



5-2019

Applications of Chemical Methodology in Environmental Science, Systems Biology, and Interdisciplinary Chemical Education

Caleb Michael Gibson

University of Tennessee, cgibso21@vols.utk.edu

Follow this and additional works at: https://trace.tennessee.edu/utk_graddiss

Recommended Citation

Gibson, Caleb Michael, "Applications of Chemical Methodology in Environmental Science, Systems Biology, and Interdisciplinary Chemical Education. " PhD diss., University of Tennessee, 2019.
https://trace.tennessee.edu/utk_graddiss/5400

This Dissertation is brought to you for free and open access by the Graduate School at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a dissertation written by Caleb Michael Gibson entitled "Applications of Chemical Methodology in Environmental Science, Systems Biology, and Interdisciplinary Chemical Education." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Chemistry.

Shawn Campagna, Major Professor

We have read this dissertation and recommend its acceptance:

Elizabeth Fozo, Michael Sepaniak, Ampofo Darko

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

**APPLICATIONS OF CHEMICAL METHODOLOGY IN
ENVIRONMENTAL SCIENCE, SYSTEMS BIOLOGY,
AND INTERDISCIPLINARY CHEMICAL EDUCATION**

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Caleb Michael Gibson
May 2019

Copyright © 2019 by Caleb Michael Gibson
All rights reserved.

DEDICATION

To Ellie, my beautiful daughter. Though you are very small, your curiosity is already evident. I hope that you grow more inquisitive as you grow bigger, leading you in the pursuit of knowledge and truth.

ACKNOWLEDGEMENTS

There are many people to whom I wish to express gratitude. Initially, I must extend many thanks to my advisor, Dr. Shawn Campagna. Without his investment in me, his continued direction, and his reassurance during the distresses of research, I would never have realized my goals. Likewise, I wish to thank my committee of Drs. Ampofo Darko, Elizabeth Fozo, and Michael Sepaniak who have helped me grow as a scientist through their collective input and constructive scrutiny of my ideas.

My greatest opportunities in science were made possible through collaborations with excellent researchers. Most prominent among these are Drs. Brynn Voy and Elizabeth Fozo who have introduced me to intriguing systems and inspired me to apply myself to pursuit of answers. Moreover, they patiently aided me in developing from a strict chemist into an interdisciplinary scientist. Other collaborators including Dr. Maria Campa have allowed me to expand my horizons and gain experience in environmental science. In addition, there are a few individuals to whom I owe gratitude for their input in the implementation of my curriculum. These are Dr. Elizabeth Fozo, April Armes, Dr. Lauren Anderson, Matt Gleason, and Megan Qualls.

My history as an academic did not begin with my graduate studies. I would be remiss if I did not reflect further into the past and acknowledge those that instilled me with curiosity and encouraged me to pursue knowledge. Foremost among these are Drs. Marc Strand, Christine Dalton, and Paul Martino. Through their mentorship, these individuals helped me find my niche in science. Others, such as Drs. John McClellan and Brian Austin, reinforced the critical thinking skills that are foundational to the pursuit of any knowledge, regardless of the field.

During my studies, I have worked alongside several group members who have aided me by providing scientific input and camaraderie. These are: Dr. Hector Castro, Dr. Stephen Dearth, Dr. Carson Prevatte, Dr. Abigail Farmer, Dr. Allen Bourdon, Brandon Kennedy, Eric Tague, Alexander Fisch, Joshua Powers, Jordan Rogerson, Maggie Lookadoo, Ashley Lato, Katarina Jones, and Katharina Hölland. In addition, I have had the

opportunity to work with some excellent undergraduate researchers including Sheev Zaver, Hannah Holst, and Brittni Woodall, who have each contributed to my projects.

This document will not fully communicate the arduous ventures of my graduate studies. This journey has, at times, been both exciting and demoralizing and I must express my appreciation towards those close friends and family who have provided constant moral support along the way. My close friend and fellow chemist Joshua Hinkle has afforded fresh perspectives and astute insights when I need them. My parents, Bruce and Diane, have always lovingly supported me with sincerity and without undue pressure. My siblings and in-laws are close friends whose confidence in me has been encouraging. My daughter Ellie has introduced a joyful motivation these past few months. Most of all, my wife Maggie has stubbornly believed in me more than I believe in myself. Her love and companionship have been prime and enduring motivators. She has been my quintessential “better half”, and I am ever grateful.

It may seem inconsequential, but I must offer earnest gratitude to the serene beauty of the Blue Ridge Mountains and the vibrant fish populations of local rivers and streams. They have provided a welcome respite for an overactive mind, allowing me to rest and refocus.

My final and most humble thanks are expressed to the Bringer of the Good News, who has extended gracious gifts of which I am far less worthy than I ever feared but far more grateful than I ever imagined.

I am wiser than this human being. For probably neither of us knows anything noble and good, but he supposes he knows something when he does not know, while I, just as I do not know, do not even suppose that I do. I am likely to be a little bit wiser than he in this very thing: that whatever I do not know, I do not even suppose I know.

-Socrates, from Plato's Apology

TABLE OF CONTENTS

Chapter One : Introduction	1
The bridges between fields: chemical techniques applied to biological systems	2
A chemical toolbox.....	2
Ecotoxicology of small molecules in the environment.....	3
Chemical biology: the interdisciplinary science.....	4
Central metabolism and ancillary networks	5
Small-molecule mitochondrial uncouplers and obesity management	7
Membrane disruptors in the intermediate molecular space	9
Addressing the need for interdisciplinarity in science education	10
Chapter Two: Impact of hydraulic fracturing on degradation potential of the biocide glutaraldehyde	11
Abstract	12
Preface	12
Introduction.....	14
Materials and methods	15
Results and discussion	22
Chapter Three: Exploring endogenous expression of <i>N</i>-Acyl amino acids	40
Abstract	41
Preface	42
Chapter 3A: A single UPLC-HRMS method for the profiling of bile acids and <i>N</i> -acyl amino acids, two classes of obesity-mediating metabolites.....	44
Chapter 3B: Expression of <i>N</i> -acyl amino acids is influenced by fasting-induced lipolysis and dietary supplementation with exogenous fatty acids	60
Chapter Four: Investigation of the toxic ZorO peptide	75
Abstract	76
Preface	76
Introduction.....	77
Methods.....	85
Results and discussion	94
Conclusions and future work	102
Chapter Five: Design and inception of interdisciplinary coursework in undergraduate chemical biology education	104
Introduction.....	105
Course material	111
Discussion	111
Outlook	113
Chapter Six: Conclusions	114
Works cited	116
Appendices	136
Appendix A: Supporting information for chapter 2	137
Appendix B: Supporting information for chapter 3	141
Appendix C: Supporting information for chapter 4	225
Appendix D: Course content for chapter 5.....	235
Vita	319

LIST OF TABLES

Table 3-1: Detection parameters for metabolite standards.....	52
Table 3-2: Differentiation of isobaric <i>N</i> -acyl amino acids using Parallel Reaction Monitoring	56
Table 3-3: Calculated concentrations of bile acids extracted from mouse duodenum...	59
Table 4-1: Attempted alpha halogenation conditions.....	87
Table 5- 1: Composition of course, highlighting the featured techniques and molecular classes.....	107
Table 5- 2: Semester schedule. Weeks marked with an asterisk are part of a multi-week sequence. Weeks marked with a double dagger are dry labs. Applicable references are cited for each week.....	108

LIST OF FIGURES

Figure 1-1: Cellular respiration	6
Figure 1-2: Effect of uncouplers on central energy metabolism. Interruption of the proton gradient (shown in black) interrupts the proton motive force and inactivates ATP synthase.....	8
Figure 2-1: Map showing the location of the PA watersheds used as microcosms source water	16
Figure 2-2: Calibration curves for glutaraldehyde (left) and its metabolite glutaric acid (right).....	18
Figure 2-3: Glutaraldehyde species detected.....	24
Figure 2-4: Biotic and abiotic degradation of glutaraldehyde and glutaric acid production over time. (a) Biotic and abiotic degradation of glutaraldehyde in HF+ and HF- microcosms. The blue dot represents the added amount of GA, 100 mg/L. (b) Biotic and abiotic production of glutaric acid in HF+ and HF- microcosms; the zoom in graph shows abiotic concentration over time. Error bars represent one standard error (n = 9).	25
Figure 2-5: TOC over time.....	27
Figure 2-6: Impacts of glutaraldehyde in abundance of 16S rRNA gene over time. The first three clusters are the HF-impacted streams, and the last three clusters represent the non-HF-impacted streams. Data point “56 days-No GA” represents bottle effect on the microcosms as no GA was added. Error bars represent one standard error (n = 3).	28
Figure 2-7: Alpha diversity measurements over time. Different richness and evenness alpha diversity estimators comparing HF+ and HF- microcosms over time; the estimators used were (a) Shannon Diversity, (b) Observed Diversity, (c) Chao1, and (d) Simpson Diversity. Red and green box plots represent HF+ glutaraldehyde (days 7 to 56) and no glutaraldehyde added (days 0 and 56 only). Blue and purple box plots represent HF- glutaraldehyde and no glutaraldehyde added. The box plots show the distribution of the data points: upper whisker to the beginning of the box is the first quartile, beginning of box to median represents the second quartile of the data, median to end of box is third quartile, and end of box to lower whisker is the fourth quartile.....	31
Figure 2-8: Principal coordinate analysis (PCoA) plot of phylogenetic microbial community changes over time in HF+ and HF- impacted microcosms amended or unamended with glutaraldehyde as described by weighted Unifrac beta diversity measurements. PC1 explains 65.4% of the variation while PC2 explains 10%.	32
Figure 2- 9: Microbial Community Shifts Over Time. A) Phylum level shifts over 56 days with and without glutaraldehyde addition, samples were average into HF-impacted and non-HF-impacted groups.	35
Figure 3-1: Synthetic scheme for <i>N</i> -acyl amino acid standards	47
Figure 3-2: Chromatographic resolution of bile acid isomers	54

Figure 3-3: Representative illustration of noncovalent <i>N</i> -acyl amino acid dimers detected by HRMS. Shown: [M-H] ⁻ and [2M-H] ⁻ ions for the eicosapentaenoyl phenylalanine standard.....	55
Figure 3-4: Detection of metabolites in biological samples from broiler chicks and mice. A) PLS-DA of the <i>N</i> -acyl amino acids detected in avian adipocytes. Data is grouped by diet. B) Graphic illustrating the number of metabolites detected in each sample type	58
Figure 3-5: PLS-DA of fasting states based on serum <i>N</i> -AAAs.....	64
Figure 3-6: Metabolites quantitated in serum of fasted chicks. A) <i>N</i> -AAAs; B) NEFA ...	65
Figure 3- 7: Percent <i>N</i> -AAA composition by fasting state. ANOVA significant <i>N</i> -AAAs are starred.....	66
Figure 3-8: Correlation matrix of <i>N</i> -AAAs detected in the serum of fasting chicks	67
Figure 3-9: Expression of <i>PM20D1</i>	69
Figure 3-10: <i>N</i> -AAAs (grouped by fatty acid) detected in chicks fed 4 distinct diets. The inset shows the levels of fatty acids in neutral lipid from adipose tissue of chicks from the same groups	69
Figure 3-11: Correlation map of <i>N</i> -AAAs detected in the tissue of chicks fed 4 distinct diets	71
Figure 4-1: Potential mechanism of ZorO toxicity	78
Figure 4-2: Photo-cross-linking	80
Figure 4-3: Structure of photo-methionine.....	81
Figure 4-4: Overview of solid-phase peptide synthesis (SPPS)	82
Figure 4-5: Synthesis of photo-methionine by Thiele et. al.....	86
Figure 4-6: Synthesis of photo-methionine by Muir et. al.	88
Figure 4-7: Synthesis of photo-methionine adapted from Muir et. al.	88
Figure 4-8: Synthesis of photo-methionine by Gibson et. al.	89
Figure 4-9: ZorO and analogs synthesized with SPPS.....	93
Figure 4-10: 1H NMR characterization of alpha-bromination attempts.....	95
Figure 4-11: MS Characterization of synthetic ZorO	97
Figure 4-12: MS characterization of synthetic R23>K ZorO	98
Figure 4-13: MS characterization of 1X-FLAG ZorO	99
Figure 4-14: Western blot of synthetic 1X-FLAG ZorO and 1X-FLAG ZorO expressed from plasmids.....	100
Figure 4-15: ZorO standard analyzed via UPLC-HRMS as a standard	101
Figure 5-1: Concept map of the course scope: firmly rooted in chemistry to answer biological inquiries.....	106

CHAPTER ONE : INTRODUCTION

The bridges between fields: chemical techniques applied to biological systems

Scientific disciplines began diverging in the eighteenth century, emerging in the nineteenth century as dedicated funding for laboratories arose.¹ Evolving specializations caused the distinctions between disciplines to intensify during the growth of academia. By the 1950s, scientists were noting—and sometimes lamenting—the resulting isolation.¹ Based on frequency of cross-discipline citations, the trend of division took a turn in the 1970's, and interdisciplinarity began to steadily rise.² Bridges were built to re-establish relationships between the disciplines and participate in interdisciplinary research. Seeking to integrate the tools, concepts, and theories of complimentary fields, interdisciplinarity allows synergistic approaches. This is typified by collaboration, holistic understanding, and coordination.¹

Chemistry is perfectly poised to participate in these joint endeavors. Often focused on synthesizing material and measuring molecular-level events, chemical techniques can be readily adapted to probe questions from life science and test biological hypotheses. Modern fields like chemical biology and ecotoxicology draw from multiple disciplines, frequently involving an explicit connection between chemical methodology and biological inquiry.

Herein, interdisciplinary chemistry is used to investigate three biological systems—each with a unique question. The questions range from cellular to ecological in scope and include both prokaryotic and eukaryotic organisms. However, they are unified by a central theme of connecting chemical tools and collaborative perspectives. In these approaches, chemical and biological data are woven together with complementarity to illuminate trends and mechanisms.

A chemical toolbox

The following chapters draw from synthetic organic chemistry as well as analytical chemistry in an integrated approach. Organic synthesis, broadly defined as the intentional construction of organic molecules using chemical reactions, is a classical tool of the chemist. The capacity to create new entities of matter and re-create naturally produced molecules are equally invaluable. In the latter instance, organic synthesis provides a

means to amass large quantities of pure biomolecules to facilitate *in situ*, *in vitro*, and *in vivo* studies. This is particularly vital when isolation of these molecules from natural sources requires astronomical opportunity cost. In the following chapters, organic synthesis is employed to facilitate simplified *in vitro* assays and verified instrumental detection of two classes of molecules.

On the analytical front, mass spectrometry (MS) is a powerful technique that is widely applied to biological systems due to its high specificity and broad adaptability. In short, it revolves around the manipulation of charged particles to detect individual molecules by their mass. Ultra-performance liquid chromatography coupled with high resolution mass spectrometry (UPLC-HRMS) allows both temporal resolution and mass resolution of biomolecules to facilitate a technique which can confidently discriminate specific analytes within a complex sample. Due to these attributes, LC-MS is advantageous over other spectroscopic techniques for high-throughput analyses of many molecules in a convoluted matrix. The broad adaptability of MS has encouraged the development of various instrumental platforms featuring unique ionization mechanisms and mass analysis methods. For biomolecules with a high composition of heteroatoms, electrospray ionization (ESI) is the preferred method of ionizing liquid eluants from UPLC. Orbitrap mass analyzers rely on the Fourier transform of a frequency signal from oscillating ions to produce highly accurate mass identifications (<5 ppm). With UPLC, ESI and Orbitraps enable a highly modular platform for the analysis of biologically active molecules—including both polypeptides and small molecules. Therefore, UPLC-ESI-HRMS is used in all three forthcoming chapters as an adaptable technique for detection and quantitation.

