Figure 16: SIC for Dimer Pentapeptide.

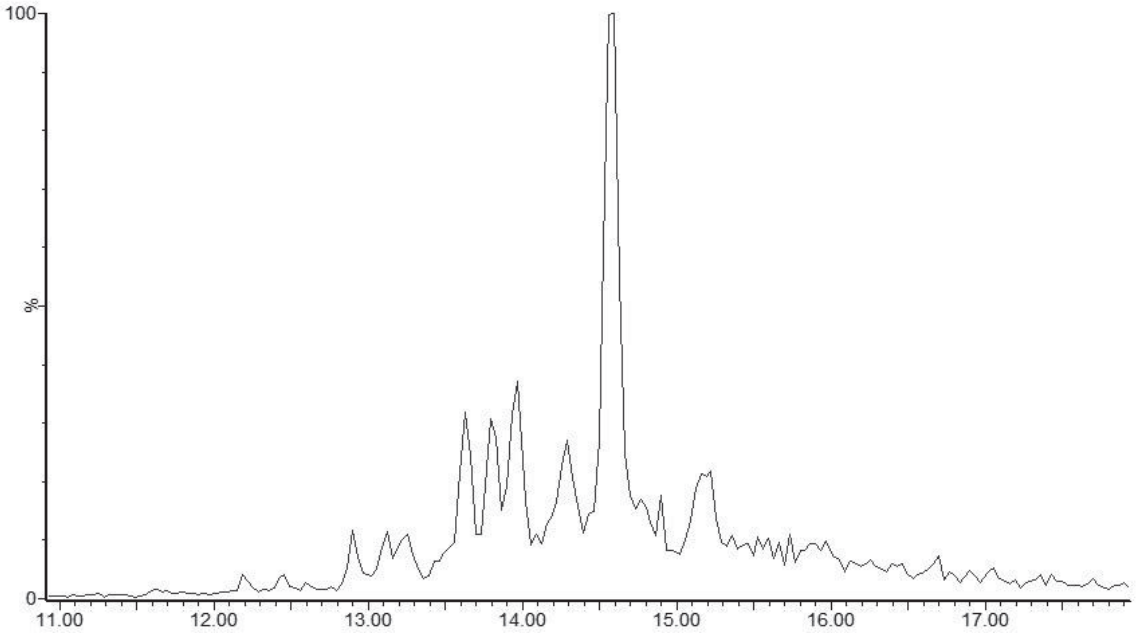


Figure 17: SIC for Trimer Pentapeptide.

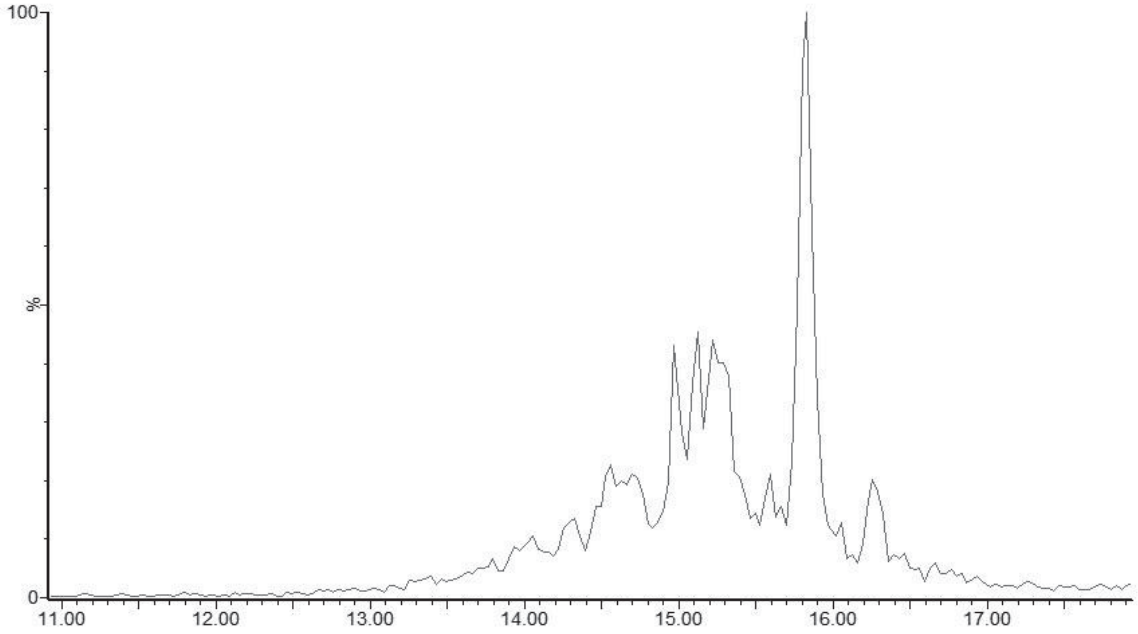


Figure 18: SIC for Tetramer Pentapeptide.

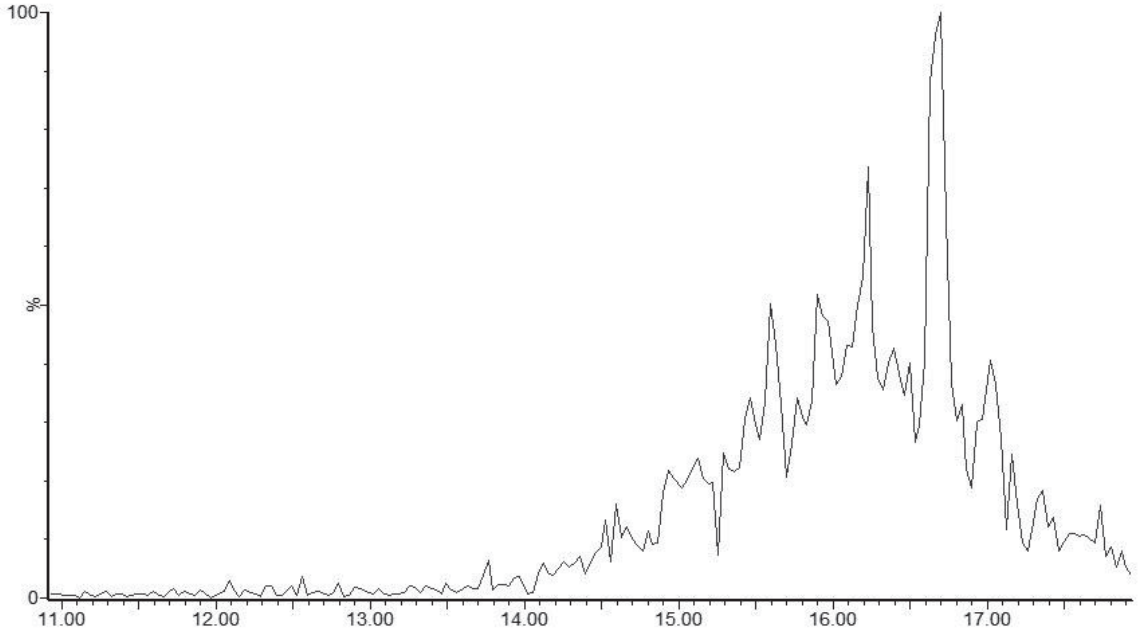


Figure 19: SIC for Pentamer Pentapeptide.

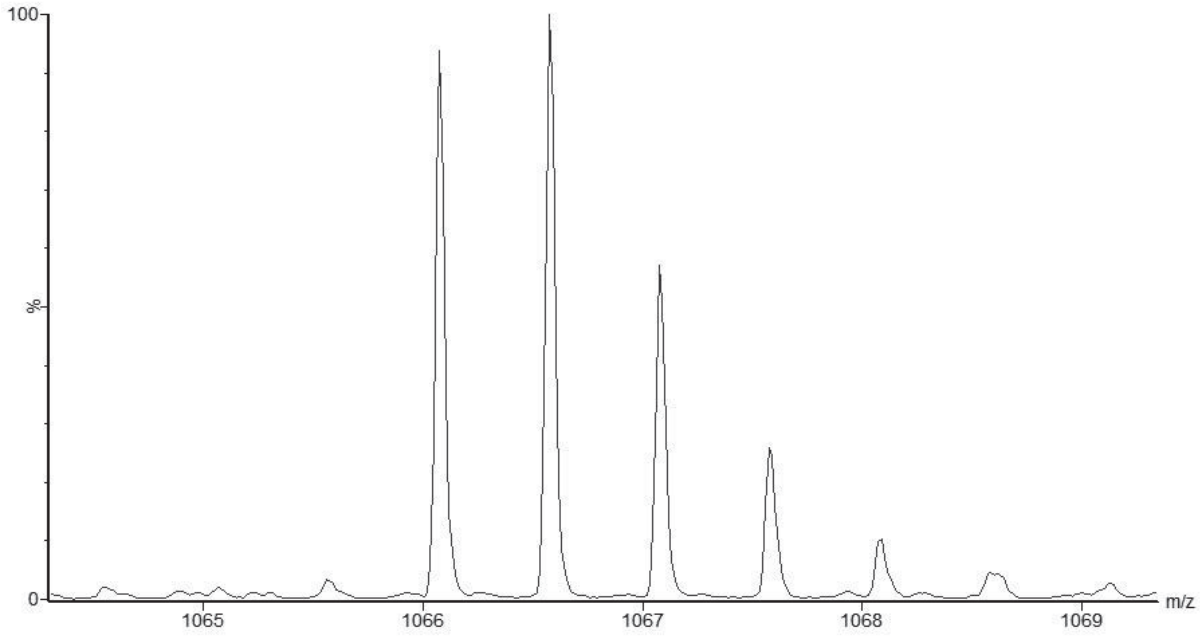


Figure 20: Mass Spectrum for Dimer Pentapeptide.

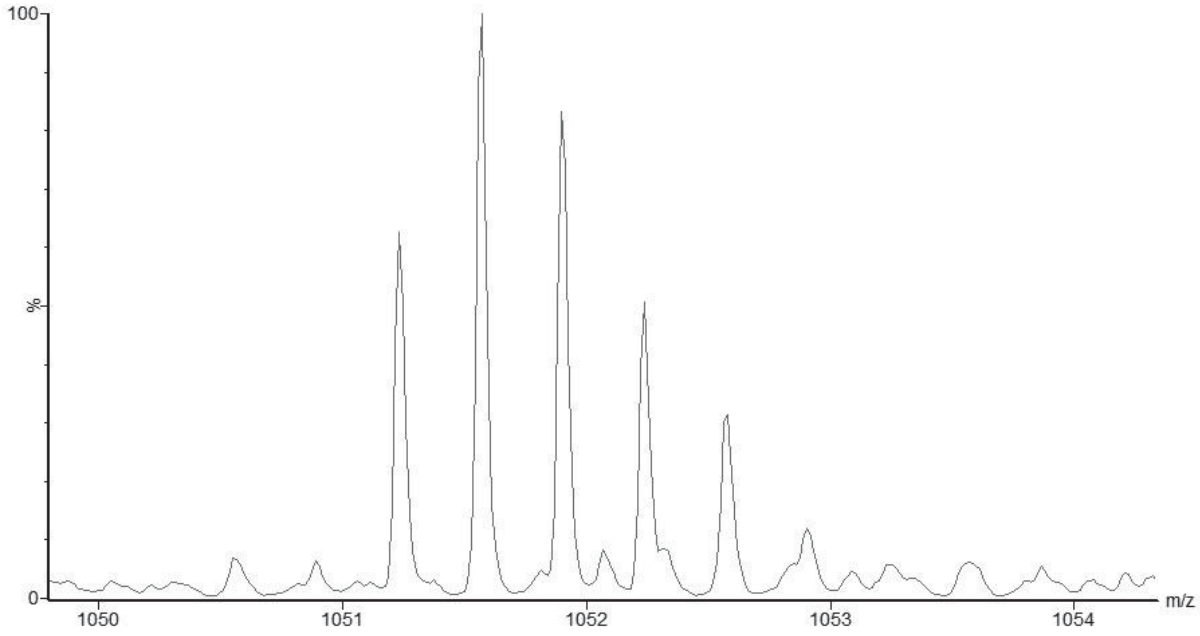


Figure 21: Mass Spectrum for Trimer Pentapeptide.