Ecotoxicology of small molecules in the environment

In the first part of this work (Chapter Two), the effect of biocide exposure on the environment is evaluated using chemical and biological data. It establishes the emerging common thread of developing chemical methodologies and applying them to address biological inquiries. This chapter relies more heavily on mass spectrometry methodology, highlighting the contribution of a chemist to the study of hydraulic fracturing ecotoxicology.

Hydraulic fracturing, first developed in 1949, has expanded to a vast industry, with nearly 2.5 million fracture treatments performed between its inception and 2010.³ The basis of hydraulic fracturing is the injection of a nonexplosive fluid into the ground to stimulate oil wells.³ Currently, most of these fluids are water-based solutions that contain a number of additives.^{3, 4} Since bacterial growth can clog equipment and inhibit gas extraction, it is common to add disinfectants, or 'biocides' to fracturing fluid.⁴ Environmental contamination with fracturing fluid, both through leaching and runoff, has been demonstrated.⁴ Thus, some of these biocides are escaping into the environment, raising questions about the impact of continued exposure on local ecosystems.

Glutaraldehyde, a small, bifunctional molecule, is a frequently-employed fracturing biocide used in nearly 30% of fracturing operations.⁴ Glutaraldehyde is a broad-spectrum biocide, and the mechanisms of its antiseptic action are still being characterized.⁵ However, it is evident that glutaraldehyde acts at cell surfaces and membranes, likely inhibiting transport and causing other structural modifications to impede cell viability.⁵ This is due, at least in part, to its ability to crosslink.⁵

Herein, a comprehensive study is conducted to evaluate the effect of prolonged exposure to fracturing fluid on the ability of a microbial environment to tolerate the biocide glutaraldehyde. This study is multi-faceted, and compiles both chemical and biological data to inform its conclusions. One specific inquiry was aimed to determine if bacterial ecosystems were adapting to more rapidly metabolize glutaraldehyde. UPLC-HRMS method development and chemical analysis were crucial for answering this question.

Chemical biology: the interdisciplinary science

Chemical biology, the collaborative scientific field that applies chemical tools in the study of and intervention in biological processes, is poised to investigate molecular regulators of cellular respiration. Chemical biologists are progressively applying techniques such as organic synthesis and mass spectrometry to elucidate the roles of molecules in metabolism. Herein, a chemical biology approach is used to investigate the expression and function of two distinct types of respiration-regulating biomolecules in two of life's

respective kingdoms. While these two investigations are distinctly unique, they can each be thoroughly studied with a similar toolbox.

Chemical biology, as a field, is largely an applied science. However, to enable this application, continual evolution of methodology is invaluable. In the quest to apply chemical tools in the pursuit of biological answers, methods must be developed, optimized, and adapted to facilitate the desired purpose. In the investigation of these regulatory molecules described above, a collection of methodologies is purposefully constructed to generate the desired data. Some of these methods are established chemistries with well-described protocols that are widely available. Others, with more niche application, required substantial optimization, modification, and even *de novo* development. While the projects described herein are introduced with substantial emphasis on the biological systems in question, an equal amount of attention is given to the development and improvement of methodology. In the current work, many of the initial conclusions are based on methodological successes that are stepping stones on the way to garnering biological data.

Central metabolism and ancillary networks

For much of recent scientific history, biochemical investigations of observed phenotypes have focused on the central dogma of molecular biology: DNA is transcribed to RNA, which is then translated to proteins with biological functions. Thus, many research efforts have focused on the expression of macromolecules, whether nucleic acid or protein. In comparison, small and medium-sized biomolecules have traditionally received less research attention. Focus is continually expanding, however, to include investigation of smaller molecules. Efforts to understand and exploit the functions of small molecules are rapidly mounting. These studies are revealing promising candidates for regulation or characterization of phenotype expression. While many of these analyses were initially focused on a few essential pathways of small-molecule metabolism, modern scientific discoveries have continued to unveil new pathways and metabolite classes—constantly reinforcing the complexity of small molecule networks and functions.

Cellular respiration is the set of central metabolic pathways that organisms use for energy production. Despite the incredible diversity among the kingdoms of life, many of the processes involved in cellular respiration are conserved across all living organisms. In aerobic respiration, glycolysis converts nutrients into 3-carbon molecules which continue through the tricarboxylic acid (TCA) cycle with the aid of other nutrient-derived metabolites. These pathways culminate with oxidative phosphorylation to synthesize adenosine triphosphate (ATP), the ubiquitous energy currency molecule. In eukaryotes, the TCA cycle and oxidative phosphorylation are localized in the mitochondria—energy production organelles. In prokaryotes, the process is similar, but occurs throughout the cell in the cytoplasm or by complexes anchored in the plasma membrane. In both cases, electrochemical gradients are essential for the function of the culminating protein complex, ATP synthase (**Figure 1-1**).

From glucose, to pyruvate, to ubiquinone, there are many key small-molecule players in cellular respiration. Together, they comprise the most recognized molecular framework in metabolism. While these are the direct facilitators of cellular respiration, other distinct classes of biomolecules can play various regulatory roles. Membrane disruptors are a broad class of molecules that can impact the central aerobic energy pathways by interference with respiration. These regulators may produce advantageous effects under optimal expression and may be leveraged to impact health and disease if properly

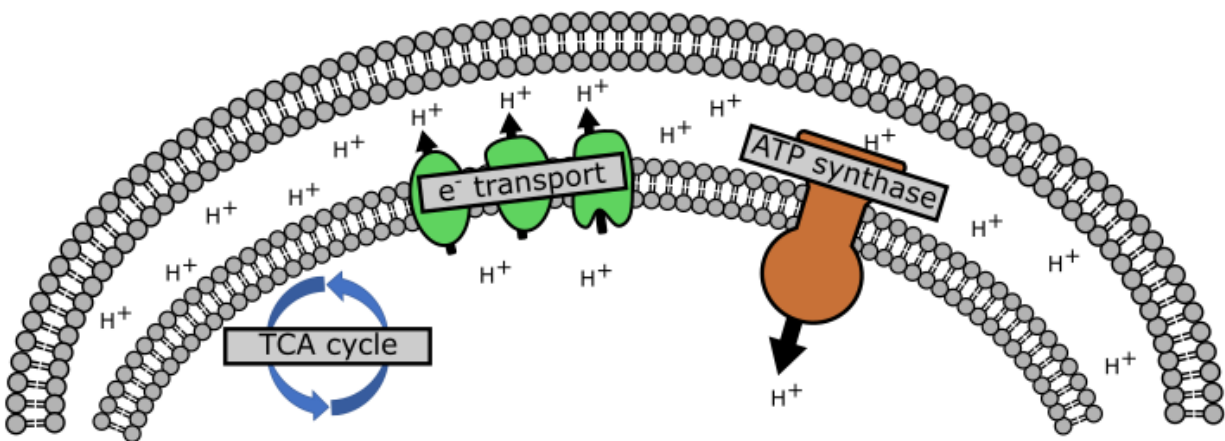


Figure 1-1: Cellular respiration

understood. While some of these regulators are macromolecular, there are current efforts to investigate small and medium-sized molecules that effect cellular respiration. In contrast to the metabolites involved in the central energy pathways that have well-characterized functions, many of these ancillary regulators are secondarily expressed and therefore less studied; however, they may be essential for understanding and leveraging adaptive metabolism. The second part of this dissertation (Chapters Three and Four) is focused on molecules that are believed to regulate cellular respiration.

Small-molecule mitochondrial uncouplers and obesity management

Mitochondrial uncouplers are molecules which disconnect nutrient metabolism from its terminal step of ATP synthesis.^{6,7} These inhibit ATP synthesis without affecting electron transport or ATP synthase. The predominant mechanisms of uncoupling include disruption of the proton gradient and membrane depolarization (**Figure 1-2**). Thus, the driving force of ATP synthase is abolished without interrupting any other respiration pathways.⁷ In animals, mitochondrial uncouplers can act as ways to promote adaptive thermogenesis and energy expenditure. Therefore, mitochondrial uncouplers attract attention as therapeutics in obesity mitigation.⁶ Until 1938, 2,4-dinitrophenol (DNP) was used by approximately 100,000 patients as weight loss drug.⁶ However, its pharmaceutical use was banned in the United States because of safety concerns despite some evidence that DNP can extend lifespan and treat metabolic diseases related to obesity. Exogenous uncouplers continue to pose risks when used as therapeutics, so endogenously-produced uncouplers illicit more excitement in the healthcare community.⁸

Some of the most well-studied endogenous uncouplers are macromolecules. Mammals use uncoupling protein 1 (UCP1), which is expressed in brown and beige adipose tissue, as a prominent means of regulating respiration to facilitate nonshivering thermogenesis.⁹¹⁰ When considered for therapeutic application, small molecule mitochondrial uncouplers are more desirable than protein uncouplers due to the comparative ease of synthesis and delivery. Avians lack UCP1 and therefore serve as suitable model organisms for investigation of small molecule mitochondrial uncouplers.

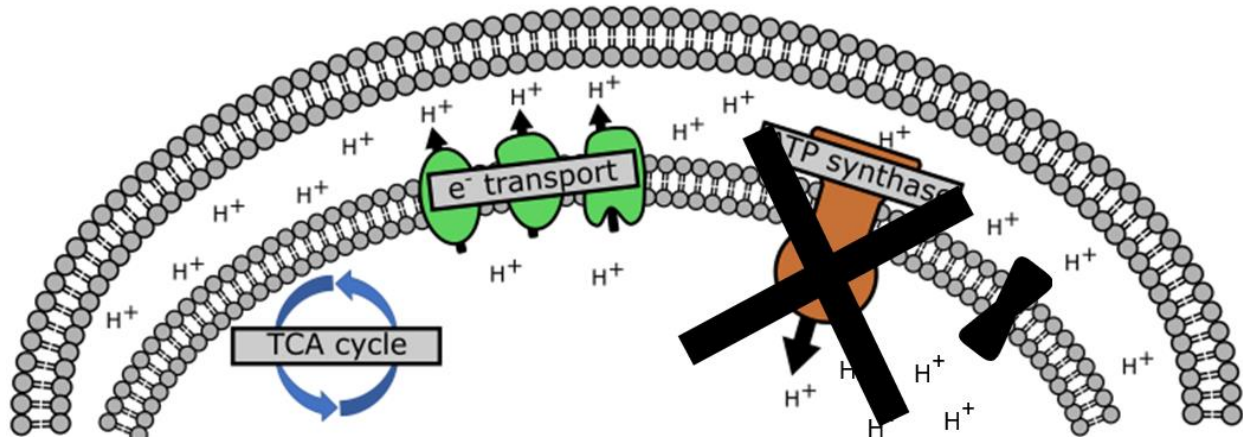


Figure 1-2: Effect of uncouplers on central energy metabolism. Interruption of the proton gradient (shown in black) interrupts the proton motive force and inactivates ATP synthase.

Some of the most well-studied endogenous uncouplers are macromolecules. Mammals use uncoupling protein 1 (UCP1), which is expressed in brown and beige adipose tissue, as a prominent means of regulating respiration to facilitate nonshivering thermogenesis.^{9, 10} When considered for therapeutic application, small molecule mitochondrial uncouplers are more desirable than protein uncouplers due to the comparative ease of synthesis and delivery. Avians lack UCP1 and therefore serve as suitable model organisms for investigation of small molecule mitochondrial uncouplers.

N-Acyl amino acids (*N*-AAAs), condensation products of amino and fatty acids, have been recently shown to uncouple mitochondria for thermogenesis.¹¹ These small molecule uncouplers are produced in many animals including mice, humans, and avians. Because of their endogenous expression, they are especially attractive candidates for obesity therapy.⁸ Given the recently emerging interest in *N*-AAAs and energy expenditure, little is known about their expression across phenotypes. Therefore, systematic studies designed to understand the ties between endogenous *N*-AAA profiles and dietary variations are valuable to both medicine and basic science. Using broiler chicks as a model organism, expression of *N*-AAAs can be monitored throughout fasting-induced lipolysis and among various fat-enriched diets. Ultimately, this should inform future investigations of *N*-AAAs in therapeutic roles.

Membrane disruptors in the intermediate molecular space

Given the emergence of metabolomics and other techniques, studies focused on small biomolecules are increasing. Yet, despite the emergent focus on small molecule metabolites, little attention has been directed towards the intermediate space between biopolymer macromolecules (>10,000 Da) and small molecules (<1,500 Da). These medium-sized molecules span a variety of structural families including polypeptides containing fewer than 50 amino acids. These peptides, which some consider too small to even bear the designation of “protein”, can no longer be overlooked.¹² Expressed by direct translation of small open reading frames (ORFs), these small proteins have been implicated in several biological roles in recent years.¹² Their localization is often pertinent to function, and they are often positioned in or at cell membranes.¹²

Z-protein often repeated (ZorO), a 29-amino acid peptide, is one such small protein. Encoded by a small ORF in wild type enterohemorrhagic *Escherichia coli* (EHEC), ZorO appears to confer toxicity to colonies.^{13, 14} When overexpressed, it initiates cell death. However, it is suspected that endogenous ZorO expression may serve to regulate cell growth in response to environmental stressors. Once this growth stasis has served its purpose, cells self-rescue by expression of an sRNA antitoxin (*orzO*) which prevents translation of ZorO-encoding mRNA.¹⁴ This toxin-antitoxin locus is absent in domesticated laboratory strains of *E. coli*, indicating that it may be valuable for growth in indigenous environments.

Like other small proteins, ZorO appears to be membrane localized.^{13, 15} Preliminary data has shown that ZorO may oligomerize in the inner membrane, creating ion channels which disrupt the proton gradient that drives ATP synthase—thus uncoupling nutrient metabolism from ATP generation. In this way, it may be acting as a prokaryotic equivalent of a mitochondrial uncoupler (**Figure 1-2**). This toxin-antitoxin locus is a valuable research target, as it may facilitate an improved understanding of pathogenic *E. coli* resilience or a selective means of combating EHEC infections. To investigate these possibilities, detailed studies are needed to fully elucidate the structure, function, expression, and localization of the ZorO peptide.

Addressing the need for interdisciplinarity in science education

As discussed above, interdisciplinary research is on the rise and has become prominent in science. However, university-level undergraduate science education has largely maintained the separation between disciplines by establishing distinct departments. As a final component of this work, an interdisciplinary laboratory course is intentionally designed to familiarize undergraduates with chemical biology. The conception of this course was built on a foundation of research experience and peer-reviewed literature, drawing from published experiments and incorporating many of the techniques introduced earlier in this dissertation. This novel course was independently developed to highlight innovative applications of chemical methodology towards the elucidation of biological structures and mechanisms. Specific focus is placed on influential chemical techniques and prominent classes of biomolecules. Since its inception, the course was piloted for several consecutive semesters and is undergoing further evolution to more effectively introduce undergraduate scientists to scientific interdisciplinarity. Ultimately, the implementation of such a course will serve the scientific community by graduating chemistry majors who are more aptly poised for either collaborative research trajectories or for employment in health, life, and forensic sciences.

**CHAPTER TWO: IMPACT OF HYDRAULIC FRACTURING ON
DEGRADATION POTENTIAL OF THE BIOCIDES GLUTARALDEHYDE**

Abstract

The environmental impacts of hydraulic fracturing, particularly those of surface spills in aquatic ecosystems, are not fully understood. The goals of this study were to: (1) understand the effect of previous exposure to hydraulic fracturing fluids on aquatic microbial community structure and (2) examine the impacts exposure has on biodegradation potential of the biocide glutaraldehyde. Microcosms were constructed from hydraulic fracturing-impacted and non-hydraulic fracturing-impacted stream water within the Marcellus shale region in Pennsylvania. Microcosms were amended with glutaraldehyde and incubated aerobically for 56 days. Microbial community adaptation to glutaraldehyde was monitored using 16S rRNA gene amplicon sequencing and quantification by qPCR. Abiotic and biotic glutaraldehyde degradation was measured using ultra performance liquid chromatography—mass spectrometry and total organic carbon. It was found that non-hydraulic fracturing-impacted microcosms biodegraded glutaraldehyde faster than the hydraulic fracturing-impacted microcosms, showing a decrease in degradation potential after exposure to hydraulic fracturing activity. Hydraulic fracturing impacted microcosms showed higher richness after glutaraldehyde exposure compared to non-impacted streams, indicating an increased tolerance to glutaraldehyde in hydraulic fracturing impacted streams. Beta diversity and differential abundance analysis of sequence count data showed different bacterial enrichment for hydraulic fracturing-impacted and non-hydraulic fracturing-impacted microcosms after glutaraldehyde addition. These findings demonstrated a lasting effect on microbial community structure and glutaraldehyde degradation potential in streams impacted by hydraulic fracturing operations.

Preface

The work presented in this chapter is published in a manuscript entitled *Impacts of Glutaraldehyde on Microbial Community Structure and Degradation Potential in Streams Impacted by Hydraulic Fracturing* (DOI: 10.1021/acs.est.8b00239). It has been reproduced with permission from *Environmental Science & Technology* (2018, 52 (10), pp 5989–5999) copyright 2018 American Chemical Society. The contribution of Caleb M. Gibson (CMG) to this publication is the UPLC-HRMS method development as well as

generation and interpretation of the chemical data. Maria Fernanda Campa (MFC) conceived the experiment with input from Stephen M. Techtmann (SMT) and Terry C. Hazen (TCH). Nikea Ulrich (NU), Regina Lamendella (RL), and Christopher J. Grant (CJG) organized the sampling expedition, and MFC and CJG collected samples. MFC, CMG, Megan Patterson (MP), and Amanda Garcia de Matos Amaral (AGMA) performed the experiments. MFC, CMG, Xiaojuan Zhu (XZ), and Shawn R. Campagna (SRC) analyzed the data.

In order to preserve the complete account, the manuscript is included in its entirety; however, special attention is given to the chemical methodology and data.

Forward

The aim of this collaborative project was to evaluate the effect of hydraulic fracturing on the local microenvironment. Namely, tolerance of and adaptation to the biocide glutaraldehyde were evaluated in microbial communities exposed to hydraulic fracturing and unexposed control groups. Initially, the hypothesis was that microbial communities exposed to hydraulic fracturing would adapt to more rapidly metabolize glutaraldehyde than unexposed communities. To test this, a UPLC-HRMS method was developed to quantify glutaraldehyde over time in cultured microenvironments. To do so, careful consideration was given to the solution equilibrium and potential reactivity of glutaraldehyde to ensure that an accurate picture of glutaraldehyde concentration was achieved.

Mass spectrometric data illustrated a trend opposite of the hypothesis, with glutaraldehyde degrading more quickly in communities unexposed to hydraulic fracturing. To confirm this observed trend, the oxidized metabolite of glutaraldehyde, glutaric acid, was also quantified. Corroborating the trend in glutaraldehyde degradation, glutaric acid production was increased in microenvironments previously unexposed to fracturing fluid. This raises an intriguing discussion about the propensity of the microbial environment to naturally remediate after prolonged exposure and the strategy for intervention.

Introduction

The use of hydraulic fracturing (HF) has grown 702% since 2007.¹⁶ Since 2011, seven shale plays have been responsible for more than 90% of the oil and gas production growth in the U.S. The most productive of these plays is the Marcellus Shale in the northeastern U.S., producing more than 18,000 mcf of natural gas per day.¹⁷ Despite the proposed economic and energy security benefits of HF, many environmental questions and potential unforeseen consequences remain. The exact mixture of chemicals and water (i.e. HF fluids) used in a HF job is proprietary and dependent on company and/or shale play geochemistry. However, HF fluids components often include gelling and foaming agents, friction reducers, crosslinkers, breakers, pH adjusters, corrosion inhibitors, iron control chemicals, clay stabilizers, surfactants, and biocides.¹⁸ Biocides are added to HF fluids to prevent the corrosion, bioclogging of pipes and equipment, and gas souring that are caused by sulfate-reducing bacteria and acid-producing bacteria. High volumes of HF fluids are injected under great pressure to crack open the shales deep beneath the surface. A portion of this fluid then resurfaces as wastewater, called “flowback” water. This flowback fluid requires special handling and disposal. Biocides have been identified as some of the most toxic chemical additives in HF fluids.^{18, 19}

The efficacy of biocides in HF operations is unclear. Previous studies report active and diverse microbial communities in flowback waters despite biocide use.²⁰⁻²⁶ Glutaraldehyde (GA) is the most commonly used biocide in HF.¹⁹ There are a number of ways GA can degrade abiotically in the environment. The compound is water miscible and does not tend to bioaccumulate. It hydrolyzes as pH increases, and it can also be photo-degraded.^{27, 28} Previous studies have shown that GA is biodegradable under aerobic and anaerobic conditions, but degradation rates can be affected by concentration, pH, salt, temperature, chemical interactions, and bacterial resistance.²⁷⁻³¹ Under aerobic conditions, GA can be biodegraded to carbon dioxide via glutaric acid, and under anaerobic conditions, the biocide is metabolized to 1,5-pentanediol.²⁸ Despite its biodegradability, GA is considered acutely toxic to both terrestrial and aquatic organisms – freshwater fish in particular—at concentrations as low as 2.5 mg/L for embryos and 4.7 mg/L in adult fish populations.³²

To date our review of the literature suggests that few if any studies have examined the fate of GA in an aquatic environment previously exposed to HF fluids. To address this gap, this study employs a combination of next generation sequencing and detailed chemical analysis. The goal of the study is to understand how GA affects aquatic microbial communities previously exposed to HF fluids and to measure the degradation of GA in an exposed aquatic system as compared to a non-exposed aquatic system.

Materials and methods

Stream selection

Streams were selected using Pennsylvania (PA) Department of Environmental Protection records and GIS surveys. The sampling area was forested and there were no physical indications of past mining activity prior to HF development in that region. The selected streams had minimum variation in watershed characteristics caused by anthropogenic impacts other than HF. There was no indication of conventional drilling, acid mine drainage, or other industrial activities. Each of the HF-impacted (HF+) streams selected had either a history of surface spills (stream names: Alex Branch (AB) and Little Laurel (LL))³³ or more than 20 well-heads (unnamed tributary (UNT) Naval Hollow) in the vicinity.³⁴ In 2009, LL received flowback from a broken pipe for over two months, to a lesser extent AB also received flowback from the same pipe. Furthermore, AB received input from an 8,000-gallon spill of water and HF fluids. Each of the HF-not impacted (HF-) streams, UNT East Elk Fork (EE), UNT West Elk Fork (WE), and Dixon Run (DR), selected as baseline, had Fracking well construction in its vicinity, but no HF activity had commenced.

There was documented use of GA in wells associated with the three HF+ streams selected according to FracFocus.org.³⁵ Detailed selection of streams, screening process, collection and description of the sites have been discussed elsewhere.³⁶⁻³⁸ Past studies surveying these and other streams in central and northwestern PA showed that the microbial community composition and indicator taxa can be used to predict HF past exposure, even years after a documented spill.^{36, 39} Indicator taxa enriched in streams exposed to HF wastewater were also present in streams with adjacent HF operations, but

no history of spills. This suggests that direct spills are not the only source of HF impacts in the aquatic ecosystem.^{36, 39} In addition to persistence of microbial indicator taxa, streams in North Dakota impacted by flowback water spills maintained the geochemical and isotopic signatures of the spill for up to 4 years after documented spills.⁴⁰

Sample collection

Stream water was collected from HF+ and HF- streams in northwestern PA in June 2015 using sterile Nalgene bottles. All streams were sampled within a two-week period, and depending on the stream, were stored for 3 or 4 weeks at 4°C until use. Geochemical parameters, including temperature, pH, conductivity, total dissolved solids, and salinity were measured at the time of sample collection with a Eutech PCSTestr, 35 Multi-parameter test probe that was calibrated weekly using a three point-calibration.⁴¹ Refer to **Figure 2-1** for a map of watersheds' location.

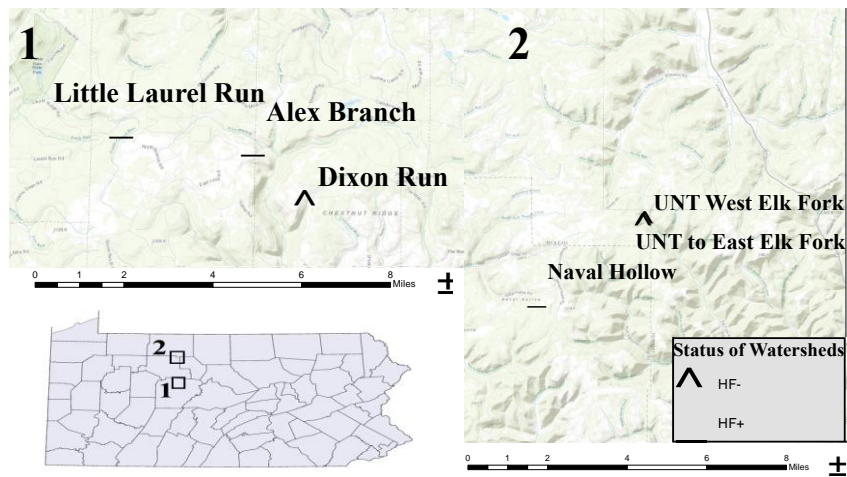


Figure 2-1: Map showing the location of the PA watersheds used as microcosms source water

Microcosm setup

The amount of biocides used in HF fluids varies widely between 10 to 800 mg/L.¹⁸ Dow Chemicals has shown a 6-log reduction of acid producing bacteria and sulfate reducing bacteria, the standard in the oil and gas industry, at a concentration of 100 mg/L of GA.⁴² Thus, microcosms were established with 260 mL of stream water amended with 100 mg/L of GA. A 50% solution of GA (CAS number 111-30-8, catalog number 340855) was bought from Sigma Aldrich (St. Louis, MO, USA). Abiotic controls were autoclaved prior to GA amendment to measure abiotic biocide degradation. Additionally, negative biological controls were setup with stream water and no GA addition to examine bottle effect on the microbial community. Both control sets had a volume of 20 mL. All microcosms were setup in triplicate and incubated for 56 days under minimal light exposure and at ambient temperature. Microcosms were uncovered only for sampling events.

Ultra-performance liquid chromatography-mass spectrometry (UPLC-MS)

Abiotic and biotic microcosms were sampled at day 1 before and after amendment of GA, and at day 7, 28, and 56. One mL of water from each microcosm was collected, filtered through a 0.2 µm Sterivex nylon filter, and frozen at -20°C until analysis at the University of Tennessee's Biological Small Molecule Mass Spectrometry Core. The samples were diluted 1:10 with HPLC grade water. A 10 µL injection volume of each sample was subjected to UPLC separation (LC Dionex Ultimate 3000) on a Synergi 2.5 µm Hydro-RP 100 Å, 100 x 2 mm column. (Phenomenex, Torrance, CA). Solvent A consisted of 0.1% formic acid in water, and solvent B was 0.1% formic acid in acetonitrile. The separation gradient featured an initial ramp from 0% to 50% B over 6.5 min, and the conditions were held constant for 1 min. This was followed by a return to initial conditions over 0.25 min and a 3.5 min equilibration at 0% B for a total runtime of 11.25 min. The flow rate was held constant at 300 µL/min. Mass spectra were recorded in positive mode with an Orbitrap Exactive Plus mass spectrometer (Thermo Scientific, Waltham, MA) under the following parameters: Positive-mode heated electrospray ionization, sheath gas flow of 25 units, aux gas flow of 8 units, capillary temperature of 300°C, aux gas heater temperature of 150°C, spray voltage of 4.2 kV, ACG target of 3x10⁶, resolution of 140,000,

and a scan range of 90 to 300 m/z . GA was detected in positive mode as the $[M+H]$ ($m/z = 101.0600$) with a retention time of 2.8 min. The GA metabolite glutaric acid (purchased from Sigma-Aldrich, CAS number 110-94-1, catalog number G3407) was measured with an identical instrument and column using an established negative-mode ion-pairing UPLC-MS method.^{43, 44} Concentrations were calculated using the standard curves in **Figure 2-2**. Average HF+ and HF- concentrations with their respective standard error were reported.

Total organic carbon (TOC)

TOC associated with GA was quantified at days 0, 7, and 56 using a Shimadzu TOC-L equipped with an ASI-L autosampler (Shimadzu, Kyoto, Japan). 1 mL of sample was filtered through a 0.2 μm Sterivex nylon filter and then diluted 1:25 or 1:10 with DI water acidified to pH 3 with HCL. This released the inorganic carbon present in the samples. Samples were collected prior to GA addition to subtract the background TOC. GA standards were run to calculate TOC associated with GA. Time point 0 sample (after addition of biocide) for the biotic microcosms was depleted during preparation. However, as the same concentration of GA was added to both biotic and abiotic microcosms, the time point 0 TOC measurement for the abiotic microcosms was used to calculate percent loss for both.

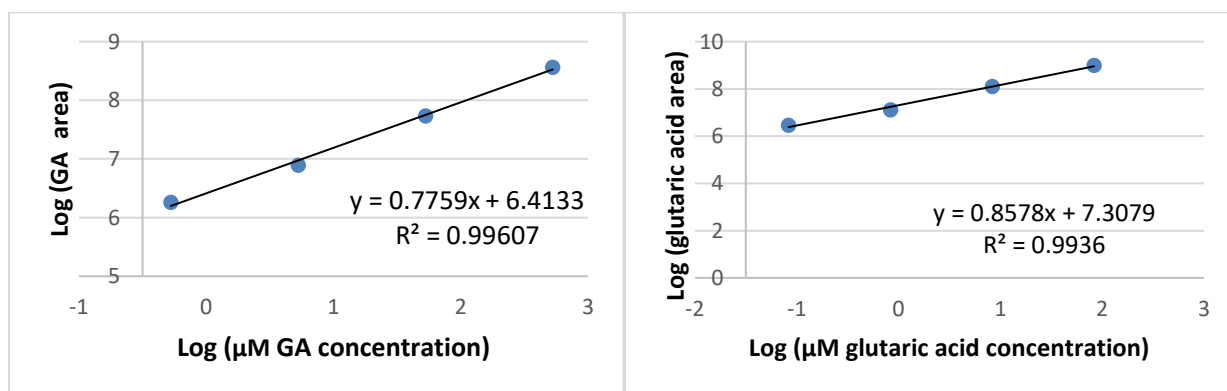


Figure 2-2: Calibration curves for glutaraldehyde (left) and its metabolite glutaric acid (right)

16S rRNA gene amplicon library preparation and sequencing

Bacterial community changes can be used as biosensors for contamination even after the contaminants are fully degraded.⁴⁵ In the case of these streams, prior exposure to GA may lead to microbial adaptation which may affect the degradation of GA. To test this hypothesis, 25 mL of water from the GA amended microcosms was filtered for DNA collection to track microbial community changes and perform qPCR for the 16S rRNA gene. Samples were collected prior to GA amendment and at days 7, 21, 35, 49 and 56. The no-GA control microcosms were sacrificially sampled at day 56 to perform the same DNA-analyses. The water collected was filtered through a 0.2 µm Sterivex nylon filter (Millipore Sigma, St. Louis, MO). Filters were frozen at -20°C until use. The plastic casing of the filter was cracked opened with sterile pliers in a biohazard hood. The filter was removed with sterile tweezers and cut with a sterile knife. The filter pieces were extracted for genomic DNA using PowerSoil DNA isolation kit (Qiagen) according to the manufacturer's manual. The v4 region of the 16S rRNA gene was amplified using the protocol described by Caporaso et al.⁴⁶ Primer dimers were removed using the Select-a-Size DNA Clean and Concentrator Kit (Zymo Research, Irvine, CA) following the manufacturer's instructions to remove fragments smaller than 300 bp. The size of the amplicons were visualized and quantified using a 2100 Bioanalyzer and DNA 1000 chip (Agilent Technologies, Santa Clara, CA). Samples were quantified using a Qubit Fluorometer (ThermoFisher Scientific) and pooled based on equimolar concentrations to a final pool of 10 nM. The concentration was verified using qPCR utilizing KAPA SYBR FAST qPCR kit (KAPA Biosystems, Wilmington, MA). Final pools of 10 nM each were run in an Illumina MiSeq (San Diego, CA) using a v2 kit (2 x 150 reads), according to manufacturer's manual.