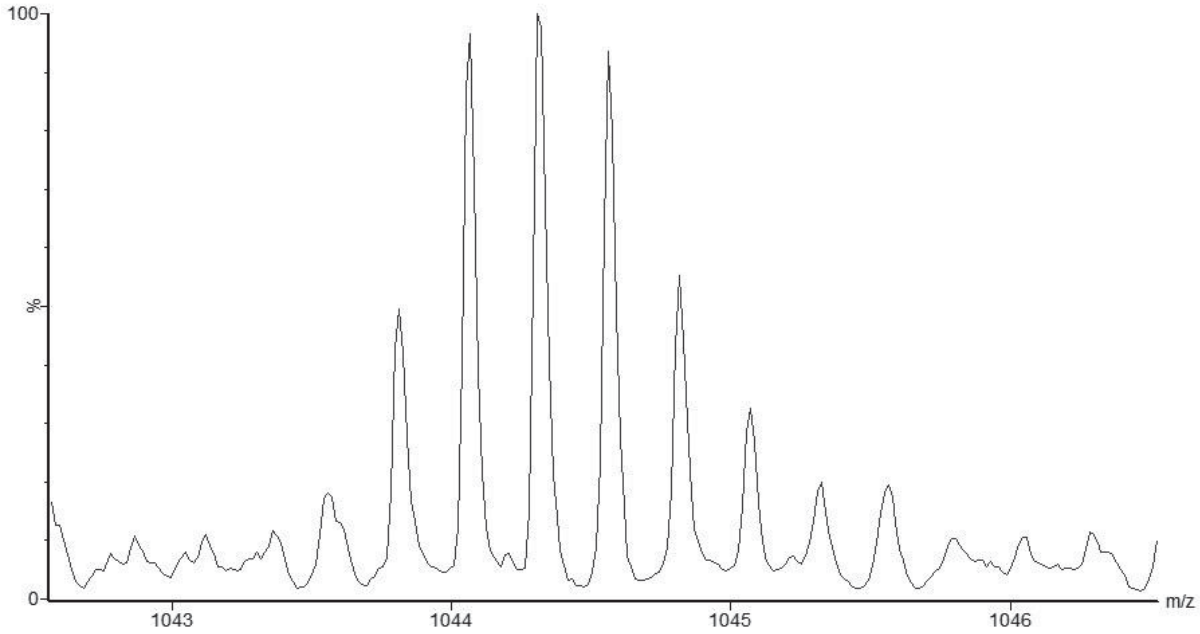


Figure 22: Mass Spectrum for Tetramer Pentapeptide.

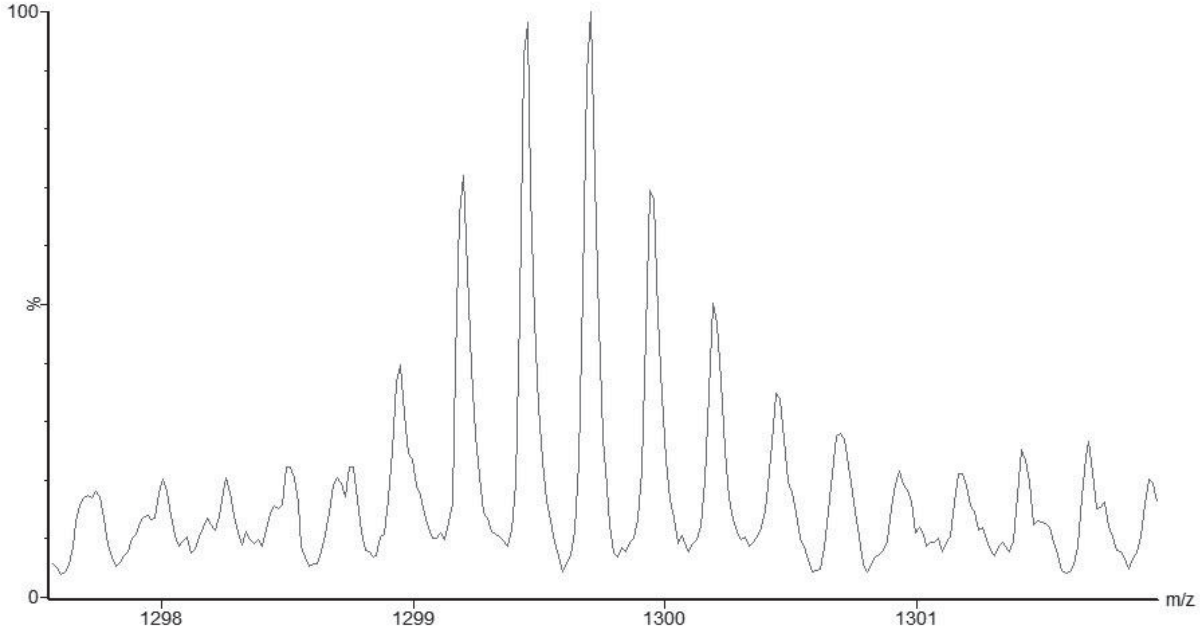


Figure 23: Mass Spectrum for Pentamer Pentapeptide.

Figures 16-19 show the SIC for unedited pentapeptides with varying degrees of crosslinking. The rightward shift of elution time as the number of crosslinks increases

demonstrates the slower elution time of larger species. Although each SIC does not have a single sharp peak corresponding to a particular elution time, each mass spectrum is taken over the time period that includes the most prominent SIC peaks (Figures 20-23). These spectra are scaled according to the most intense peak in each distribution, so the heights of the peaks in different spectra are not representative of actual intensities. In general, species with fewer crosslinks are found to be more abundant when intensities are compared after corrections for various factors. The distribution of peaks within each mass spectrum depends on the size of the PG fragment. For smaller fragments with less crosslinking, such as dimers and trimers (Figure 20-21), the second peak is the most intense peak. For larger fragments with more crosslinking (Figures 22-23), the third peak is the base peak. For larger fragments, the isotopic distribution is wider and spread more evenly between the peaks, and the base peak is less prominent. Therefore, before comparing the intensity of different species, the base peaks must be corrected according to the unique distribution of each species.

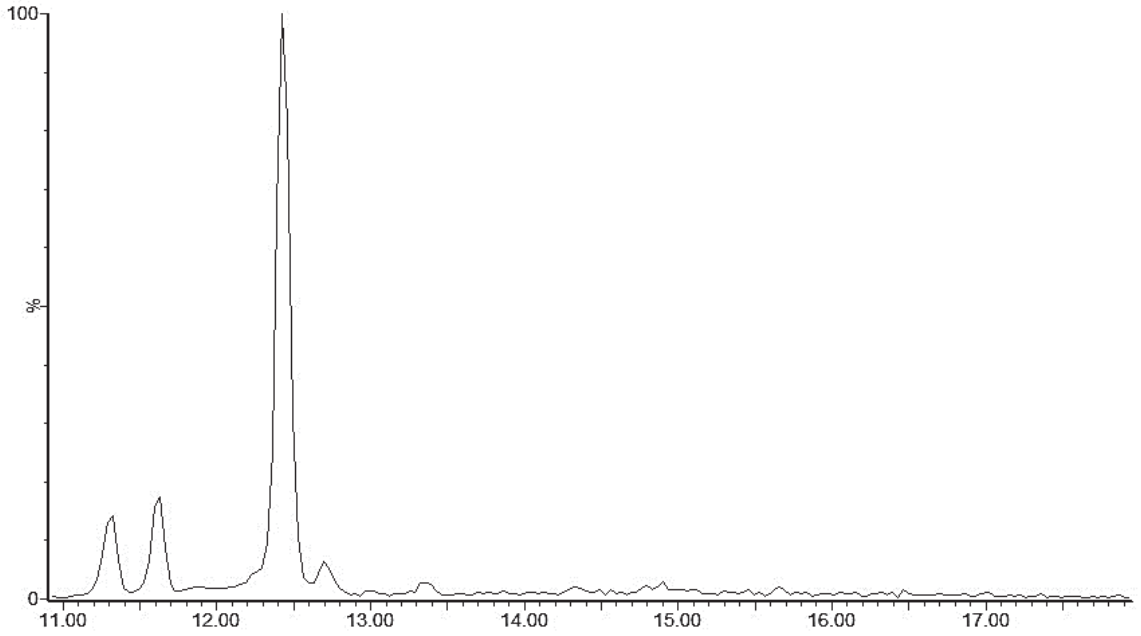


Figure 24: SIC for Isotopically Labeled Dimer Pentapeptide.

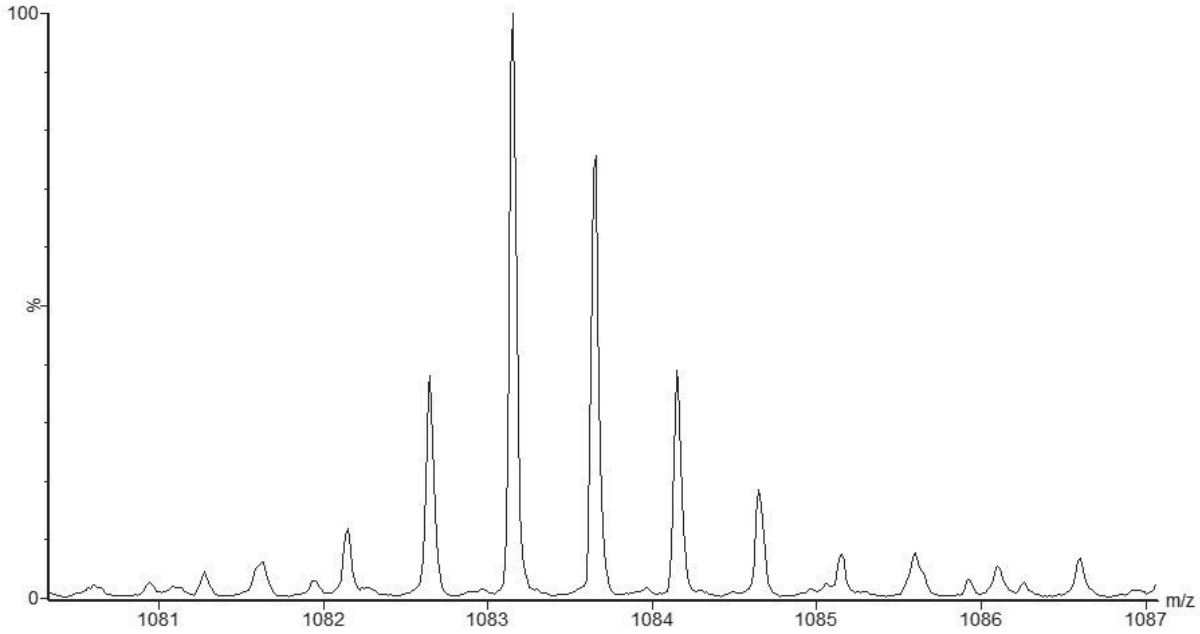


Figure 25: Mass Spectrum for Isotopically Labeled Dimer Pentapeptide.

Figures 24 and 25 show the SICs and mass spectra for isotopically labeled dimer pentapeptides. The labeled species elutes at about 12.4 minutes, as compared to the

unlabeled species (Figure 16), which elutes at about 12.6 minutes. Since the labeled species elutes more quickly than the unlabeled species due to the label's greater molecular mass, the chromatographic separation allows for the distributions of species to be visualized with minimal concerns about overlap in the mass spectrum. In the mass spectra for the labeled species, the isotopic distribution exhibits a tailing effect to the left of the base peak due to incomplete labeling of PG species. In this way, the distributions for light species are easily distinguished from those for heavy species.