Quantification of bacterial 16S rRNA gene

qPCR amplification was performed using the universal bacterial primers Bac1055YF⁴⁷, ⁴⁸ and Bac1392R.⁴⁸ Each 10 µl reaction was loaded using a QIAgility (Qiagen, Hilden, Germany) automated PCR loading robot. Each reaction contained 2 µl of template DNA, 3.94 µl of water, 4.00 µl of Applied Biosystems Power SYBR GreenPCR Master Mix (ThermoFisher Scientific, Waltham, MA), 0.03 µl of 300 nM forward primer and 300 nM

reverse primer respectively. qPCR reactions were performed in a QuantStudio 12K Flex Real-Time PCR system (ThermoFisher Scientific) using the qPCR cycle parameters described in Ritalahti et al.⁴⁷

16S rRNA gene amplicon sequencing data analyses

Data analyses were performed using the QIIME pipeline (version 1.9.1)⁴⁶ and the Phyloseq⁴⁹ package in R.⁵⁰ Briefly, the forward and reverse raw reads were joined using the assembler fastqjoin⁵¹ embedded in QIIME. De-multiplexing and quality filtering was performed at an average Q-score of more than 19. The sequences were then chimera filtered using the UCHIME method and applying the USEARCH program.^{52, 53} Both *de novo* and reference-based chimera detection were used. For the reference-based detection, the Greengenes database (version May 2013)⁵⁴ filtered to up to minimum 97% sequence identity was used. Open reference OTU picking was performed using the command `pick_open_reference_otus.py` using the UCLUST method⁵² using the Greengenes database as described above. Representative sequences for each OTU were aligned using the PyNAST method⁵⁵ and taxonomy was assigned to each representative sequence using the RDP classifier⁵⁶ trained against the Greengenes database.^{54, 57, 58} OTUs were then filtered to remove sequences with counts below 0.005%. The samples were then rarefied to 1,220 sequences. The BIOM file generated was then converted to JSON file and diversity analyses were performed using Phyloseq⁴⁹ using R.⁵⁰

Statistics

Geochemical parameters were compared between HF+ and HF- microcosms using a T-test. GA degradation over time was compared between HF+ and HF- microcosms to test if degradation rates changed based on impact status. Degradation between biotic and abiotic samples was also compared to test if the main driver of degradation was biotic or abiotic. To do this the biocide concentration was log10 transformed and a baseline of 100mg/L was used for day 0. A complete randomized design (CRD) with a split-split plot was used. Impact statuses (HF+ and HF-) were assigned to the whole plot and applied to two levels of conditions (biotic and abiotic) for the sub-plot. Microcosms' samples were

taken for measurement at days 7, 28, and 56 (sub-sub-plot). Data was then divided between biotic and abiotic. A CRD with repeated measures was applied to each. The same was performed with glutaric acid concentrations. The mixed effect ANOVA method was employed to analyze the data using SAS 9.4, and least squares means separated with a Bonferroni method. The alpha level was set at $P = 0.05$. A Pearson correlation of pH and GA concentrations was performed for day 56.

The 16S rRNA gene abundance was compared to understand the effect previous exposure to HF fluids have on aquatic microbial community structure after GA addition. This was done using a CRD with split plot using impact status (HF+ vs. HF-) as the whole plot factor and time (days) as the split plot factors using a mixed effect ANOVA model (R nlme package⁵⁹). The least squares means were computed and separated with Bonferroni method (R emmeans package⁶⁰). 16S rRNA gene copies/mL were log₁₀ transformed to meet normality and variance assumptions for ANOVA. To compare the no-biocide control at day 0 and at the end of the experiment (day 56), the same model was run. To determine differences between HF+ and HF- at day 0, an independent sample T test was run with data for only that time point.

Alpha diversity, beta diversity and DESeq2⁶¹ analyses were performed using un-rarefied OTU table. Difference in community evenness and richness between HF+ and HF- streams was measured using Simpson, Chao1, Observed diversity, and Shannon alpha diversity metrics. Alpha diversity results were used as a proxy for microbial resistance and resilience against GA. Microbial community alpha diversity values were rank transformed and compared using the same model as for 16S rRNA gene copies/mL. Beta diversity measures were calculated using weighted UniFrac distance matrix⁶² and visualized using a principal coordinate analysis (PCoA). Finally, microbial community beta diversity was compared using a nested PERMANOVA using the adonis command in the VEGAN⁶³ R package.

The DESeq2⁶¹ R package was used to find microbial taxa enriched through time by comparing each time point (day 7, 21, 35, 29, and 56) to the day 0 No-GA control. Day 56 was also compared to the day 56 No-GA control, and both day 0 and day 56 No-GA

controls were also compared. At each time point comparison between HF+ and HF- was performed to identify differentially enriched taxa between HF impact status. The same was performed with HF- streams. The Wald test⁶¹ was performed using the parametric fit-type and a Benjamini & Hochberg adjusted P-value with an alpha < 0.01. Biomarkers without Family level taxa classification were removed.

Results and discussion

The objective of this study was to understand the lasting effect of HF impacts on the biocide resistance and degradation potential of surface water microbial communities. To do this, GA—the most common biocide used in the HF industry—was added to microcosms of water from streams impacted and not impacted by HF as determined by previously published studies.^{36-38, 64}

Physiochemical parameters of stream water in situ

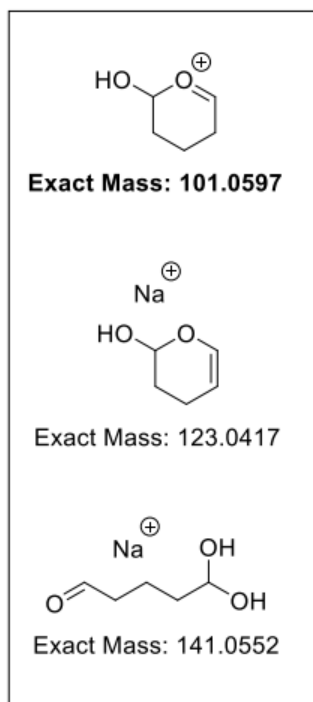
Temperature, pH, conductivity, total dissolved solids, and salinity were measured at the time of sample collection and results are shown in Table S1. HF+ streams had an average temperature of 16.8°C and HF- streams had an average temperature of 12°C. HF+ streams had an acidic pH averaging 4.9, while HF- had a neutral pH of 6.5. The average conductivity for HF+ streams was 26.3 µS/cm, and for HF- streams 33.7 µS/cm. The average total dissolved solids for HF+ was 23.67 ppm and for HF- 23.93. Finally, the average salinity for HF+ streams was 19.2 ppm and 18.9 ppm for HF- streams. There were no statistically significant differences in the physiochemical parameters between HF+ and HF-. It is worth noting that others have documented higher conductivity in surface waters impacted by unconventional oil and gas extraction (UOG) wastewaters⁶⁵ as UOG wastewaters are high in salinity.^{20, 26} The streams described by Akob et al.⁶⁵ were impacted by their proximity to UOG wastewater disposal facility, which suggests that the high salinity could have been caused either by a recent spill or constant inflow of wastewater to the streams. In that study, the pathway of contaminants to disposal facility could not be assessed. However, a 5 year-long study of these 6 streams and others in northwestern PA consistently showed that pH was the only statistically different measured

parameter between the impacted and not impacted streams³⁹, indicating that a one-time spill is not enough to alter conductivity for a long time as input waste is diluted overtime.

GA speciation

Given the probability that GA exists in equilibrium with many different hydrated forms, searches were conducted for any other GA related chromatographic peaks. In both the standard and experimental samples, a peak corresponding to the sodium adduct of the GA hydrate ($[M+H_2O+Na+] = 141.0526$ m/z) was observed (**Figure 2-3**) at an identical retention time to the molecular ion (2.8 min), indicating that hydration/dehydration was occurring in-source. Likewise, the sodium adduct of molecular ion was observed ($[M+H_2O+Na+] = 123.0417$ m/z, r.t. 2.8 min) (**Figure 2-3**). However, these additional peaks were minor components compared to the observed molecular ion, with areas and intensities at least 1 order of magnitude lower to the molecular ion. The detection of multimeric forms of GA has been addressed in previous reports.⁶⁶ Ferrer and Thurman⁶⁶ analyzed GA (among other HF additives) and detected peaks for GA oligomers. The oligomers are formed in solution by aldol condensation instead of in-source, as evidenced by their separate retention times. In the current study, GA standards and samples produced a peak corresponding to the sodium adduct of the singly hydrated aldol dimer ($[M+H_2O+Na+] = 223.0941$ m/z, r.t. 4-5 min) (**Figure 2-3**). Additionally, a peak was observed that corresponded to the mass of the sodium adduct of a doubly hydrated aldol dimer ($[M+2H_2O+Na+] = 241.1038$ m/z, r.t. 4-5 min) formed in-source (**Figure 2-3**). In all cases, the area and intensity of the dimer peak was at least 2 orders of magnitude lower than that of the parent molecular ion of GA. The mass range of the experiment excluded the sodium adduct of the doubly hydrated aldol trimer observed in past reports⁶⁶, and no other forms of the trimer were detected. Despite the presence of other detected forms of GA, the chromatographic peak for the chosen molecular ion of 101.0600 m/z is the best means of quantitation (**Figure 2-3**). If any environmental variable between streams influenced the detected amount of GA, then the changes can be reflected in the abiotic controls, which displayed constant GA among streams and over time as discussed below.

Median retention time = 2.8 - 2.9 min



Median retention time = 4.3 - 4.4 min

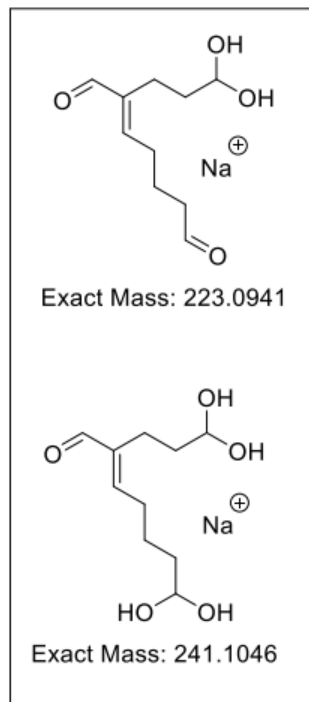


Figure 2-3: Glutaraldehyde species detected

GA abiotic and biotic degradation over time measured with UPLC-MS

It was observed that abiotic degradation of GA was negligible and independent of HF impact status and the difference in GA concentrations between HF+ and HF- abiotic microcosms through time was not statistically significant. The final concentration of GA in the abiotic HF+ control was 101.9 ± 4.2 mg/L and 106.79 ± 5.1 mg/L in abiotic HF- control (**Figure 2-4a**). Additionally, biotic degradation of GA was detected in both HF+ and HF- microcosms. The final concentration of GA in the HF+ biotic microcosms was 47.3 ± 5.2 mg/L and in the HF- biotic microcosms was 31.7 ± 3.8 mg/L. The difference in degradation over time was found statistically significant ($P < 0.05$). The HF- communities degraded GA faster by day 56, a 68.3% removal of GA with half-life of 33.8 d, while HF+ experienced a 52.7% removal with half-life of 51.9 d.

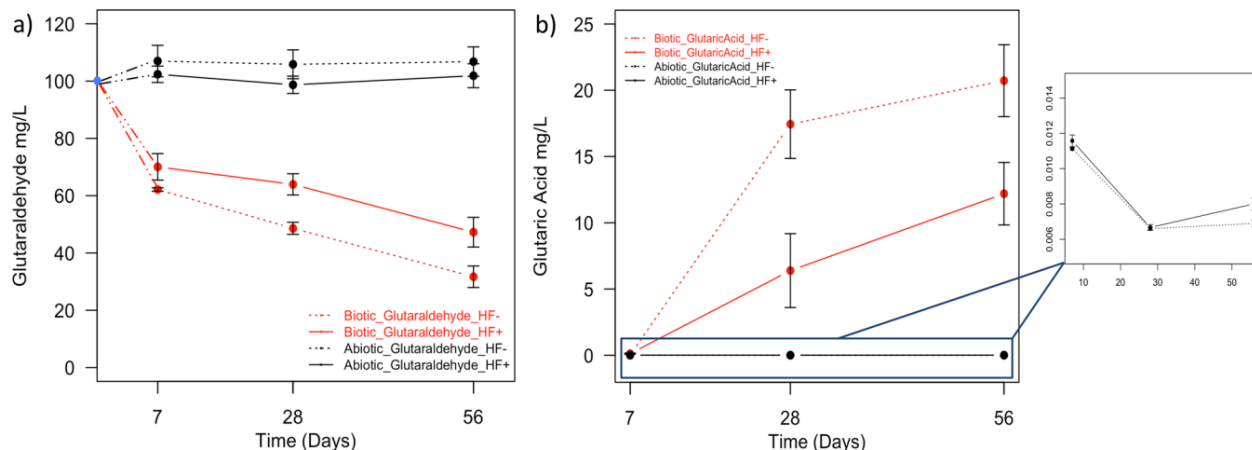


Figure 2-4: Biotic and abiotic degradation of glutaraldehyde and glutaric acid production over time. (a) Biotic and abiotic degradation of glutaraldehyde in HF+ and HF- microcosms. The blue dot represents the added amount of GA, 100 mg/L. (b) Biotic and abiotic production of glutaric acid in HF+ and HF- microcosms; the zoom in graph shows abiotic concentration over time. Error bars represent one standard error (n = 9).

Glutaric acid is a known degradation product of the oxidation of GA²⁷. Glutaric acid was produced in the microcosms, validating the GA degradation measurements. Minimal production of glutaric acid was observed in the abiotic microcosms, with pronounced production in the biotic microcosms (**Figure 2-4b**). By day 56, abiotic HF+ microcosms produced 8.0 ± 1.0 $\mu\text{g/L}$ of glutaric acid and abiotic HF- microcosms produced $6.9 \pm .5$ $\mu\text{g/L}$. This difference was statistically significant ($P < 0.05$). Meanwhile, by day 56, 12.2 ± 2.4 mg/L of glutaric acid were produced in the HF+ biotic microcosms and the HF- biotic microcosm produced 20.7 ± 2.7 mg/L. The difference between the abiotic and biotic glutaric acid production and the difference between biotic HF+ and biotic HF- glutaric acid production over time were also statistically significant ($P < 0.05$). The steady increase of glutaric acid in the biotic microcosms as compared to the abiotic microcosms shows that the main pathway of GA depletion after day 7 is microbially mediated.

Other studies have shown abiotic degradation of GA in oxic and anoxic conditions, but their experimental conditions included soil, where GA can be lost to sorption.^{28, 30} However, in this study the rate of biotic degradation in both HF+ and HF- microcosms

was slower than the rates reported in the review by Leung.²⁸ Leung describes the degradation of lower concentrations (0.9 to 50 mg/L) than the study described here (100mg/L) and GA degradation was indirectly quantified in the review using oxygen, carbon dioxide, or dissolved organic carbon measurements as proxies for GA degradation.²⁸ Leung reported a variable GA half-life of 0.4-24 d, due to enhanced microbial inhibition at higher GA concentrations, which increases the half-life of GA. Another study measuring the biodegradation of GA in combination with 5 other HF chemicals also showed an increase in GA half-life at increasing concentrations.³⁰ In that study, microcosms containing 100 mg/L of GA did not experience more biodegradation than the abiotic controls, indicating complete microbial inhibition, with an extrapolated half-life of more than 93 d. The addition of 5 other HF chemicals could have exacerbated microbial toxicity, particularly as the inoculum in those microcosms came from pristine soil with no previous exposure to HF chemicals.³⁰ In this study, the HF+ source water had prior exposure to HF, and there was no competing chemical interactions or toxicity to inhibit microbes other than GA.

GA associated TOC in biotic and abiotic microcosms

It was observed that TOC decreased in the first 7 days (day 0 to 7) for both abiotic and biotic microcosms (Figure S6). After the initial TOC reduction, abiotic microcosms stayed constant, and by day 56 there was 8.64% removal in HF+ and 7.04% removal in HF-. In contrast, the biotic microcosms observed a higher TOC removal by day 56, 57.06% removal in HF+ and 62.81% removal in HF-. These findings agree with the trends observed with direct GA and glutaric acid measurements by UPLC-MS, showing a pronounced difference between biotic and abiotic degradation, and HF- microcosms degrading GA faster than HF+.

The decrease in TOC after GA addition may suggest a decrease of GA in the first 7 days in both biotic and abiotic microcosms. This correlates to what McLaughlin et al.²⁹ observed in their microcosms with agricultural topsoil and synthetic surface water. However, they attributed this effect to GA absorption into the soil, either by physiosorption or chemisorption.

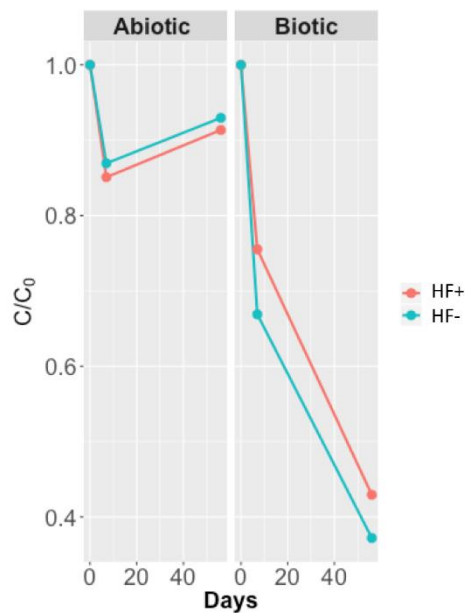


Figure 2-5: TOC over time

Because the microcosms described here did not have sediment as a confounding variable, it is likely that the observed initial depletion was from less prominent reversible GA hydrates forming in solution (**Appendix A1**). The results indicate that GA persists longer in a sediment free aquatic environment than in a sediment-water matrix such as the one described by McLaughlin et al.²⁹ as their reported half-life for GA was 10 d. Furthermore, previous HF impacts and pH variations may increase GA persistence in the environment.

qPCR

The abundance of 16S rRNA genes was determined from initial samples before addition of GA (**Figure 2-6**). All the pre-GA treatment 16S rRNA gene concentrations were on the order of 10^4 gene copies/ mL, averaging 4.03×10^4 gene copies/mL in the HF+ streams and 4.38×10^4 gene copies/ mL in HF- streams.

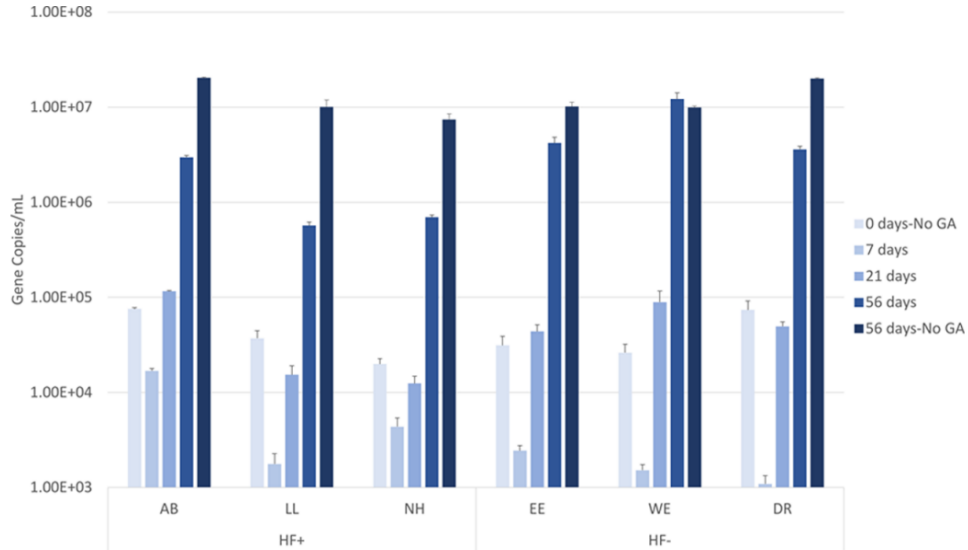


Figure 2-6: Impacts of glutaraldehyde in abundance of 16S rRNA gene over time. The first three clusters are the HF-impacted streams, and the last three clusters represent the non-HF-impacted streams. Data point “56 days-No GA” represents bottle effect on the microcosms as no GA was added. Error bars represent one standard error (n = 3).

The difference between HF+ and HF- was not statistically significant. Seven days after addition of GA, 16S rRNA gene copy number observed a \log_2 fold change (FC) in all microcosms independent of HF impacted status. HF+ microcosms showed a smaller change with an average of $-2.92 \log_2$ FC compared to $-4.62 \log_2$ FC in HF- microcosms. However, by day 21 the bacterial population recovered, returning to the original order of magnitude and with all streams surpassing the original gene copies except for NH (HF+) and DR (HF-), which were slightly lower. HF+ streams had an average of $-0.45 \log_2$ FC from the original gene copies/mL on day 21, while HF- streams had surpassed the original concentration with an average of $0.56 \log_2$ FC. Finally, by day 56 all of the microcosms underwent 16S rRNA gene enrichment, exhibiting a higher enrichment on HF- microcosms. Additionally, HF+ microcosms underwent a $4.79 \log_2$ FC from day 0, while HF- was $7.18 \log_2$ FC. The difference through time (day 7 to 56) between HF+ and HF- was statistically significant ($P > 0.05$). In contrast, at day 56 the no-GA controls had a similar \log_2 FC, independent of previous HF status. No-GA HF+ microcosms had an average of $8.23 \log_2$ FC while no-GA HF- has an $8.34 \log_2$ FC, which was not statistically

significant. When compared the same time points, microcosms with no GA had higher 16S rRNA gene copies/mL at day 56 than the GA-amended microcosm. This can be attributed to the GA-free microcosms not experiencing inhibited growth and having sufficient nutrients from the source water to promote growth. Thus, without GA addition, the biomass of the microbial communities increased to the same final gene copies/mL, showing that the difference in gene copies/mL between the GA-amended HF+ and HF- microcosms can be attributed to the microbial community response to GA.

Quantification of the 16S rRNA gene also showed that HF+ microcosms were able to tolerate and resist the biocide better than HF- microcosms at day 7, the critical response phase to GA biocidal action (**Figure 2-6**). However, both HF+ and HF- microbial communities recovered rapidly after 21 days suggesting adaptation by certain microbial populations and enrichment of those microbes able to tolerate and resist GA in both the HF+ and HF- water, especially as GA's concentration decreases over time. Furthermore, the differences in 16S rRNA gene copies over time showed that HF+ and HF- microcosms had a distinct adaptation and tolerance to GA.

Microbial community changes between HF+ and HF- over time

Richness, as measured by Shannon, Observed Species, and Chao1, showed that before GA amendment HF- streams were more diverse than HF+ streams ($P < 0.05$) (**Figure 2-7**) while the difference was not significant for Simpson alpha diversity measurements. Seven days after addition of GA, HF+ maintained higher richness and evenness than HF-, a significant trend observed with Chao1, and Observed diversity measurements ($P < 0.01$ through the duration of the experiment) but not with Simpson and Shannon. The interaction between impact status (HF+ and HF-) and days was not significant. A comparison of no-GA control microcosms at day 0 and day 56 showed that there were no significant changes in alpha diversity (Observed, Chao1, and Simpson) over time except with Shannon diversity ($P < 0.05$). Thus, the control (no GA added) at day 56 maintained high diversity, comparable to the diversity before GA addition, independent of HF-impact status. This shows that the diversity differences observed after GA addition are not confounded by the bottle effect.

The overall alpha diversity found in this study's HF- microcosms was higher than the HF+ microcosms preamendment of GA. This is in agreement with the in situ study that examined these streams and other streams in the region.³⁹ After amendment of GA, HF+ microcosms maintained higher richness than HF- streams when calculating diversity with metrics that focus on unique OTUs (Observed) and the importance of rare OTUs (Chao1), whereas evenness seems to be decreasing through time as a couple of taxa dominated over time in both HF+ and HF- microcosms as seen by similar Simpson and Shannon diversity trends between the groups (**Figure 2-7**). High diversity and richness in a community after a perturbation is a sign of adaptation to chronic exposure to perturbations.⁶⁷ This shows that more unique members of the HF+ microbial community were able to tolerate and resist the biocide than HF- microbial communities.

Beta diversity was calculated using weighted UniFrac distance matrix. Data was ordinated using a Principal Coordinate Analysis (PCoA) as described. Clustering by PC 1 explains 65.4% of the variation in the microbial community, while clustering by PC 2 explains 10%. Results showed a visible clustering by days and impact status (HF+ and HF-) in the GA added microcosms by both PC1 and PC2, while the no-GA microcosm mostly clustered by PC 2 (**Figure 2-8**). Statistically significant differences were observed between HF+ and HF- microbial communities ($P < 0.01$), treatments (GA vs No Biocide with $P < 0.001$), treatments through time ($P < 0.001$), the interaction between impact status (HF+ and HF-) and treatments ($P < 0.02$), and the interaction between impact status, treatments, and days ($P < 0.03$). Results showed that the microbial community response to the biocide in these microcosms included phylogenetically distinct organisms based on previous exposure to HF activity.

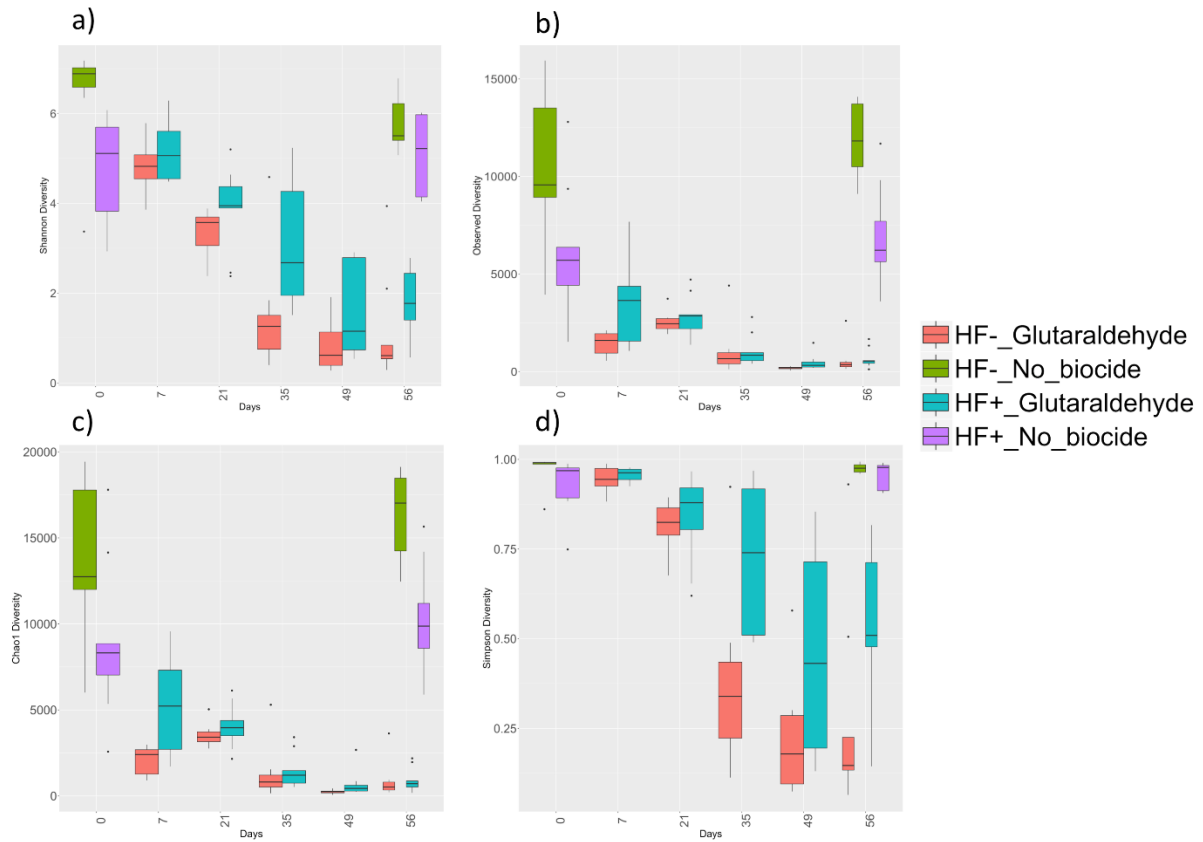


Figure 2-7: Alpha diversity measurements over time. Different richness and evenness alpha diversity estimators comparing HF+ and HF- microcosms over time; the estimators used were (a) Shannon Diversity, (b) Observed Diversity, (c) Chao1, and (d) Simpson Diversity. Red and green box plots represent HF+ glutaraldehyde (days 7 to 56) and no glutaraldehyde added (days 0 and 56 only). Blue and purple box plots represent HF- glutaraldehyde and no glutaraldehyde added. The box plots show the distribution of the data points: upper whisker to the beginning of the box is the first quartile, beginning of box to median represents the second quartile of the data, median to end of box is third quartile, and end of box to lower whisker is the fourth quartile