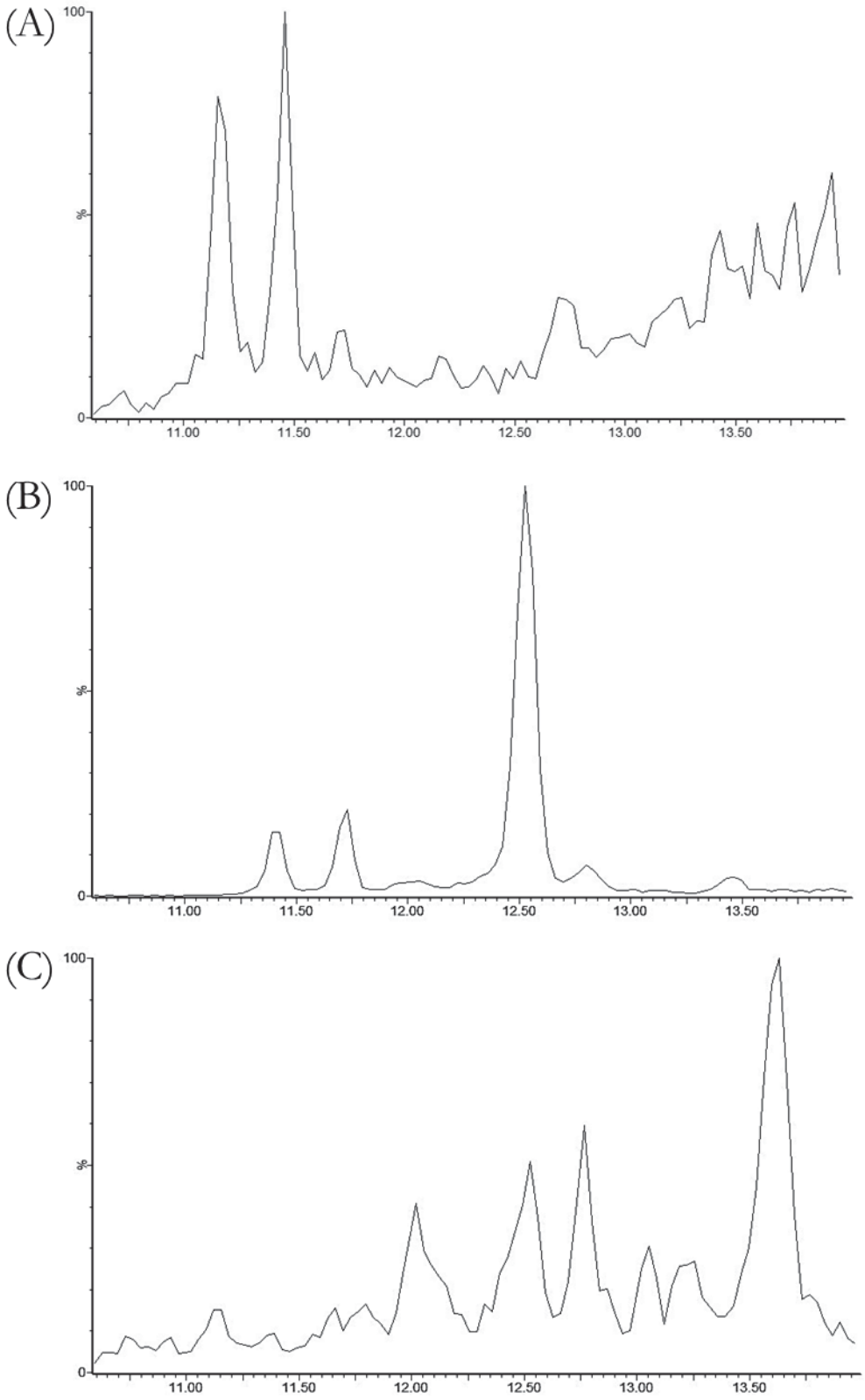


Figure 26: SICs for Dimer Pentapeptides at Different Acetylation States. (A) -1 Ac, (B) 0 Ac, (C) +1 Ac.

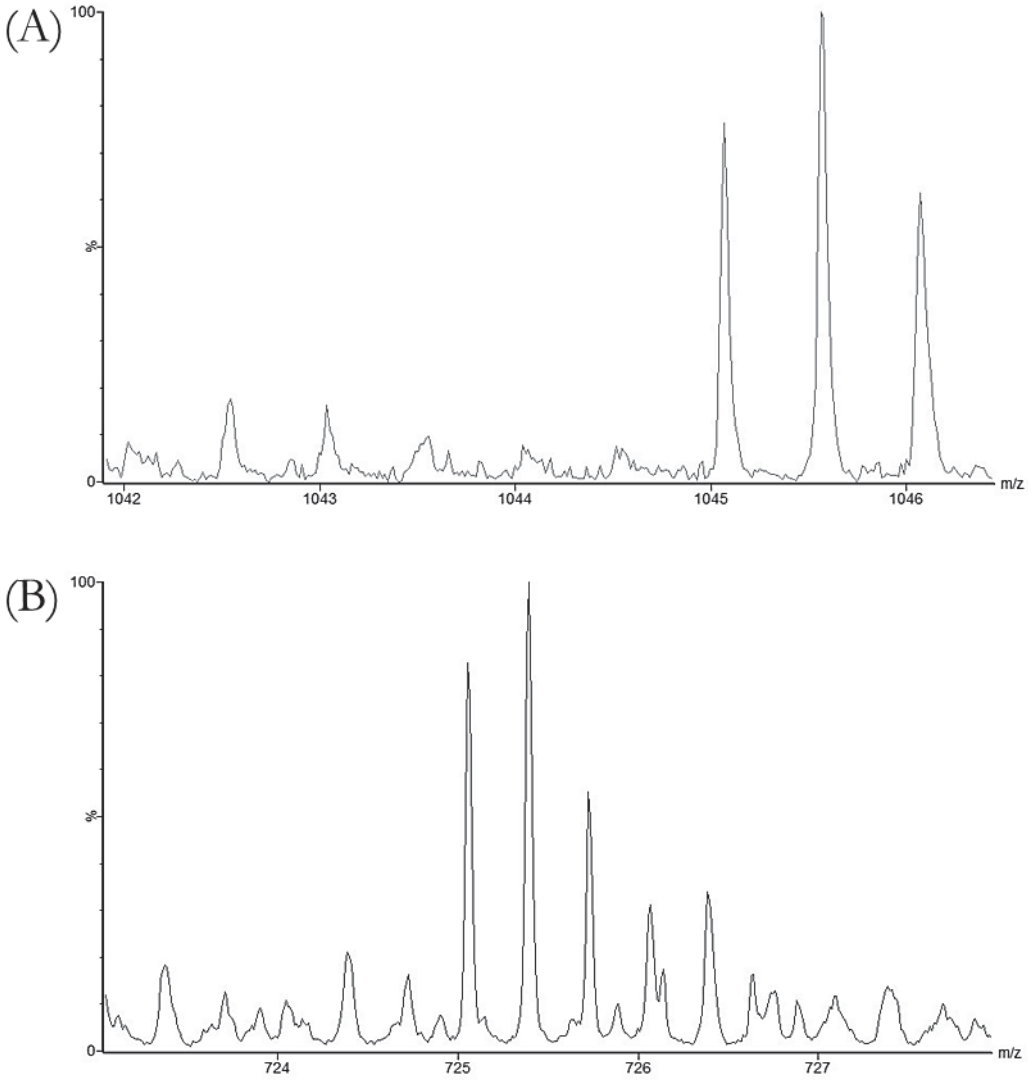


Figure 27: Spectra for Dimer Pentapeptides at Different Acetylation States. (A) -1 Ac, (B) +1 Ac. The spectrum for 0 Ac is shown in Figure 20.

Figures 26 and 27 show the SIC and spectra for dimer pentapeptides at various acetylation states. In terms of elution time, the SICs clearly demonstrate that N-deacetylation of NAG causes a species to elute more quickly and O-acetylation of NAM causes a species to elute more slowly. In general, species with an acetylation number of 0 Ac were found in greatest abundance, followed by those with +1 Ac and then -1 Ac.

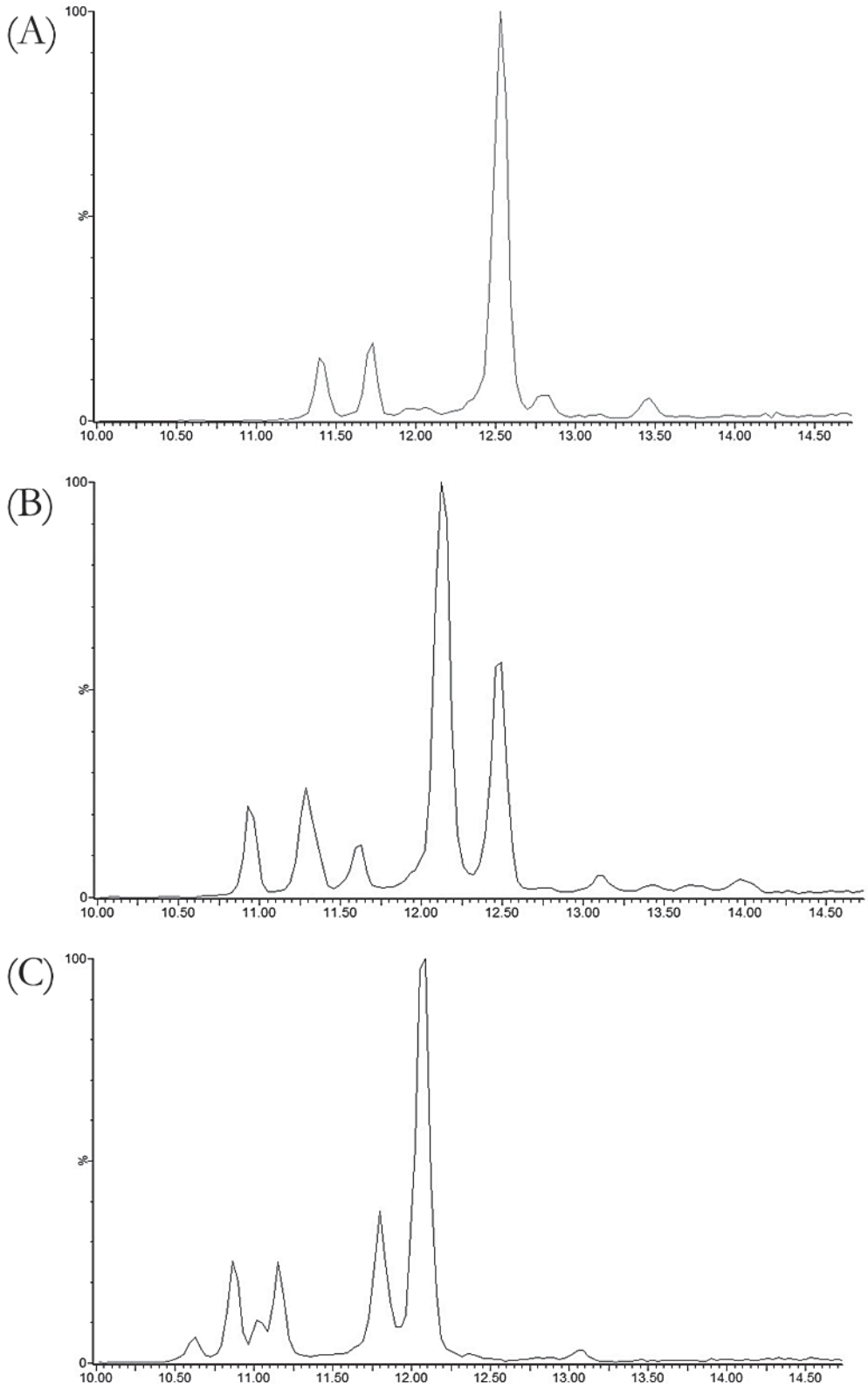


Figure 28: SICs for Dimers in Different Alanylation States. (A) Pentapeptide, (B) Tetrapeptide, (C) Tripeptide.