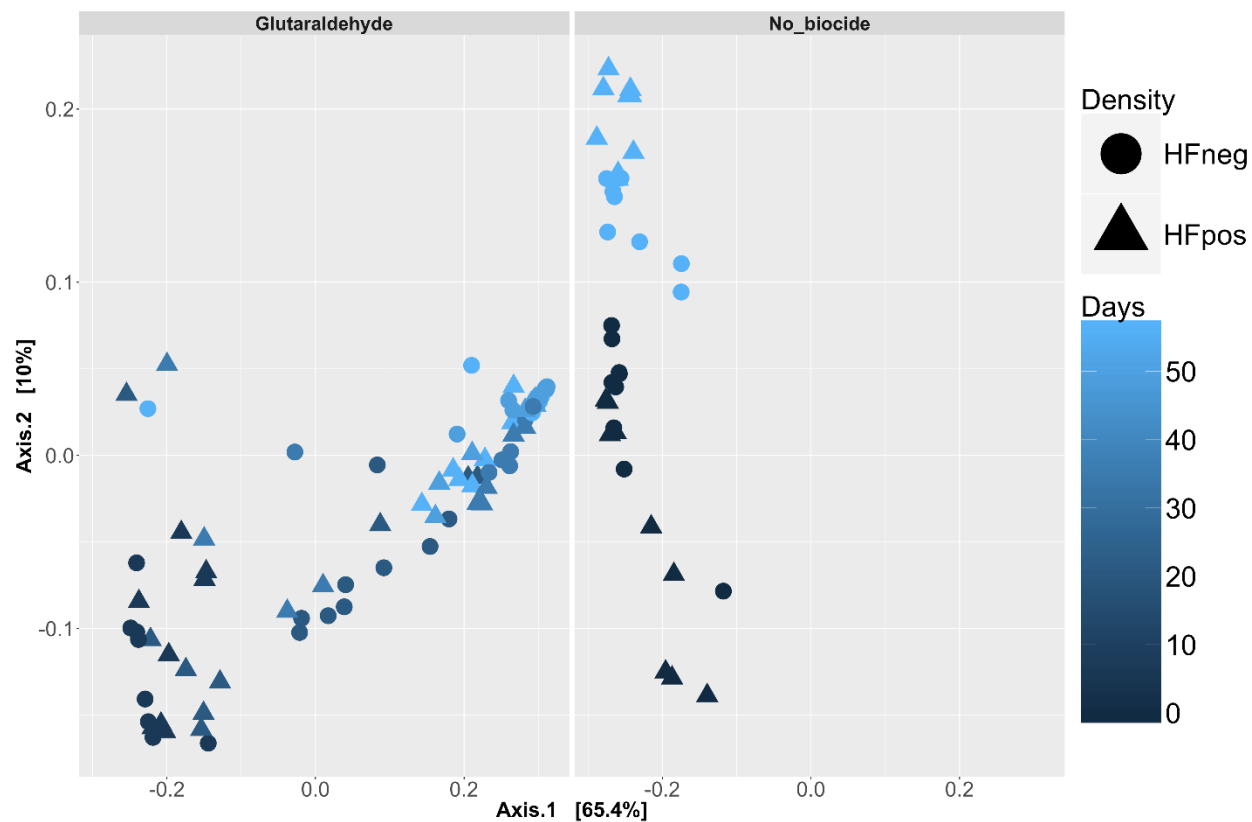


Figure 2-8: Principal coordinate analysis (PCoA) plot of phylogenetic microbial community changes over time in HF+ and HF- impacted microcosms amended or unamended with glutaraldehyde as described by weighted Unifrac beta diversity measurements. PC1 explains 65.4% of the variation while PC2 explains 10%.

Differentially enriched taxa over time in HF+ and HF- microcosms

Overall, many members of the original microbial community in HF+ and HF- microcosms were not able to tolerate GA over time as seen by a decrease in diversity (**Figure 2-7**) and by an increase in differentially abundant OTUs between day 0 and the next 4 sampling events (days 7, 21, 35, and 49). By the last sampling event, day 56, the number of differentially abundant OTUs decreases, a sign of population resilience, and/or GA reaching concentrations below inhibition level.

Specifically, 7 days after addition of GA, 239 OTUs were differentially enriched. Twenty-seven OTUs experienced a positive \log_2 FC while 212 OTUs experienced a negative \log_2 FC and, hence, were inhibited by exposure to GA. The highest \log_2 FC corresponded to an OTU identified as the genus *Myroides* (19.09 \log_2 FC), followed by OTU identified as *Robinsoniella* (18.64 \log_2 FC). Interestingly, 6 OTUs corresponding to the marine clade SAR406 were also enriched (all corresponding to Family A714017 but different or unclassified genus). However, all of these enriched OTUs were in low abundance (<2%) except for *Alcanivorax* (2.77 \log_2 FC). There were 71 differentially enriched OTUs between HF+ and HF- prior to the addition of GA. Seven days after addition, only one OTU was differentially enriched between HF+ and HF- identified as *Psychroserpens* (7.80 \log_2 FC). However, it was at low abundance (below 2%). By day 21, there were 315 OTUs differentially enriched as compared to the original pre-GA population. Eight OTUs were enriched at this time point. The only OTUs with abundance of more than 2% of the population were *Idiomarina* (4.90 \log_2 FC), *Methylobacterium* (2.78 \log_2 FC), and *Bacillus* (2.06 \log_2 FC). There were not significant differences in enrichment between HF+ and HF- that passed the stringent 2 \log_2 FC cutoff that was imposed.

By day 35, there were 407 OTUs differentially enriched as compared to the original, day 0 microbial population. These OTUs were classified as *Amphritea* (5.29 \log_2 FC), *Methylobacterium* (5.19 \log_2 FC), and *Beijerinckia* (3.23 \log_2 FC). Three OTUs were differentially enriched in HF+ vs HF- at day 35. The genus *Acinetobacter* had a 3.60 \log_2 FC in HF-, while *Beijerinckia* and *Janthinobacterium* had an 8.17 and 3.94 \log_2 FC, respectively, in HF+. By day 49, there were 419 differentially enriched OTUs as compared

to the pre-GA microbial population. Only four OTUs were positively enriched at day 49; those OTUs correspond to *Myroides* (14.00 log₂ FC), *Robinsoniella* (10.61 log₂ FC), *Methylobacterium* (6.02 log₂ FC), and *Beijerinckia* (2.93 log₂ FC). One OTU was differentially enriched in HF+ vs HF- at day 49. The genus *Beijerinckia* had an 8.97 log₂ FC in HF+ as compared to HF-.

By day 56, there were 174 differentially enriched OTUs. Of those, 66 were enriched in day 56 as compared to day 0. The ones with more than 2% abundance were *Methylobacterium* (12.19 log₂ FC), *Beijerinckia* (10.20 log₂ FC), *Mycobacterium* (7.81 log₂ FC), *Alcanivorax* (5.74 log₂ FC), *Stenotrophomonas* (5.24 log₂ FC), *Bacillus* (3.48 log₂ FC), *Idiomarina* (3.28 log₂ FC), and *Burkholderia* (3.04 log₂ FC). Only one OTU identified as the genus *Beijerinckia* (9.36 log₂ FC) was enriched in HF+ microcosms as compared to the HF-. Day 56 GA-microcosms were also compared to no-GA microcosms at day 56. There were 263 enriched OTUs, of those 44 were enriched in the GA microcosms: *Methylobacterium* (10.31 log₂ FC), *Alcanivorax* (5.81 log₂ FC), *Mycobacterium* (5.67 log₂ FC), *Beijerinckia* (5.21 log₂ FC), *Idiomarina* (4.42 log₂ FC), and *Bacillus* (3.13 log₂ FC); day 0 and day 56 no-GA microcosms were also compared to see how the community changed over time due to bottle effect. There were 209 differentially enriched OTUs. It is worth noting that *Bacillus* (-2.77 log₂ FC) and *Idiomarina* (-5.10 log₂ FC) were suppressed at day 56 no-GA as compared to day 0 and that *Myroides* (5.09 log₂ FC) experienced an enrichment.

These enrichments over time suggest which OTUs were driving the response to GA. *Alcanivorax* was a dominant first responder, and after an adaptation period, *Idiomarina*, *Methylobacterium*, and *Bacillus* responded as well. *Methylobacterium* differential enrichment continued until the end of the experiment, dominating in abundance (71% in HF+ and 84% in HF- microcosms at day 56, (**Figure 2-9**), indicating that it was able to adapt to GA presence and dominate.

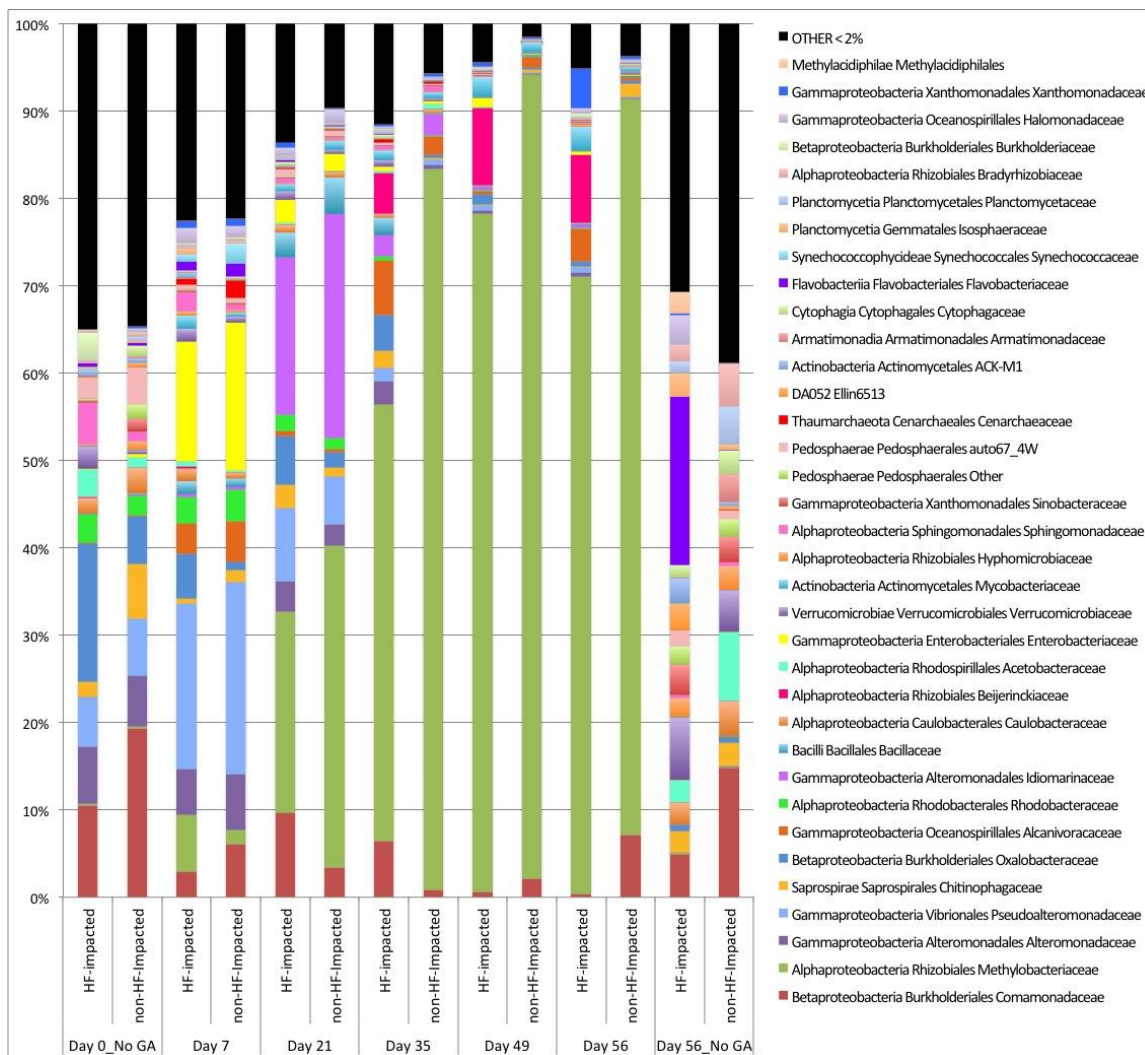


Figure 2- 9: Microbial Community Shifts Over Time. A) Phylum level shifts over 56 days with and without glutaraldehyde addition, samples were average into HF-impacted and non-HF-impacted groups.

It is worth noting that it was not enriched right after GA addition, possibly indicating that a lag period was needed for adaptation. By day 35, other than *Methylobacterium*, *Beijerinckia* is worth highlighting, as it was preferentially enriched in HF+ microcosms. The trend of *Methylobacterium* and *Beijerinckia* continued until the end of the experiment. In addition, by day 56, *Alcanivorax* and *Idiomarina* were enriched when comparing both day 56 with day 0 no-GA and with day 56-no-GA.

Studied members of the enriched genus can provide better understanding of the interactions at play. *Alcanivorax* are commonly found in hydrocarbon-impacted marine environments and have been observed to degrade alkanes and other hydrocarbons and use them as their sole carbon source,⁶⁸ and the alkane degradation pathway employs aldehyde dehydrogenases,^{69, 70} which may help this genus thrive and possibly help degrade GA. Furthermore, isolated strains of *Alcanivorax* spp. were shown to be resistant to antimicrobials by the use of efflux pumps,⁷¹ which could also facilitate tolerance for GA. *Idiomarina* is frequently detected in hydrocarbon-rich environments such as oil spills;⁶⁸ HF produce water and flowback,²³ but their role and/or mechanisms in hydrocarbon degradation is unknown. It is possible that enrichment of *Idiomarina* is also associated with the aldehyde dehydrogenases. *Alcanivorax* and *Idiomarina* are members of the *Gammaproteobacteria* class, which observed enrichment after a week of exposure to GA (**Figure 2-9**); enrichments of this class have been observed in aquatic environments after perturbations from hydrocarbon sources, sewage runoff, antimicrobials, and other anthropogenic sources.⁷² Most of the enriched *Gammaproteobacteria* families are known to be halotolerant such as *Alteromonadaceae*,⁷³ *Pseudoalteromonadaceae*,⁷⁴ *Alcanivoracaceae*,⁷⁵ *Idiomarinaceae*,⁷⁶ and *Halmonadaceae*⁷⁷ (**Figure 2-9**). Moreover, Vikram et al.²⁵ showed that genes needed for responding to osmotic stress, membrane integrity, and protein transport are up-regulated when the bacteria are exposed to HF produced and flowback water, and this up-regulation was correlated with increased bacterial tolerance to biocide exposure. Another recent study indicated that, in pathogens, GA resistance can be mediated through an increase in efflux pumps, which will increase the rate of export of the biocide.⁷⁸ It has also been reported that efflux pump encoding genes increase in downstream UOG impacted surface water, which may be a bacterial response

mechanism to stress caused by HF chemicals and high salinity.⁷⁹ This could help explain why *Gammaproteobacteria* associated with saline aquatic environments are enriched after GA addition, since the mechanisms to control osmotic stress might be a key genetic trait for microbial survival against GA.

These microcosms experiments did not explore the impacts of high salinity with respect to microbial response and degradation of GA. High salinity might affect the tolerance to GA as shown by Vikram et al.;²⁵ however, as shown by this work, higher tolerance does not translate to higher degradation. Another study showed inhibited biotic degradation of GA in a mixture with 30 000 mg/L NaCl and two other HF chemicals on agricultural top soil as compared to GA alone, while the abiotic degradation of the GA, NaCl, and HF chemicals was faster than GA alone.²⁹ Degradation of low concentration (1.5–3.0 mg/L) GA has also been shown in seawater and its native organisms.²⁸ Halotolerant microbes seem to be able to degrade GA; however, it is unclear how salt would affect degradation rates in freshwater streams in the case of a HF fluid spill containing GA and the high salt content associated with HF flowback.