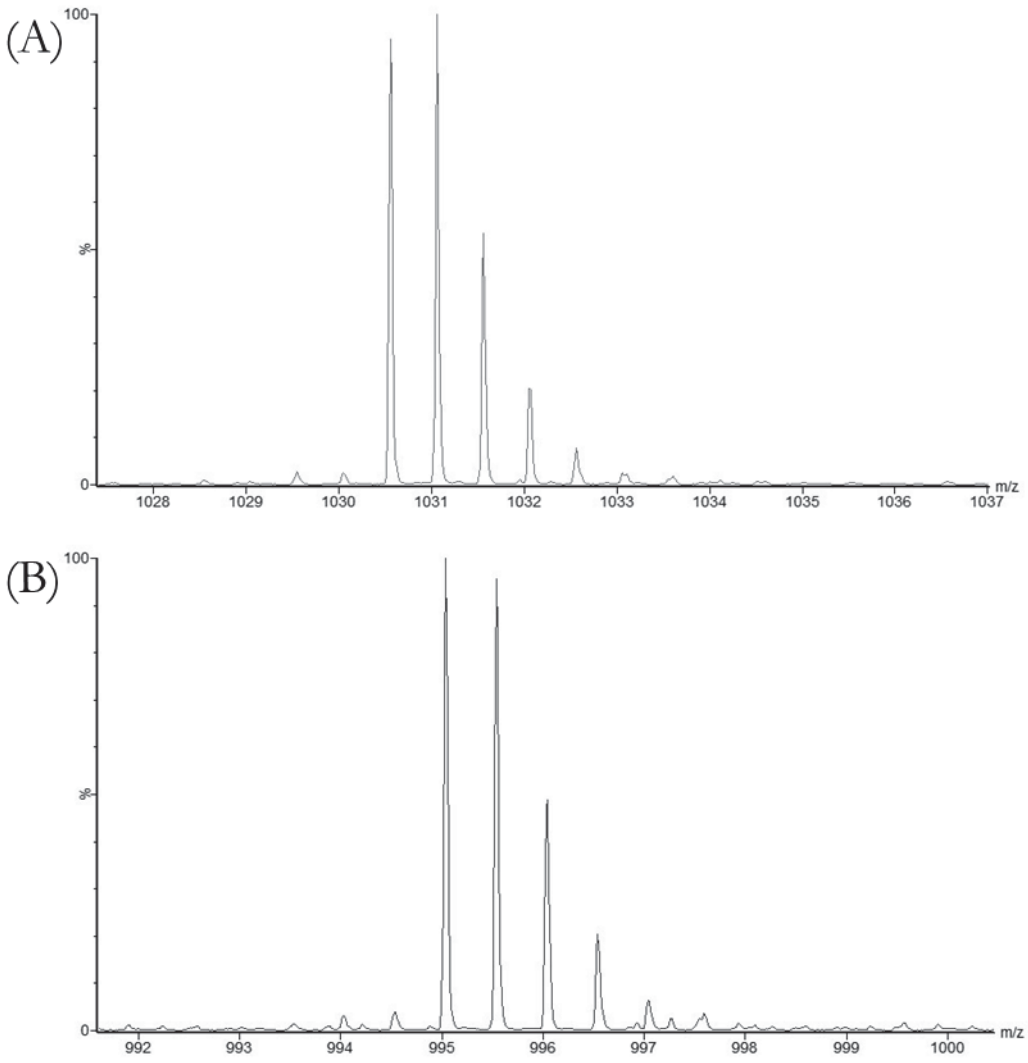


Figure 29: Spectra for Dimers in Different Alanylation States. (A) Tetrapeptide, (B) Tripeptide. Pentapeptide is shown in Figure 20.

Figures 28 and 29 show the SIC and mass spectra for dimers in different alanylation states due to carboxypeptidase editing. The trend in elution time shows that the smaller tripeptides elute more quickly than the larger pentapeptides. The distributions of peaks in the mass spectra for tetrapeptides and tripeptides follow the trend that is expected based on PG fragment size; in fact, the first peak is the base peak for dimer tripeptides because of their smaller size.

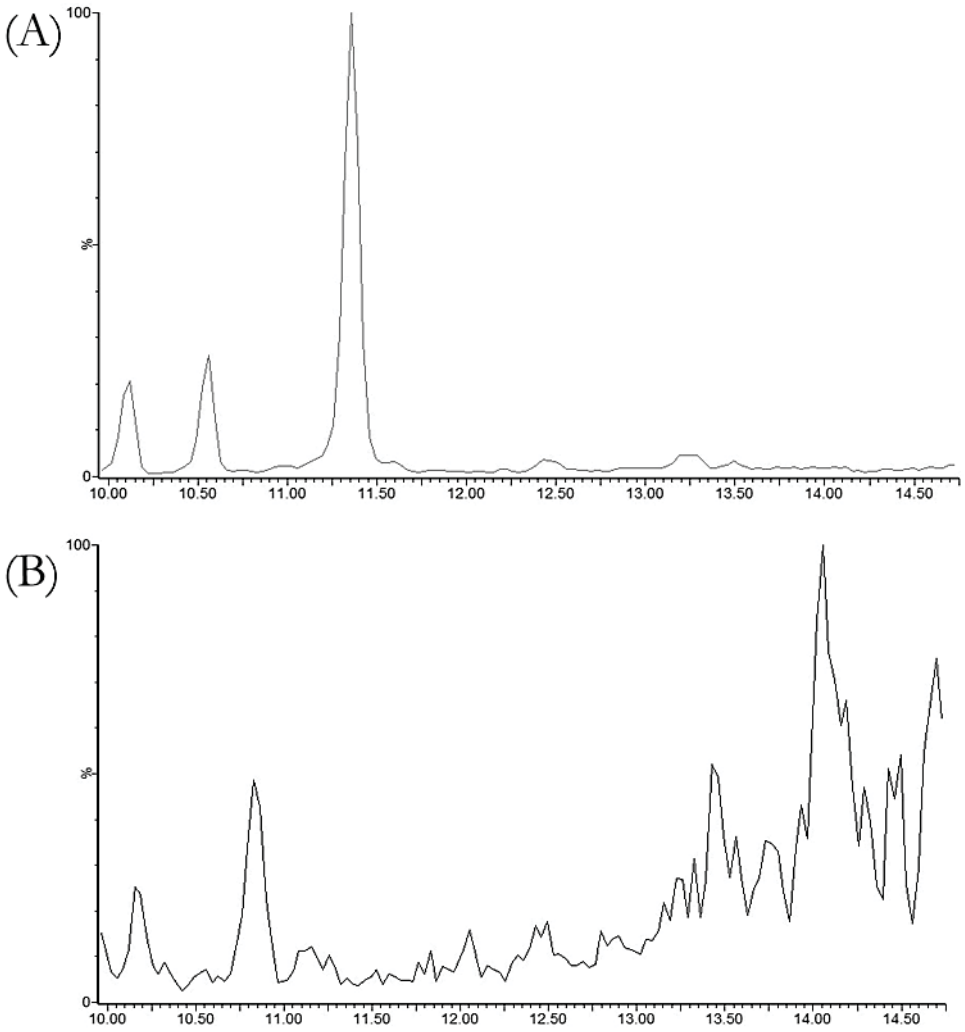


Figure 30: SICs for Dimers with Missing Crosslinking Alanine. (A) -1 Ala (4 amino acids in the subunit), (B) -2 Ala (3 amino acids). It should be noted that the -2 Ala (3 amino acids) elutes around 10.7 minutes.

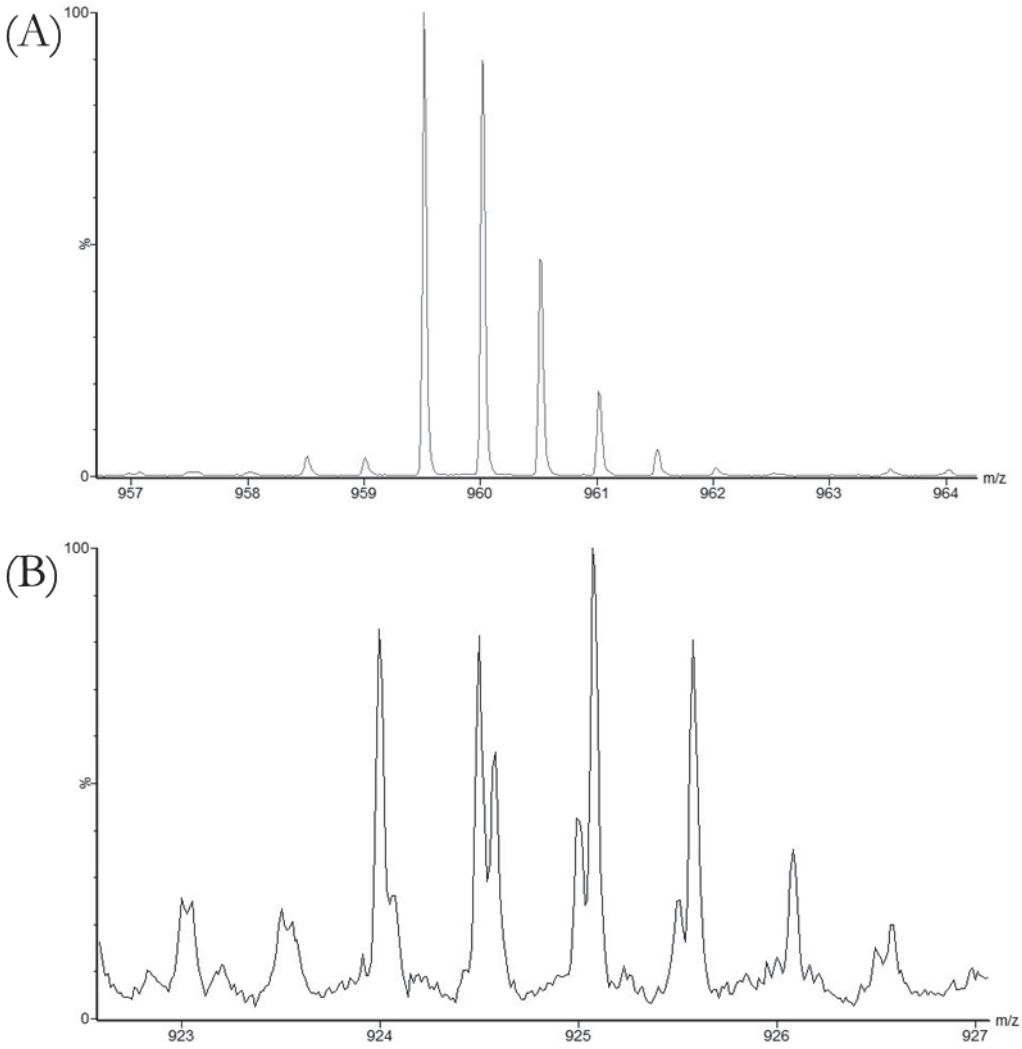
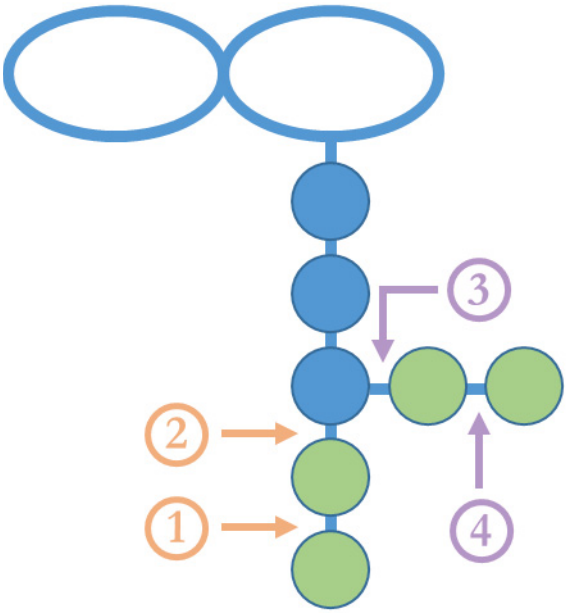


Figure 31: Spectra for Dimers with Missing Crosslinking Alanine. (A) -1 Ala (4 amino acids in the subunit), (B) -2 Ala (3 amino acids).



	1	2	3	4
-2 Ala	-	-		
-1 Ala	-			
	-	-	+	
0 Ala	-	-	+	+
	-	-		
	-		+	
1 Ala	-		+	+
			+	
2 Ala			+	+

Figure 32: Activity of Enzymes for Various Alanylation Numbers. (1) represents D,D-carboxypeptidase, (2) represents L,D-carboxypeptidase, (3) represents BppA1, and (4) represents BppA2. Note that (1) and (2) decrease the number of alanine, while (3) and (4) increase the number of alanine.

Figures 30 and 31 show the SIC and mass spectra for species with missing crosslinking alanine. Before considering these species, it is important to understand how PG fragments in different alanylation states are found. Figure 32 shows how alanylation number is dependent upon the activity of four enzymes at their respective splice sites. D,D-carboxypeptidase and L,D-carboxypeptidase act in sequence to remove first and second terminal D-Ala from the peptide stem, while BppA1 and BppA2 act in sequence to add the first and second L-Ala of the crosslinking bridge (Bouhss et al., 2001)(Bouhss et al., 2002). An alanylation state of -2 Ala must represent a tripeptide with two D-Ala missing from the stem and two L-Ala missing from the crosslinking bridge. The species with an alanylation

state of -1 Ala may be either a tripeptide where the BppA2 enzyme fails to add the second crosslinking L-Ala or a tetrapeptide where BppA1 and BppA2 fail to add the crosslinking L-Ala-L-Ala bridge.

The species with -2 Ala primarily eluted at about 10.7 minutes, earlier than tripeptides with a normal crosslinking bridge, which elute at about 12.1 minutes. The species with -1 Ala eluted at about 11.4 minutes, significantly earlier than tripeptides and tetrapeptides possessing normal crosslinking bridges, which elute at about 12.1 and 12.3 minutes, respectively (Figure 30). The mass spectra for -1 and -2 Ala show similar distributions, but the species with -2 Ala is more difficult to observe on the spectra due to its lower abundance (Figure 31).

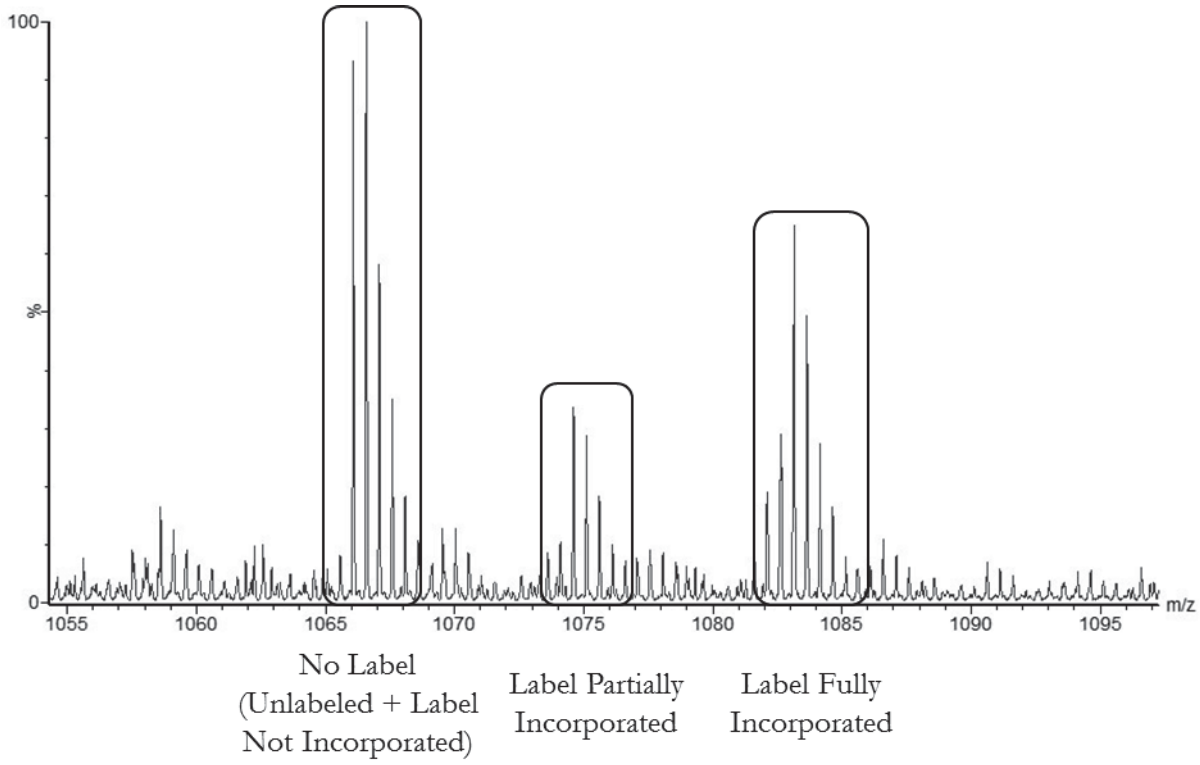


Figure 33: Spectrum Showing Different Degrees of Label Incorporation. "No Label" includes contributions from both labeled and unlabeled samples. "Label Partially Incorporated" and "Label Fully Incorporated" represent species from the labeled sample.

Due to the differences in distributions displayed in the mass spectra, it is clear that a correction is needed for the isotopic distribution. In addition, a broader view of each spectrum demonstrates that there is a third distribution of peaks between the distributions of the unlabeled and labeled species. For simplicity, the spectrum for a dimer pentapeptide is shown in Figure 33. The intermediate distribution corresponds to species that did not fully incorporate the labeled lysine into the PG. Using the correction scheme outlined in the Methods section, the ratio between fully and partially incorporated species was used to

correct for partial incorporation and adjust the intensities of both unlabeled and labeled species.

For larger species, the needed correction was more significant, and Figure 34 depicts the factor of error for pentamer pentapeptides at various levels of incorporation. Figure 34 also illustrates how the factor of error is affected by the percent enrichment of the labeled lysine. From the mass spectra in this experiment, the enrichment was calculated to be 98.3%, and the incorporation was calculated to be 83.3%, corresponding to a necessary correction factor greater than three for pentamer pentapeptides. Based on these values, appropriate corrections were made in order to compare the intensities of PG fragments identified in this experiment.

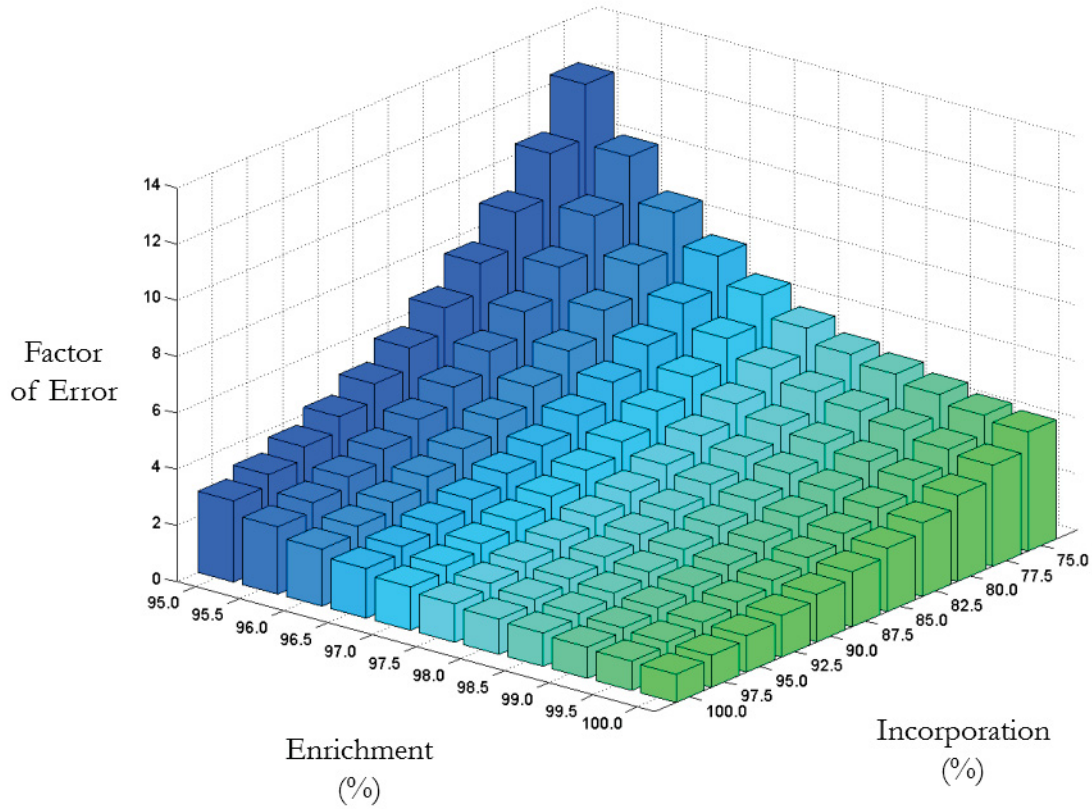


Figure 34: Factor of Error for Various Degrees of Enrichment and Incorporation for a Pentamer Pentapeptide. Enrichment is calculated by comparing expected and observed isotopic distributions. Incorporation is calculated by comparing the abundance of species with fully and partially incorporated labeled lysine. $Factor\ of\ Error = \frac{Intensity_{corrected}}{Intensity_{uncorrected}}$

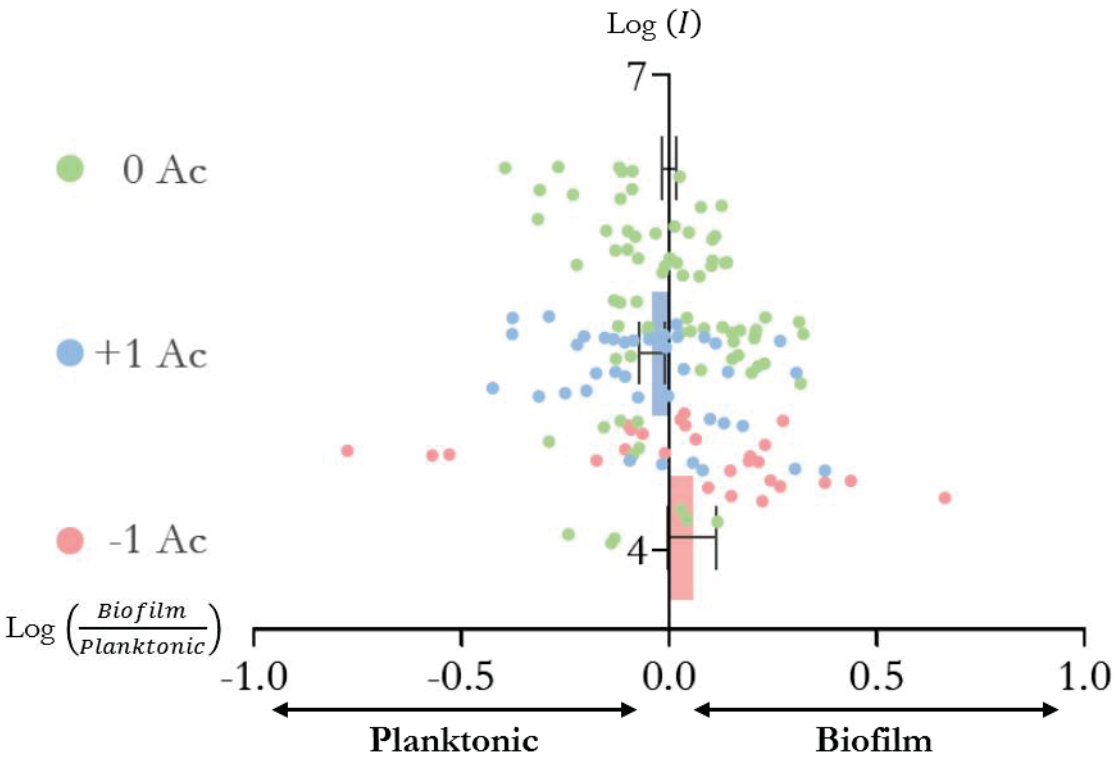


Figure 35: Acetylation Profile. Each data point represents a SILAC pair for a single peptidoglycan species. The x-value indicates a fragment's propensity to be found in either the planktonic or biofilm phase, and the y-value indicates the abundance of the fragment. The y-value is only applicable to the scatter plot data points. The direction and magnitude of the bars indicate the degree to which each acetylation state is favored by either the planktonic or biofilm phase.