The increase in *Alphaproteobacteria* (accounting for more than 90% of the microbial community in the microcosms after day 49 of GA amendment, **Figure 2-9**) as the microbial system adapted to the GA perturbation suggests that this bacterial class is better at tolerating the GA as a stressor in the long term compared to *Beta* and *Gammaproteobacteria*. *Alphaproteobacteria* are known to experience horizontal gene transfer more frequently than other *Proteobacteria*, and their extensive genomes are known to have a larger number of mobile elements.⁸⁰ This may contribute to the higher “memory effect” or adaptation detected in the HF+ aquatic microbial community with genetic material being shared between the sediment’s sessile microbial community, the epilithic bacteria from rocks, and the free-floating microbes collected for the microcosm setups.⁸¹ Moreover, *Alphaproteobacteria* are Gram-negative and, therefore, are known to be more resilient to antimicrobials because of their outer membrane, as compared to Gram-positive bacteria.⁸² The genus *Methylobacterium* was the most abundant *Alphaproteobacteria* in both HF+ and HF- streams; however, it is more dominant in HF- streams, representing 84% of the population by day 56. The

Family *Methylobacteriaceae* are commonly found in the environment growing on single carbon compounds, the microbe's sole energy source, in addition to more complex carbon compounds.⁸³ Enrichment of methylotrophs has also been observed in studies pertaining to triclosan and quaternary ammonium antimicrobials and other environmental pollutants like hydrocarbons and chlorinated compounds as these bacteria are able to cometabolize these pollutants through the production of methane monooxygenase.⁸⁴ ⁸⁵ However, *Methylobacteriaceae* response might be antimicrobial specific and dependent on oxygen availability as a study utilizing anaerobic microcosm inoculated with UOG impacted and unimpacted sediment described a significant decrease in abundance of after the addition of the biocide DBNPA.⁸⁶ Another interestingly enriched *Alphaproteobacteria* was the genus *Beijerinckia*, preferentially enriched in HF+ microcosms. These genera are members of the order *Rhizobiales*, which has similarly been detected in streams adjacent to UOG disposal facilities.⁶⁵ Isolated members of this genera have been shown to be nitrogen-fixing, nonsymbiotic, chemo-heterotrophic bacteria capable of degrading recalcitrant aromatic compounds because of their methanotrophic capabilities.⁸⁷

Overall, the microbial communities of HF+ and HF- microcosms had different phylogenetic responses to the addition of GA even though *Methylobacteriaceae* was the most dominant taxa in both. The phylogenetic differences are driven by lower abundance microbes that respond to GA based on past HF activity exposure. HF- had a more prominent negative response to GA, as seen by biomass and richness loss. This suggests HF fluids exposure causes different microbial responses and adaptation to the biocide GA.

A long list of studies have described the adaptation of microbes to chemical stressors, which they then use as energy sources or acquire the ability to cometabolize.⁸⁸ An increase in this effect has previously been observed in ecosystems that were exposed to contaminants;⁸⁹ however, adaptation did not provide a degradation advantage to GA in the HF+ microcosms. This suggests the difference in degradation rates might not be biotic alone but rather driven by abiotic–biotic interactions. HF- microcosms source water had a more neutral pH (average pH = 6.5) compared to the acidic pH of the three HF+ streams

(average pH = 4.9), and the pH was negatively correlated (Pearson = -0.83) to the concentration of GA at day 56. Thus, higher pH experienced more biodegradation of GA. Higher pH difference has the potential to affect the availability of reactive sites in the microbial cell walls surface, causing a faster biocidal effect (and a faster depletion/deactivation of GA).⁵ These factors may explain why GA decreases more rapidly over time in the HF- streams. The site with the most GA depletion by day 56 was EE (**Appendix A2**), the HF- stream with the highest pH (pH = 7.3).

However, the microcosms did not maintain constant pH over the incubation period; independent of HF impact status, source water location, or biotic or abiotic conditions, all of the microcosms pH increased over time (**Appendix A3**). While GA is more stable at lower pH, its bactericidal properties are impaired in acidic environments where there are fewer available active sites on the cell wall. This effect of pH will require more future studies, but it is still our hypothesis that it is affecting GA degradation in a number of ways.

As explained above, the microbial community from HF+ microcosms was shown to better tolerate the biocide, however it did not degrade the biocide faster than HF- microcosms. Therefore, a further study of the biodegradation mechanisms associated with varying pH is needed to better understand the nuances of the abiotic-biotic interactions driving GA biodegradation.

Environmental Implications of this study

This study shows that there are long lasting effects in streams impacted by HF, which need to be considered for environmental impact assessment and bioremediation strategies. Abiotic factors such as acidified pH can affect the microbial community's ability to respond to a second or continuous exposure to HF waste, causing HF chemicals to be more persistent in the environment than expected. As HF practices keep expanding worldwide, this knowledge can help bioremediation efforts to optimize natural attenuation.

**CHAPTER THREE: EXPLORING ENDOGENOUS EXPRESSION OF N-
ACYL AMINO ACIDS**

Abstract

Adaptive thermogenesis has re-emerged as an important component of energy expenditure. *N*-acyl amino acids (*N*-AAAs) can contribute to adaptive thermogenesis by uncoupling mitochondrial respiration. Hypothesizing that lipolysis promotes the formation of *N*-AAAs, the connection was studied in an avian model (chickens) because lipolysis is readily induced by acute (~ 4 hr) food withdrawal. An avian model also allowed evaluation of *N*-AAA production in the absence of brown adipose tissue, which was reported to contribute the enzyme (peptidase M20 domain containing 1 (*PM20D1*)) that catalyzes their synthesis. Serum samples were collected from 21 d broiler chicks that were fed *ad libitum*, fasted for 4, 8, or 12 h, or fasted for 12 h and then re-fed for 2 h (n=10/group). Samples were profiled for levels of 27 *N*-AAA species using a novel UPLC-HRMS method. In short, a single Ultra-Performance Liquid Chromatography—High Resolution Mass Spectrometry (UPLC-HRMS) method was developed and validated with 27 *N*-acyl amino acid standards and 20 bile acid standards. (Given their shared relevance for obesity research and their similar physio-chemical properties, a single analytical method for the simultaneous detection of these metabolites is desirable). Reversed-phase chromatography and Orbitrap mass spectrometry were used in tandem to detect the 47 metabolites in a 20.5-minute method. The standards were detected with a high degree of mass accuracy (less than 2 ppm for all *N*-AAAs and less than 7 ppm for all bile acids). In addition, protocols for the extraction of these metabolites from biological media were tested.

Fasting progressively increased the serum pool of *N*-AAAs ($p=1.8 \cdot 10^{-6}$), in parallel with increased non-esterified fatty acid (NEFA) levels ($p=0.0003$). Refeeding for 2 h rapidly restored the *N*-AAA pool to fed levels ($p=0.99$; fed vs. 12 h fast-refed). Percent abundance of some species (e.g., oleoyl-l-valine; $p=2.8 \cdot 10^{-4}$) increased with fasting while others (e.g., stearoyl-l-leucine; $p=1.9 \cdot 10^{-5}$) declined, suggesting molecule-specific regulation of *N*-AAA synthesis. Expression of *PM20D1* in adipose tissue (subcutaneous and abdominal) and in liver was not affected by fasting. Effect of diet on *N*-AAA expression was also evaluated using the same boiler chick model. Four unique diets, containing a supplement of flaxseed, canola, fish, or animal (lard) oil were administered to chicks, which were then sacrificed for analysis of *N*-AAA in tissue. The fatty acid composition of each diet was

directly reflected in the tail composition of each *N*-AAA profile. This suggests that substrate availability may be a prominent means of *N*-AAA regulation.

Preface

The following chapter is adapted from two manuscripts: a methods paper by Gibson et. al. submitted to *Metabolomics* and an applied biological paper by Gibson et. al. in final preparation for submission to the *American Journal of Physiology-Endocrinology and Metabolism*.

Contributions for methods paper

CMG and Eric D Tague (EDT) developed the methods and analyzed samples. Sheev M. Zaver (SMZ) synthesized standards and performed extractions. CMG characterized synthetic standards. Brittni M. Woodall (BMW) and SMZ assisted with method validation. Emmanuelle T. Torchon (ETT), Brynn H Voy (BHV), and Hans-Rudolf Berthoud (HRB) provided biological samples. BHV and HRB prompted method creation and biological rationale. CMG, EDT, Shawn R. Campagna (SRC), and BHV wrote the manuscript.

Contributions for biological paper

CMG and SMZ ran samples. ETT and BHV conducted the animal studies and provided samples. SMZ extracted and prepared samples. BHV conceptualized the experiments. CMG and BHV performed statistical analyses and interpreted the data. CMG, BHV, and SRC wrote the manuscript.

Forward

As previously discussed, small molecule mitochondrial uncouplers possess therapeutic potential for the treatment of obesity and related metabolic disease. Exogenous uncouplers like DNP have been historically applied in therapeutic roles with mixed response, due to their potential side-effects. In contrast to these foreign uncouplers, *N*-acyl amino acids are endogenously produced metabolites that induce thermogenesis, promote glucose homeostasis, and increase energy expenditure. Long and coworkers captured the attention of health professionals when they published an initial report

describing the uncoupling abilities of *N*-AAAs in 2016.¹¹ Their findings prompted a letter published in *Science* magazine, enthusiastically entitled “The Quest to Burn Fat, Effortlessly and Safely”.⁸ Adaptive thermogenesis, regulated endogenously, represents a rich area of research for both basic and applied science, and *N*-AAAs are at the forefront of the discussion.

Since these molecules are therapeutically promising and conserved across many animal species, research opportunities are abundant. Efforts to fully ascertain the natural biological roles of *N*-AAAs are essential. To this end, a method was developed that allowed the analysis of *N*-AAAs by UPLC-HRMS. Given the related biological relevancies, efforts to profile *N*-AAAs were combined with concurrent endeavors to measure bile acids. Successful method development yielded a single general method for detecting both classes of molecules from biological matrices with the desired resolution. While orthogonal efforts to study bile acid expression have been undertaken, the following work is focused the measurement and biological roles of *N*-AAAs.

Using broiler chicks as a model system for studying adiposity and uncoupling, animal experiments were conducted with the aim of elucidating the endogenous expression of *N*-AAAs. A group of chicks was subjected to incremental fasting to acutely induce lipolysis. A parallel experiment involved dietary supplementation with various fatty acids. In both cases, samples were collected from the chicks and analyzed *via* UPLC-HRMS to determine changes in *N*-AAA profile.

All UPLC-HRMS data was interpreted using statistical tools. Notable among these is partial least squares discriminant analysis (PLS-DA), a multivariate analysis algorithm which can be applied to classification problems to accomplish descriptive modeling.⁹⁰⁻⁹² PLS-DA reduces data dimensionality to assign components and constructs a predictive model. As a result, it allows the variance across data groups to be visualized and identifies the variables responsible for the differences. In addition, post hoc statistical power analysis was employed to illustrate that the pair-wise differences between means were significant for the given number of samples. The power level denotes the probability of correctly assigning real differences. This is used in conjunction with the student’s *t*-test,

which communicates the probability of inadvertently assigning false differences. Lastly, analysis of variance (ANOVA) is used as a way to generalize the t-test and compare means across three or more groups.

Chapter 3A: A single UPLC-HRMS method for the profiling of bile acids and *N*-acyl amino acids, two classes of obesity-mediating metabolites

Introduction

Obesity is a growing epidemic. From 2011 to 2014, the prevalence of obesity in the United States was 36% in adults and 17% in children.⁹³ As an underlying cause of diabetes, cardiovascular disease, cancer, hypertension, and many other widespread medical conditions, obesity is a common culprit for morbidity and mortality.⁹⁴ In addition to the direct health effects of obesity, the indirect societal effects include an impending inundation of health care resources.⁹⁴ The epidemic has inspired focused research initiatives to increase fundamental scientific knowledge and investigate potential therapies for combating obesity and related illnesses. In recent years, *N*-acyl amino acids and bile acids, two classes of small, amphipathic biomolecules have emerged as possessing potential weight-regulating activities. Further investigation of these families of molecules may inform the global fight against obesity.

N-acyl amino acids are condensation products of fatty acids and amino acids. Sometimes designated as elmiric acids, these compounds are endogenously produced and have been shown to possess various physiologic and pharmacological effects.⁹⁵ To date, approximately 70 of these naturally occurring elmiric acids have been identified.⁹⁵ Initially investigated for their potential analgesic and anti-inflammatory properties, *N*-acyl amino acids have been implicated in vasodilation, cell migration, and selective inhibition of cancer cell proliferation.⁹⁵⁻¹⁰⁰ More recently, a class of *N*-acyl amino acids featuring lipophilic amino acids has been demonstrated in brown adipocytes of mice to bind directly to mitochondria and uncouple cellular respiration.¹¹ Administration of these endogenously produced lipoamino acids improved glucose homeostasis and promoted energy expenditure.¹¹ The enzyme implicated as being responsible for catabolism and anabolism of these *N*-acyl amino acids (PM20D1) is secreted by adipocytes in fat tissue.¹¹ Due to

their role in activating mitochondria for thermogenesis, these compounds attract immense attention as promising candidates for the safe combat of obesity.^{8, 11} Predictably, as endogenously-produced uncouplers of respiration, they are more attractive therapeutic candidates than foreign chemical uncouplers. Despite the promise they possess, the endogenous expression of these thermogenesis-promoting *N*-acyl amino acids has only begun to be studied. These compounds are an interesting class of potential biomarkers for investigating molecular variations across various dietary and metabolic states such as fasting and lipolysis.

Bile acids are cholesterol-derived steroid acids found predominantly in the bile of vertebrates.¹⁰¹ A structurally diverse group of steroids, bile acids exist in various forms that are related through reduction/oxidation, esterification, and amidation.¹⁰¹ They occur naturally in both free forms (such as cholic acid) and conjugated forms which have been condensed with amines (primarily taurine).¹⁰¹ Collectively, bile acids aid in the digestion of dietary fats and absorption of lipids. This is due to an distinctive ability to emulsify products of lipolysis by forming solubilizing micelles around free fatty acids, monoglycerides, and cholesterol.^{102, 103} A growing body of evidence indicates that specific bile acids also regulate metabolism by acting as ligands for farnesoid X receptor (FXR) and takeda G receptor 5 (TGR5).^{104, 105} The farnesoid X receptor is a nuclear receptor that transcriptionally controls lipid metabolism by forming a heterodimer with retinoic X receptor α , and the takeda G receptor 5 exerts its metabolic effects by activating protein kinase A upon binding of bile acids at the cell membrane. These metabolic actions of bile acids may underlie some of the benefits of Roux-en-Y gastric bypass surgery, which increases levels of circulating bile acids. Independent studies have demonstrated that bile acids may help mediate type-II diabetes and promote weight loss after gastric bypass.¹⁰⁶⁻¹⁰⁸ In addition, bile diversion independent from surgical rearrangement of the GI tract has been shown to initiate weight loss and improve homeostasis for glucose and lipids.¹⁰⁹ The benefits of bile diversion are comparable to those of bariatric surgery and are accompanied by elevated levels of bile acids, which may be responsible for the observed effects.¹⁰⁹ Most recently, bile acids have been identified among the most important biomarkers for achieving a stable metabolic state after bariatric surgery.¹¹⁰ Together, this data suggests that bile acids are also important molecular facilitators of weight regulation.

Because of the pharmacological significance of bile acids and *N*-acyl amino acids in lipolysis, glucose homeostasis, thermogenesis, and weight control, methods for the detection and quantitation of these biomolecules are desired. Since their roles in metabolism may be related and given that these two classes of compounds occupy a similar mass range, possess similarly amphiphilic properties, and share a preference for negative-mode ionization, it is plausible that they may be simultaneously extracted from biological samples and analyzed using a single Ultra-Performance Liquid Chromatography—High Resolution Mass Spectrometry (UPLC-HRMS) method.

The research interest in these classes of molecules is emerging, so relatively few methods for their detection have been reported. To our knowledge, no method for the simultaneous detection of both classes has been reported. The LC-MS detection of several endogenous *N*-acyl amino acids was reported in 2010 using a targeted lipidomics method.¹¹¹ Another targeted approach using triple quadrupole mass spectrometry was recently used to profile endogenous elmiric acids in serum of mice.¹¹ Bile acids have been measured via LC-MS using an array of methods which generally employ tandem MS in targeted approaches.¹¹²⁻¹¹⁸ At least one of these MS/MS methods used select deuterated bile acids as internal standards for enhanced quantitation.¹¹⁵ Compared to other LC-MS techniques, the increased performance offered by UPLC-HRMS provides increased resolution both in the time domain and the mass domain. In contrast to the targeted methods, an untargeted approach allows information to be collected across all *m/z*'s instead of selecting *m/z*'s of interest. This is beneficial because the data can be reanalyzed later, and more compounds can be identified as future biological investigations are undertaken. Herein, a UPLC-HRMS method designed for the untargeted profiling of both bile acids and *N*-acyl amino acids is described, validated using synthetic and commercial standards, and applied to biologically relevant sample types.