Acetylation is found to be more abundant in the planktonic phase, while deacetylation is more abundant in the biofilm phase (Figure 35). Data points for acetylated species are more densely clustered on the planktonic side, indicating that acetylation is less favored when bacteria is in the biofilm phase. Deacetylation is favored in the biofilm phase, as indicated by the data points for species with an acetylation number of -1, which apart from a few outliers are clustered on the biofilm side of the graph.

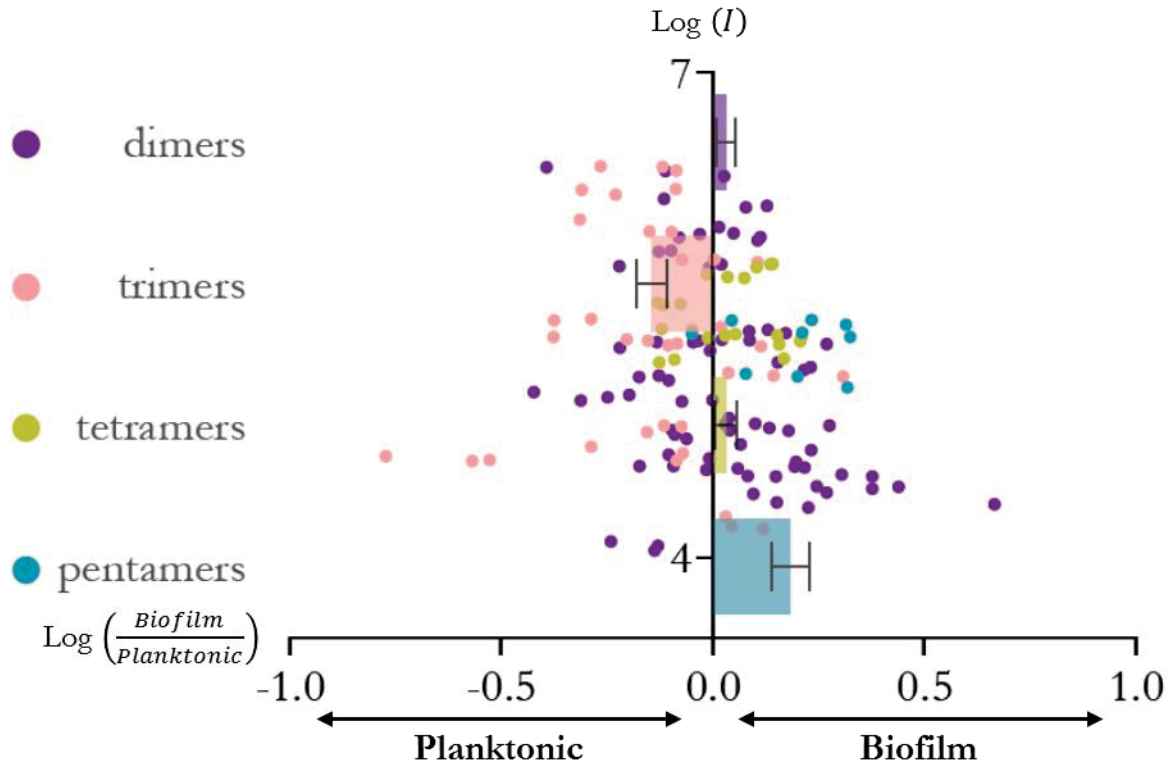


Figure 36: Crosslinking Profile. The scatter and bar data represent the degree to which each multimer is favored by either the planktonic or biofilm phase.

Species with a greater number of crosslinks are found to be favored in the biofilm phase (Figure 36). Though dimers and tetramers are found in slightly greater abundance in the biofilm phase, the trimer and pentamer data emphasizes a trend toward greater crosslinking in the biofilm phase. For the trimers, most of the high intensity species are identified in the planktonic phase, while for the pentamers, all but one data point fall on the biofilm side of the graph.

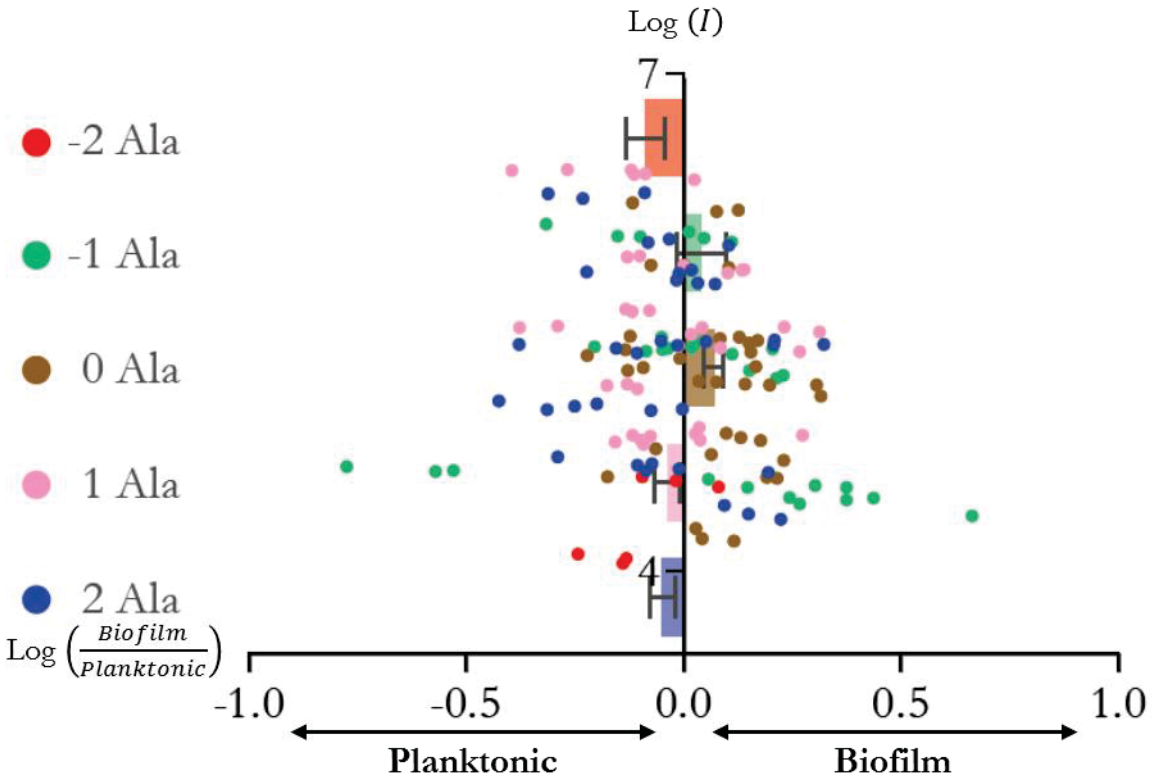


Figure 37: Alanylation Profile. The scatter and bar data represent the degree to which each alanylation state is favored by either the planktonic or biofilm phase.

The trend in alanylation illustrated in Figure 37 is more complicated due to the activity of both carboxypeptidases and BppA enzymes. Carboxypeptidase activity appears to be favored by the biofilm phase for peptidoglycan with normal crosslinkers, as demonstrated by the trend in data from 2 Ala to 1 Ala to 0 Ala. Although there are significantly fewer data points for -2 Ala, and these data points appear at a notably lower intensity, the data clearly shows that -2 Ala species are favored by the planktonic phase. This indicates that the failure of BppA enzymes to add crosslinking alanine is more common in the planktonic phase. The trend toward the planktonic phase for -1 Ala and -2 Ala fragments, species in which one or both of the BppA enzymes has failed to add crosslinking L-Ala, also indicates that in the rare

occasion that BppA enzymes are dysfunctional, carboxypeptidase activity on defective peptidoglycan is enhanced.

Overall, using the SILAC technique, it was found that biofilm exhibits a trend toward species with decreased O-acetylation, increased N-deacetylation, increased crosslinking, increased carboxypeptidase activity, and normal crosslinking bridge formation (Table 5). Aside from the trend in O-acetylation, these results aligned with initial hypotheses regarding biofilm's PG structure.

Table 5: Summary of Modifications Favored in the Planktonic or Biofilm Phase.

	Planktonic	Biofilm
O-Acetylation	✓	
N-Deacetylation		✓
Crosslinking		✓
Carboxypeptidase Activity		✓
Normal Crosslinking Bridge		✓

CHAPTER FOUR

Discussion

This experiment outlines a novel approach for applying the SILAC technique to macromolecules such as peptidoglycan. After corrections for molecule size, isotopic enrichment, and label incorporation, the abundance of modifications to peptidoglycan structure in biofilm and planktonic phases of *E. faecalis* was directly compared in a single mass spectrum. The following trends in peptidoglycan modification in the biofilm phase were identified: 1) decreased O-acetylation, 2) increased N-deacetylation, 3) increased crosslinking, 4) increased editing by carboxypeptidases, and 5) increased crosslinker biosynthesis.

Decreased O-Acetylation

Though O-acetylation, a common post-synthetic modification on the C6 hydroxyl group of NAM, is known to increase bacteria's resistance to antibiotics and host defenses (Moynihan et al., 2014) and was expected to be favored by the biofilm phase, biofilm actually exhibited less O-acetylation than planktonic phase bacteria.

Several explanations may account for biofilm's decrease in O-acetylation. The increased density of bacteria in biofilm may hinder lytic transglycosylases and other host defenses from accessing bacteria within the biofilm, which would minimize the need for O-acetylation in the biofilm phase. Since O-acetylation is metabolically costly because of its direct, extensive use of metabolite to modify cell wall, O-acetylation may not be favored in biofilm due to the decreased necessity for this modification. Because WTA is important for

the functioning of biofilm and attaches to the same oxygen on NAM that the acetyl group is added to, WTA may occupy many of the sites that are normally acetylated in planktonic bacteria. Unfortunately, PG with attached WTA were not observed because they are outside the mass range examined in this experiment (Brown et al., 2013). In addition, decreased O-acetylation makes peptidoglycan more hydrophilic, enhancing its interaction with WTA and poly-N-acetylglucosamine (PNAG), another essential component in the biofilm matrix. These interactions facilitate attractions between the cells of biofilm and between the cells and matrix of biofilm.

Increased N-Deacetylation

Biofilm's increased propensity toward N-deacetylation of NAG is consistent with biofilm's increased resistance to lysozyme. Though the N-deacetylation is generally a less common modification than O-acetylation, an increase in the proportion of N-deacetylated peptidoglycan was observed in the biofilm phase.

N-deacetylation introduces an additional positive charge into the peptidoglycan, which may affect the binding to host structural proteins such as fibrinogen and increase the bacteria's resistance to cationic antimicrobial peptides (Moynihan et al., 2014). As with decreased O-acetylation, increased N-deacetylation increases the hydrophilicity of biofilm peptidoglycan and enhances attractions between cells through interactions with WTA and PNAG.

Increased Crosslinking

As expected, the data shows a clear preference for crosslinking in the biofilm phase. Because of the more sessile nature and lower turnover of peptidoglycan in biofilm, biofilm

peptidoglycan is more likely to be crosslinked than peptidoglycan of the more metabolically active planktonic bacteria. Since biofilm exhibits a less rapid turnover of cells, its rate of peptidoglycan biosynthesis is also likely to be decreased, giving penicillin binding proteins (PBPs) a greater opportunity to carry out the transpeptidation reaction in biofilm.

Increased Editing by Carboxypeptidase

Biofilm exhibits an increased propensity toward editing by carboxypeptidases. Like biofilm's crosslinking, this trend can be attributed to the slower turnover of biofilm peptidoglycan, which provides more opportunity for carboxypeptidases to cleave the peptidoglycan stem. The positive correlation between carboxypeptidase and transpeptidase activity confirms that carboxypeptidase editing occurs after PBP transpeptidation.

Increased Crosslinker Biosynthesis

Failure of BppA1 and BppA2 to add crosslinking alanine is more common in the planktonic phase, which can likely be attributed to the rapid turnover of planktonic peptidoglycan before the actions of BppA1 and BppA2 are complete. Because of this higher turnover in the planktonic phase, the supply of components of cell wall biosynthesis, such as BppA1 and BppA2, may not be sufficient to meet the demand for these components. In contrast, the slower peptidoglycan turnover of the mature, dormant cells of biofilm provides BppA1 and BppA2 with greater opportunity to add the crosslinking alanine.

Though missing crosslinking alanine due to incomplete BppA activity is not a common modification, the trend toward the planktonic phase in species missing both stem and crosslinking alanine suggests a link between carboxypeptidase activity and BppA inactivity. In the interest of creating a regular peptidoglycan lattice, species that do not

contain crosslinking alanine cannot undergo transpeptidation to yield a normal PG structure. These defective species can be marked against further transpeptidation by selective carboxypeptidase activity. Selective carboxypeptidase activity for species missing their crosslinking alanine would prevent irregularities in the cell wall caused by further crosslinking of these defective peptidoglycan.

Importance of PG Modifications to Biofilm

Research on various PG modifications has shown that PG structure is important for bacterial resistance to antibiotics and host defenses. Many antibiotics target certain elements of PG assembly. For example, beta-lactam antibiotics inhibit transpeptidation by PBPs, which ultimately prevents crosslinking. Another example of a drug that targets PG assembly is vancomycin, which binds to PG precursor lipid II and blocks progression of new cell wall biosynthesis by sterically hindering PBPs and transglycosylases. In order to design drugs that are effective in inhibiting biofilm formation, it is important to understand how peptidoglycan structure affects biofilm. Since O-acetylation is not as prevalent in the biofilm phase as it is in the planktonic phase, inhibiting acetyltransferases may not be efficacious against biofilm. However, biofilm's increased hydrophilicity due to decreased O-acetylation and increased N-deacetylation should be noted in the design of new antibiotics that will be active against biofilm. Though hydrophobicity is generally beneficial for the duration of antibiotic action, hydrophobicity must be balanced with hydrophilicity in order to facilitate entry of the antibiotic into the biofilm matrix. The biofilm matrix also contains two very hydrophilic molecules, WTA and PNAG, reinforcing the necessity of developing a drug with proper hydrophilicity so that it can access its target. Since biofilm exhibits increased

crosslinking, targeting PBPs may also be an important element for drugs with efficacy against biofilm.

Future Prospects

In this study, SILAC methods were used to compare the composition of labeled biofilm and unlabeled planktonic bacteria, but this project could be extended to include other combinations of labels, strains, and growth phases. Permutations could be further expanded by using lysine with different numbers of labeled atoms, which would allow for three or more samples to be compared simultaneously. Lysine incorporation calculations provide a model for correcting data obtained using amino acids with an even lower rate of incorporation, such as alanine. Use of alanine would permit a single study to be used to monitor the abundance of WTA and PG, which both contain D-Ala. Additional flexibility in experimental design would enable the use of SILAC to trace biofilm's response to various combinations of cephalosporins, such as ceftaroline and ceftobiprole, and other beta lactams.

The SILAC correction techniques explored in this experiment can also be applied to larger molecules composed of multiple repeating units, such as those encountered in glycobiology. Since mass spectrometry is the preferred method for glycobiology, SILAC is advantageous for direct quantification of different post-translational modifications on glycoproteins, which are important considerations in the clinical setting. Thus, SILAC methods explored in this study introduce novel approaches to questions that have not previously been studied with mass spectrometry.

APPENDIX

```
% Enterococcus faecalis library, matching, and corrections

clear all
clc

% Initializing the matrix
Modification_variables = 19;
Number_of_variations = 3000000;
Library(Number_of_variations,Modification_variables) = 0;

% Variable parameters
crosslink_minimum = 1;
crosslink_maximum = 4;
sugar_missing_minimum = 0;
alanine_minimum = -2;
lactate_minimum = 0;
O_acetylation_minimum = 0;
N_deacetylation_minimum = 0;
label_minimum = 0;
query_charge_limit = 8;
starting_charge_column = 9;

% Mass parameters
mass_proton = 1.0072764668;
start_mass = 968.4782475338;
cross_link_mass = 1020.4975201831;
alanine_mass = 71.0371137878;
lactate_addition_mass = 0.9840155848;
O_Ac_mass = 42.0105646863;
N_DAc_mass = 41.0032882341;
sugar_missing_mass = 203.079372533;
label_medium = 8.0502139672;
label_heavy = 17.0706895263;

% Library Generation
Library_count = 0;

% Looping crosslinks
for c = crosslink_minimum:crosslink_maximum

    % Looping label
    % for light, label=0
    % for medium, label=1
    % for heavy, label=2
    label_maximum = 0;

    for label = label_minimum:label_maximum;

        if label == 0
            difference = 0;
            number_enrichment = 0;
        elseif label == 1
            difference = label_medium;
            number_enrichment = 8;
```

```
elseif label == 2
   difference = label_heavy;
   number_enrichment = 17;
end

% Looping missing sugars
sugar_missing_maximum = 1 + c;
for sugar_missing = sugar_missing_minimum:sugar_missing_maximum

   % Looping alanine
   alanine_maximum = 2;
   for ala = alanine_minimum:alanine_maximum;

      % Looping lactate
      lactate_maximum = floor(ala/2);
      if lactate_maximum < 0
         lactate_maximum = 0;
      end
      for lactate = lactate_minimum:lactate_maximum

         % Looping O-Acetylation
         O_acetylation_maximum = c + 1;
         for O_Ac = O_acetylation_minimum:O_acetylation_maximum

            % Looping N-Deacetylation
            N_deacetylation_maximum = c + 1 + sugar_missing;
            for N_DAc = N_deacetylation_minimum:N_deacetylation_maximum

               Library_count = Library_count + 1;
               Library(Library_count,1) = c;
               Library(Library_count,2) = ala;
               Library(Library_count,3) = O_Ac;
               Library(Library_count,4) = sugar_missing*(-1);
               Library(Library_count,5) = label;
               Library(Library_count,6) = number_enrichment * (c + 1);
               exact_mass = start_mass +(c*cross_link_mass)+(ala*alanine_mass)+(O_Ac*O_Ac_mass) +
                 (N_DAc*(-1)*N_DAc_mass)+ ((sugar_missing*(-1))*sugar_missing_mass) + ((c +
                 1)*(difference)) + (lactate*lactate_addition_mass);
               Library(Library_count,7) = exact_mass;
               initial_charge = N_DAc + 1;
               Library(Library_count,8) = initial_charge;
               Library(Library_count,18) = N_DAc*(-1);
               Library(Library_count,19) = lactate;

               % Filling out the m/z columns
               for charge_count = 0:query_charge_limit
                  Library(Library_count, (starting_charge_column + charge_count)) = ((exact_mass +
                   charge_count * mass_proton) / (initial_charge + charge_count));
               end

            end

         end

      end

   end
```

```
        end

    end

end

% Reading the combinatorically generated library
target_index_size = Library_count;
target = Library(1:target_index_size,:);

% Correcting Library for glufib

%defining glufib values
glufib_theoretical = 785.8457;
glufib_observed = 785.8700;

%array of zeros
target_glufib_array = zeros(target_index_size,18);

%adding values to array
target_glufib_array(:,1:8) = target(:,1:8);
target_glufib_array(:,9:17) = (target(:,9:17)) * (glufib_observed/glufib_theoretical);
target_glufib_array(:,18:19) = target(:,18:19);

for target_glufib_rows = 1:target_index_size

    for target_glufib_columns = 9:17

        if target_glufib_array(target_glufib_rows, target_glufib_columns) >= 2000
            target_glufib_array(target_glufib_rows, target_glufib_columns) = 0;
        end

    end

end


%Generating distribution array (peak heights of light species)

% Initializing the chemical formula variables (creating 2 separate arrays of zeros as
% placeholders)
chemical_formula = zeros (target_index_size,5);
peak_heights = zeros (target_index_size,13);

% Loop for defining chemical formula and determining peak distribution for each chemical formula
for entry_number = 1:target_index_size

    %Caculating number of C,H,N,O,S for each library entry
    chemical_formula(entry_number,1) = 39 + target(entry_number,1)*42 + target(entry_number,2)*3 +
     target(entry_number,3)*2 + target(entry_number,4)*(8) + target(entry_number,18)*2;
    chemical_formula(entry_number,2) = 70 + target(entry_number,1)*72 + target(entry_number,2)*5 +
     target(entry_number,3)*2 + target(entry_number,4)*(13) + target(entry_number,18)*1;
    chemical_formula(entry_number,3) = 9 + target(entry_number,1)*10 + target(entry_number,2)*1 +
     target(entry_number,3)*0 + target(entry_number,4)*(1) + target(entry_number,18)*0;
    chemical_formula(entry_number,4) = 19 + target(entry_number,1)*19 + target(entry_number,2)*1 +
     target(entry_number,3)*1 + target(entry_number,4)*(5) + target(entry_number,18)*1;
```

```
    distribution = isotopicdist([chemical_formula(entry_number,1) chemical_formula(entry_number,2)
     chemical_formula(entry_number,3) chemical_formula(entry_number,4)
     chemical_formula(entry_number,5)]);

    %Putting data into array (rows are library entries, columns are relative peak heights of up to ten peaks)
    distribution_length = length (distribution);
    if distribution_length >= 10
       distribution_length = 10;
    end
    for distribution_number = 1:distribution_length
       peak_heights (entry_number,distribution_number) = distribution (distribution_number,2);
    end

end

%initializing binomial distribution array (zeros as placeholders)
binomial_dist_array = zeros (4,target_index_size);
percent_enrichment_medium = 0.995;
percent_enrichment_heavy = 0.983;

for binomial_dist_count = 1:target_index_size

    if target(binomial_dist_count,5) > 0
       %specifying medium or heavy
       if target(binomial_dist_count,5) == 1
          temporary_percent_enrichment = percent_enrichment_medium;
       else
          temporary_percent_enrichment = percent_enrichment_heavy;
       end

       %generating binomial distribution array
       for binomial_dist_peaks = 0:3
          %calculating binomial coefficient
          binomial_coefficient =
            factorial(target(binomial_dist_count,6))/(factorial(binomial_dist_peaks)*(factorial(target(binomial_dist
            _count,6)-binomial_dist_peaks)));
          binomial_expression = (temporary_percent_enrichment^(target(binomial_dist_count,6)-
            binomial_dist_peaks))*((1-temporary_percent_enrichment)^(binomial_dist_peaks));
          binomial_dist_array(binomial_dist_peaks + 1, binomial_dist_count) =
            binomial_coefficient*binomial_expression;
       end

    else
          binomial_dist_array(1,binomial_dist_count) = 1;
    end

end

%initializing combined array
combined_array = zeros (target_index_size,16);

%combining distribution and binomial distribution
for library_entry = 1:target_index_size

    %initializing product array
    product_array = zeros(16,4);
```

```
    for distribution_entry = 1:10

        for binomial_entry = 1:4

            %defining each element in product array
            product_array(distribution_entry + 3,binomial_entry) = peak_heights(library_entry,distribution_entry)
             * binomial_dist_array(binomial_entry,library_entry);

        end

    end

    %putting values in combined array (sum of products)
    for rows = 1:13

        for term = 0:3
            combined_array(library_entry,rows) = combined_array(library_entry,rows) + product_array(rows +
             term, 1 + term);
        end

    end

end

%making array of correction factors
correction_factor_array = zeros(target_index_size,1);

for correction_factor_count = 1:target_index_size
    correction_factor_array(correction_factor_count,1) = max(combined_array(correction_factor_count,:));
end


% Reading the manually picked m/z targets of interest
[mz_query] = xlsread('PG match list',4,'G3:Q28');
mz_query_index_size = length(mz_query);

% Defining search parameters and initializing the targets
tolerance = 50; % in ppm
match = zeros(5000,30);
match_index = 1;

% Search algorithm
for mz_query_index = 1:mz_query_index_size

    % Setting the m/z query targets

    %observed
    current_observed_mz_01 = mz_query(mz_query_index,3);
    current_observed_mz_02 = mz_query(mz_query_index,6);
    current_observed_mz_03 = mz_query(mz_query_index,9);

    %observed corrected for glufib
    current_query_mz_01 = mz_query(mz_query_index,4);
    current_query_mz_02 = mz_query(mz_query_index,7);
    current_query_mz_03 = mz_query(mz_query_index,10);

    current_query_charge = mz_query(mz_query_index,1);
```

```
current_query_intensity_01 = mz_query(mz_query_index,5);
current_query_intensity_02 = mz_query(mz_query_index,8);
current_query_intensity_03 = mz_query(mz_query_index,11);

% Moving the target from one to another while making the comparison
for target_index = 1:target_index_size

    % Taking the difference
    target_charge_column = current_query_charge + 8 - (target(target_index,8) - 1);
    mz_difference_01 = current_query_mz_01 - target(target_index,target_charge_column);
    mass_difference_01 = mz_difference_01*current_query_charge;
    ppm_difference_01 = (mass_difference_01/target(target_index,7))*1000000;

    target_charge_column = current_query_charge + 8 - (target(target_index,8) - 1);
    mz_difference_02 = current_query_mz_02 - target(target_index,target_charge_column);
    mass_difference_02 = mz_difference_02*current_query_charge;
    ppm_difference_02 = (mass_difference_02/target(target_index,7))*1000000;

    target_charge_column = current_query_charge + 8 - (target(target_index,8) - 1);
    mz_difference_03 = current_query_mz_03 - target(target_index,target_charge_column);
    mass_difference_03 = mz_difference_03*current_query_charge;
    ppm_difference_03 = (mass_difference_03/target(target_index,7))*1000000;

    %Determining minimum ppm difference
    ppm_difference_array = zeros(3,1);
    ppm_difference_array(1,1) = abs(ppm_difference_01);
    ppm_difference_array(2,1) = abs(ppm_difference_02);
    ppm_difference_array(3,1) = abs(ppm_difference_03);
    ppm_difference_min = min(ppm_difference_array(1:3,1));

    % Comparing the difference and adding to the matches if found
    if ppm_difference_min < tolerance && current_query_charge >= target(target_index,8)
        match(match_index,1) = target_index;
        match(match_index,2) = mz_query(mz_query_index,2);
        match(match_index,3:10) = target(target_index,1:8);
        match(match_index,11) = current_query_charge;
        match(match_index,12) = current_observed_mz_01;
        match(match_index,13) = current_observed_mz_02;
        match(match_index,14) = current_observed_mz_03;
        match(match_index,15) = current_query_mz_01;
        match(match_index,16) = current_query_mz_02;
        match(match_index,17) = current_query_mz_03;
        match(match_index,18) = target(target_index,target_charge_column);
        match(match_index,19) = ppm_difference_01;
        match(match_index,20) = ppm_difference_02;
        match(match_index,21) = ppm_difference_03;
        match(match_index,22) = current_query_intensity_01;
        match(match_index,23) = current_query_intensity_02;
        match(match_index,24) = current_query_intensity_03;
        match(match_index,25) = correction_factor_array(target_index,1);
        %observed monoisotopic peak intensities
        match(match_index,26) = current_query_intensity_01 / correction_factor_array(target_index,1);
        match(match_index,27) = current_query_intensity_02 / correction_factor_array(target_index,1);
        match(match_index,28) = current_query_intensity_03 / correction_factor_array(target_index,1);
        match(match_index,29) = target(target_index,18);
        match(match_index,30) = target(target_index,19);
```

```
            match_index = match_index + 1;
        end

    end

end


% lysine incorporation
HH_HL_ratio = 2.50;
prob_H = (2 * HH_HL_ratio) / (1 + (2 * HH_HL_ratio));

% initializing corrected match array
match_corrected = zeros(match_index, 30);
match_corrected(:, 1:25) = match(1:match_index, 1:25);
match_corrected(:,29) = match(1:match_index, 29);
match_corrected(:,30) = match(1:match_index, 30);

% adjusting all heavy and light intensities
multimer_min = crosslink_minimum + 1;
multimer_max = crosslink_maximum + 1;

for counting_multimers = multimer_min:multimer_max
    HH_per_x = (prob_H)^counting_multimers;
    LL_per_x = (1 - prob_H)^counting_multimers;

    % correcting all heavy intensities
    for match_entry = 1:match_index

        if match(match_entry, 7) == 2 && match(match_entry,3) == (counting_multimers - 1)
            match_corrected(match_entry, 26) = match(match_entry,26) / HH_per_x;
            match_corrected(match_entry, 27) = match(match_entry,27) / HH_per_x;
            match_corrected(match_entry, 28) = match(match_entry,28) / HH_per_x;
        end

    end

    %correcting all light species by searching corresponding heavy and subtracting heavy contribution
    for specie_to_be_matched = 1:match_index

        for searching_light = 1: match_index

            if match(searching_light, 7) == 0

                for finding_corresponding_heavy = 1:match_index

                    if match(finding_corresponding_heavy, 7) == 2

                        for counting_c = 1:match_index

                            if match(counting_c, 3) == match(specie_to_be_matched ,3)

                                for counting_ala = 1:match_index

                                    if match(counting_ala, 4) == match(specie_to_be_matched, 4)

                                        for counting_z = 1:match_index
```

```
                    if match(counting_z, 11) == match(specie_to_be_matched, 11)

                        if counting_c == counting_ala && counting_ala == counting_z

                            if match(specie_to_be_matched,7) == 0 && match(specie_to_be_matched,3)
                             == (counting_multimers - 1)
                               match_corrected(specie_to_be_matched, 26) =
                                 match(specie_to_be_matched, 26) – (match(counting_z, 26) *
                                 (LL_per_x/HH_per_x));
                               match_corrected(specie_to_be_matched, 27) =
                                 match(specie_to_be_matched, 27) – (match(counting_z, 27) *
                                 (LL_per_x/HH_per_x));
                               match_corrected(specie_to_be_matched, 28) =
                                 match(specie_to_be_matched, 28) – (match(counting_z, 28) *
                                 (LL_per_x/HH_per_x));
                            end

                        end

                    end

                end

            end

        end

    end

end
```

BIBLIOGRAPHY

Archer, N. K.; Mazaitis, M. J.; Costerton, J. W.; Leid, J. G.; Powers, M. E.; Shirtliff, M. E. *Staphylococcus aureus* Biofilms: Properties, Regulation and Roles in Human Disease. *Virulence* **2011,** *2 (5)*, 445-459.

Benachour, A.; Labjouzi, R.; Jeune, A. L.; Hébert, L.; Thorpe, S.; Courtin, P.; Chapot-Chartier, M.; Prajsnar, T. K.; Foster, S. J.; Mesnage, S. The Lysozyme-Induced Peptidoglycan N-Acetylglucosamine Deacetylase PgdA (EF1843) is Required for *Enterococcus faecalis* Virulence. *Journal of Bacteriology* **2012**, *194 (22)*, 6066-6073.

Bjarnsholt, T. The Role of Bacterial Biofilms in Chronic Infections. *APMIS* **2013**. *121*, 1-58.

Bouhss, A.; Josseaume, N.; Allanic, D.; Crouvoisier, M.; Gutmann, L.; Mainardi, J.-L.; Mengin-Lecreulx, D.; van Heijenoort, J.; Arthur, M. Identification of the UDP-MurNAc-Pentapeptide:L-Alanine Ligase for Synthesis of Branched Peptidoglycan Precursors in *Enterococcus faecalis*. *Journal of Bacteriology* **2001**, *183 (17)*, 5122–5127.

Bouhss, A.; Josseaume, N.; Severin, A.; Tabei, K.; Hugonnet, J.-E.; Shlaes, D.; Mengin-Lecreulx, D.; Heijenoort, J. van; Arthur, M. Synthesis of the L-Alanyl-L-alanine Cross-bridge of *Enterococcus faecalis* Peptidoglycan. *Journal of Biological Chemistry* **2002**, *277 (48),* 45935–45941.

Brown, S.; Santa Maria, J. P.; Walker, S. Wall Teichoic Acids of Gram-Positive Bacteria. *Annual Review of Microbiology* **2013**, 67.

Brock, T. D.; Madigan, M. T.; Martinko, J. M. *Biology of Microorganisms*, 7th ed.; Pearson Prentice Hall: Upper Saddle River, 1994.

Centers for Disease Control and Prevention. Antibiotic Resistance Threats in the United States, 2013. http://www.cdc.gov/drugresistance/threat-report-2013/ (accessed October 15, 2014).

Costeron, J.; Stewart, P. Bacterial Biofilms: A Common Cause of Persistent Infections. *Science* **1999**, *284*, 1318-1322.

Flemming, H.-C.; Wingender, J. The Biofilm Matrix. *Nature Reviews Microbiology* **2010**, *8 (9)*, 623–633.

Foster, E. E. The Effects of Vancomycin on the Structure of Cell Wall Peptidoglycan of *vanB*-type Vancomycin-Resistant *Enterococcus faecalis* by Liquid Chromatography-Mass Spectrometry. Honors Thesis, Baylor University, Waco, TX, 2015.

Geipel, U. Pathogenic Organisms in Hip Joint Infections. *International Journal of Medical Sciences* **2009**, *6 (5)*, 234-240.

Guiton, P. S.; Hung, C. S.; Kline, K. A.; Roth, R.; Kau, A. L.; Hayes, E.; Heuser, J.; Dodson, K. W.; Caparon, M. G.; Hultgren, S. J. Contribution of Autolysin and Sortase A During *Enterococcus faecalis* DNA-Dependent Biofilm Development. *Infection and Immunity* **2009**, *77 (9)*, 3626–3638.

Ipsen, A. Efficient Calculation of Exact Fine Structure Isotope Patterns via the Multidimensional Fourier Transform. *Analytical Chemistry* **2014**, *86*, 5316-5322.

Lupoli, T. J.; Tsukamoto, H.; Doud, E. H.; Wang, T.-S. A.; Walker, S.; Kahne, D. Transpeptidase-Mediated Incorporation of D-Amino Acids into Bacterial Peptidoglycan. *Journal of American Chemical Society* **2011**, *133 (28)*, 10748-10751.

McKessar, S. J.; Berry, A. M.; Bell, J. M.; Turnidge, J. D.; Paton, J. C. Genetic Characterization of *vanG*, a Novel Vancomycin Resistance Locus of *Enterococcus faecalis*. *Antimicrobial Agents and Chemotherapy* **2000**, *44 (11)*, 3224–3228.

Moynihan, P. J.; Sychantha, D.; Clarke, A. J. Chemical Biology of Peptidoglycan Acetylation and Deacetylation. *Bioorganic Chemistry* **2014**, *54*, 44–50.

Ong, S.-E.; Mann, M. A Practical Recipe for Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC). *Nature Protocols* **2007**, *1 (6)*, 2650–2660.

Stoodley, P.; Fux, C.; Dirchx, P. Montana State University Center for Biofilm Engineering. Sites of Primary and Secondary Biofilm Infection. 2003. http://www.biofilm.montana.edu/resources/images/healthmedical-biofilms/sites-primary-and-secondary-infection.html (accessed November 18, 2015).

Tu Quoc, P. H.; Genevaux, P.; Pajunen, M.; Savilahti, H.; Georgopoulos, C.; Schrenzel, J.; Kelley, W. L. Isolation and Characterization of Biofilm Formation-Defective Mutants of *Staphylococcus aureus*. *Infection and Immunity* **2007**, *75 (3)*, 1079–1088.

Vollmer, W. Structural Variation in the Glycan Strands of Bacterial Peptidoglycan. *FEMS Microbiology Reviews* **2008**, *32*, 287-306.

Yeagley, A. A.; Su, Z.; McCullough, K. D.; Worthington, R. J.; Melander, C. N-Substituted 2-aminoimidazole Inhibitors of MRSA Biofilm Formation Accessed through Direct 1,3-bis(tertbutoxycarbonyl) Guanidine Cyclization. *Organic and Biomolecular Chemistry* **2012**, *11 (1)*, 130–137.

Zimmerli, W.; Trampuz, A.; Ochsner, P. E. Prosthetic-Joint Infections. *New England Journal of Medicine* **2004**, *351 (16)*, 1645–1654.