Experimental procedures

Chemicals

N-dodecanoyl valine (DDV), *N*-dodecanoyl leucine (DDL), *N*-dodecanoyl phenylalanine (DDF), *N*-palmitoleoyl valine (PMOV), *N*-palmitoyl valine (PMV), *N*-palmitoleoyl leucine

(PMOL), *N*-palmitoyl leucine (PML), *N*-linoleoyl valine (LINV), *N*-oleoyl valine (OLV), *N*-stearoyl valine (STV), *N*-linoleoyl leucine (LINL), *N*-oleoyl leucine (OLL), *N*-stearoyl leucine (STL), *N*-eicosapentaenoyl valine (EPAV), *N*-palmitoleoyl phenylalanine (PMOF), *N*-palmitoyl phenylalanine (PMF), *N*-eicosanoyl valine (EIV), *N*-eicosapentaenoyl leucine (EPAL), *N*-eicosanoyl leucine (EIL), *N*-docosahexaenoyl valine (DHAV), *N*-linoleoyl phenylalanine (LINF), *N*-oleoyl phenylalanine (OLF), *N*-stearoyl phenylalanine (STF), *N*-docosahexaenoyl leucine (DHAL), *N*-eicosapentaenoyl phenylalanine (EPAF), *N*-eicosanoyl phenylalanine (EIF), and *N*-docosahexaenoyl phenylalanine (DHAF) were synthesized in-house using a method previously reported.¹¹ Characterization data for all the standards included in **Appendix B**. In short, fatty acids were activated by conversion to acyl chlorides using oxalyl chloride (1.2 eq.), concentrated, and coupled to amino acids (1.5 eq.) in a one-pot, two-step sequence (**Figure 3-1**). Synthetic lipoamino acids were purified via flash chromatography using 1:9 tetrahydrofuran in hexanes before UPLC-HRMS analysis.

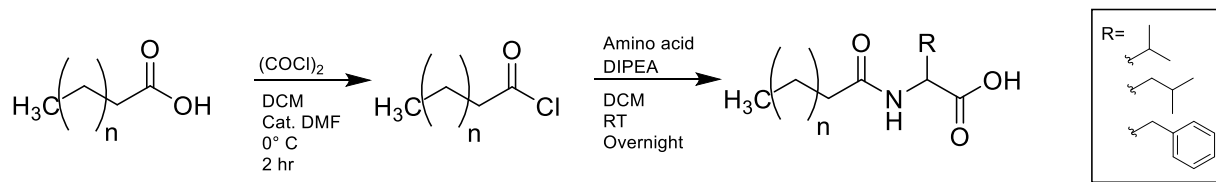


Figure 3-1: Synthetic scheme for *N*-acyl amino acid standards

Bile acid standards were purchased from various sources, predominantly as salts: Cholic acid (CA) and deoxycholic acid (DCA) were purchased from MP Biomedicals (Santa Ana, California). Taurocholic acid (TCA), taurodeoxycholic acid (TDCA), and taurochenodeoxycholic (TCDCA) acid were purchased from Millipore Sigma (Darmstadt, Germany). Chenodeoxycholic acid (CDCA), hyodeoxycholic acid (HDCA), ursodeoxycholic acid (UDCA), α -muricholic acid (α MCA), β -muricholic acid (β MCA), glycolithocholic acid (GLCA), glycocholate (GCA), glycochenodeoxycholate (GCDA), glycodeoxycholate (GDCA), tauroolithocholate (TLCA), taurohyodeoxycholate (THDCA),

tauroursodoexycholate (TUDCA), tauro- α -muricholate (TaMA), tauro- ω -muricholate (TwMA), and tauro- β -muricholate (TbMA) were purchased from Steraloids (Newport, RI). Deuterated bile acid standards cholic acid-d4 (dCA), glycochenodeoxycholic acid-d4 (dGCDC), chenodeoxycholic acid-d4 (dCDCA), glycocholic acid-d4 (dGCA), and taurocholic acid-d4 (dTCA) were purchased from Cambridge Isotope Laboratories (Tewksbury, MA).

Formic acid, HPLC grade water, and HPLC grade acetonitrile were purchased from Thermo Fisher Scientific (Waltham, MA).

UPLC–HRMS method

For each sample, a 10 μ L injection volume was subjected to UPLC separation on a Dionex UltiMate 3000 (Thermo Fisher Scientific, Waltham, MA). A Synergi 2.5 μ m Hydro-RP 100 Å, 100 \times 2 mm column (Phenomenex, Torrance, CA) was used at 40°C in a temperature-controlled compartment for the duration of the run. The mobile phase was comprised of a two-solvent system: Solvent A consisted of 0.1% formic acid in water, while solvent B was 0.1% formic acid in acetonitrile. Gradient: $t = 0$ min, 100% solvent A, 0% solvent B; $t = 13$ min, 0% solvent A, 100% solvent B; $t = 14$ min, 0% solvent A, 100% solvent B; $t = 14.5$ min, 100% solvent A, 0% solvent B; $t = 20.5$ min, 100% solvent A, 0% solvent B. The flow rate was a constant 0.300 mL per minute. The chromatograph was coupled to an Orbitrap Exactive Plus mass spectrometer (Thermo Fisher Scientific, Waltham, MA) equipped with a heated electrospray ionization (HESI) source. Negative-mode HESI and mass-spectrometric analysis were performed according to the following parameters: sheath gas flow of 25 units, aux gas flow of 8 units, capillary temperature of 300 °C, aux gas heater temperature of 150 °C, spray voltage of 4.2 kV, ACG target of 3×10^6 , resolution of 140 000, and a scan range of 150–1000 m/z .

Data analysis

Post-processing was conducted by converting *.raw* files to *.mzml* and integrating Extracted-Ion Chromatograms (EICs) using Metabolomic Analysis and Visualization Engine (MAVEN), an open-source software for processing LC-MS data.¹¹⁹⁻¹²¹ Partial

Least Squares Differential Analysis (PLS-DA) was performed using a series of R packages in RStudio version 1.1.423.¹²²⁻¹²⁴ Box and whisker plots were also generated with R (version 3.4.3).

Differentiation of isobaric N-acyl amino acids using parallel reaction monitoring

To further distinguish the pairs of isobaric elmeric acids, tandem MS was used to fragment the parent molecular ions. Standards were analyzed using identical instrumental parameters on a hybrid Orbitrap Q Exactive (Thermo Fisher Scientific, Waltham, MA) using parallel reaction monitoring (PRM). The parent molecular ions of 400.2857 and 426.3014 m/z were selected using the quadrupole mass filter with an isolation window of $\pm 0.4 m/z$ and fragmented using a collision energy of 35 eV. Product ions were detected, and their chromatograms were integrated using Xcalibur Quan Browser (Thermo Fisher Scientific, Waltham, MA) for comparison and quantitation.

Sample preparation

Bile acid standards were prepared in 1:1 water:ethanol prior to the creation of a combined stock solution. Four stock solutions were used to individually analyze BAs with identical m/z to determine if isomers were separable via chromatography. Calibration levels were as follows (in μM): 0.001, 0.005, 0.025, 0.050, 0.100, 0.250, 1.0, 2.5, 10, 20, and 100. A mixture of internal standards was spiked into all calibration standards and biological samples at a final concentration of 50 nM. *N*-acyl amino acid standards were suspended in 10% chloroform in methanol for serial dilution and calibration. *N*-acyl amino acid standards were prepared from stock solutions containing mixtures compounds at the following concentrations (μM): 0.001, 0.005, 0.010, 0.050, 0.100, 0.500, 1.0, 5.0, 10.0, 50, and 100.

Given the amphiphilic properties of both bile acids and *N*-acyl amino acids, some liberty with regards to extraction from biological matrices can be available. However, certain extraction protocols increase sensitivity and yield maximized ion intensities and peak areas. *N*-acyl amino acids and bile acids were extracted from test matrices using two previously-reported protocols: a) a 40:40:20 methanol/water/acetonitrile mixture with 0.1

M formic acid¹²⁵ and b) a biphasic solvent mixture containing 0.1 N aqueous HCl/MeOH/chloroform in a 1:1:1 ratio¹²⁶. In the former procedure, samples were resuspended in 300 μ L Milli-Q water after extraction and concentration. In the latter, all samples were resuspended in 300 μ L 10% chloroform in MeOH. Both methods were able to extract *N*-acyl amino acids and bile acids, but the latter method yielded increased intensities for *N*-acyl amino acids. Biological samples used for bile acid analysis were extracted using the acidic acetonitrile extraction.

Exact quantitation of bile acids

Calibration plots were created by integrating the areas under the curves for all external standards and internal standards. Not all external standards had a matched internal standard due to the limited commercial supply of isotopically labeled bile acids. Thus, internal standards were chosen to match the compound closest in *m/z* and chromatographic retention. The pairings of deuterated standards with analytes are provided in the Supporting Information. The log of the ratios for the external standard's area were compared to the internal standard's area ($\log[A_{\text{ex}}/A_{\text{is}}]$) and then plotted against the log of ratios for the external standard's concentration compared to the internal standard's concentration ($\log[C_{\text{ex}}/C_{\text{is}}]$). Linear regression analysis was used to interpolate area of standards from biological data.

Animal methods

Broiler chicks were used as a model system for detection of *N*-acyl amino acids from biological matrices. From seven to thirty days of age, chicks were fed diets in which fatty acids were provided from flaxseed oil, canola oil, fish oil, or lard, each of which has a distinctive fatty acid profile.¹²⁷ At 30d of age, abdominal adipose tissue was harvested and snap-frozen from five chicks in each group. Approximately 100 mg samples of adipose tissue from each of the chicks was homogenized and subjected to a previously described extraction protocol.¹²⁶ Samples were analyzed via UPLC-HRMS on the Exactive Plus (Thermo Fisher Scientific, Waltham, MA). Bile acids are most abundant in the intestinal tract of mammals, making this the ideal sample type to validate the quantitation of multiple bile acids species. A set of mice (*n*= 15) were fed a diet containing

10% fat. The duodenums from mice were removed and bile acids extracted using the acidic acetonitrile method.¹²⁵ Samples were analyzed as previously described. All tissue extracts were spiked with deuterated standard mixture at a final concentration of 50nM for quantitation using isotope dilution.

Results and discussion

Separation

The chromatographic conditions employed by the described method allow for optimized analysis of the 27 *N*-acyl amino acids by providing well-resolved, discrete peaks with preferred gaussian shapes (**Table 3-1**). Of the 20 bile acid standards analyzed, 17 spectral features were detected (**Table 3-1**). Many of the bile acids investigated are simple isomers either differing by the location of a single hydroxyl or a change in stereochemistry. The chromatographic conditions facilitate the complete resolution of several isobaric compounds, which is demonstrated by the three distinct spectral features at 407.2803 *m/z*. CA, aMCA, and bMCA all have the same *m/z*, thus separation in the mass domain was not possible. However, as seen in **Figure 3-2**, the standards are clearly resolved from each other in the time domain. As expected, CA should be retained differently than the muricholic acids (α and β) because the hydroxyl changes from the 6th position on the B ring to the 12th position on the C ring of the steroid structure (**Figure 3-2**). Surprisingly, the two diastereomers aMCA, and bMCA, have the same *m/z* as CA and are well resolved with retention times of 8.2 min and 8.4 min respectively. There are additional isomers with an *m/z* of 407.2803 that were not analyzed, but this separation scheme shows great potential in resolving those isomers as well. Conversely, some isomers were unable to be separated within the analysis, leading to joint identifications seen in Table 1. An example of this can be seen when comparing CDCA (hydroxyl at 7 position), DCA (hydroxyl at the 12 position), and HDCA (hydroxyl at the 6 position). Even though they are all constitutional isomers, CDCA and DCA have identical retention times at 10.3 minutes while HDCA has a retention time of 9.2 minutes (data not shown). Furthermore, the separation provided by reverse-phase UPLC reduces the risk of suboptimal detection of the elmiric acids and bile acids by resolving them from the polar analytes in the sample matrix, which could contribute to ion suppression.

Table 3-1: Detection parameters for metabolite standards

Compound	Formula	RT	Species	Calculated m/z	Observed m/z	$\Delta m/z$ (ppm)	Linear Range (μM)
N-acyl Amino Acids							
N-dodecanoyl valine	C ₁₇ H ₃₃ NO ₃	11.3	[M-H] ⁻	298.2388	298.2385	1.01	0.005-50.0
N-dodecanoyl leucine	C ₁₈ H ₃₅ NO ₃	11.6	[M-H] ⁻	312.2544	312.2545	0.32	0.005-50.0
N-dodecanoyl phenylalanine	C ₂₁ H ₃₃ NO ₃	11.7	[M-H] ⁻	346.2388	346.2385	0.87	0.005-50.0
N-palmitoleoyl valine	C ₂₁ H ₃₉ NO ₃	12.7	[M-H] ⁻	352.2857	352.2857	0.00	0.005-50.0
N-palmitoyl valine	C ₂₁ H ₄₁ NO ₃	13.6	[M-H] ⁻	354.3014	354.3014	0.00	0.005-50.0
N-palmitoleoyl leucine	C ₂₂ H ₄₁ NO ₃	13.0	[M-H] ⁻	366.3014	366.3014	0.00	0.005-50.0
N-palmitoyl leucine	C ₂₂ H ₄₃ NO ₃	13.8	[M-H] ⁻	368.3170	368.3174	1.09	0.005-50.0
N-linoleoyl valine	C ₂₃ H ₄₁ NO ₃	12.9	[M-H] ⁻	378.3014	378.3015	0.26	0.005-50.0
N-oleoyl valine	C ₂₃ H ₄₃ NO ₃	13.7	[M-H] ⁻	380.3170	380.3171	0.26	0.005-50.0
N-stearoyl valine	C ₂₃ H ₄₅ NO ₃	14.6	[M-H] ⁻	382.3327	382.3327	0.00	0.005-50.0
N-linoleoyl leucine	C ₂₄ H ₄₃ NO ₃	13.3	[M-H] ⁻	392.3170	392.3168	0.51	0.005-50.0
N-oleoyl leucine	C ₂₄ H ₄₅ NO ₃	13.9	[M-H] ⁻	394.3327	394.3322	1.27	0.005-50.0
N-stearoyl leucine	C ₂₄ H ₄₇ NO ₃	14.8	[M-H] ⁻	396.3483	396.3484	0.25	0.005-50.0
N-eicosapentaenoyl valine	C ₂₅ H ₃₉ NO ₃	12.3	[M-H] ⁻	400.2857	400.2852	1.25	0.005-50.0
N-palmitoleoyl phenylalanine	C ₂₅ H ₃₉ NO ₃	12.9	[M-H] ⁻	400.2857	400.2857	0.00	0.005-50.0
N-palmitoyl phenylalanine	C ₂₅ H ₄₁ NO ₃	13.8	[M-H] ⁻	402.3014	402.3014	0.00	0.005-50.0
N-eicosanoyl valine	C ₂₅ H ₄₉ NO ₃	15.6	[M-H] ⁻	410.3640	410.3641	0.24	0.005-50.0
N-eicosapentaenoyl leucine	C ₂₆ H ₄₁ NO ₃	12.6	[M-H] ⁻	414.3014	414.3012	0.48	0.005-50.0
N-eicosanoyl leucine	C ₂₆ H ₅₁ NO ₃	16.1	[M-H] ⁻	424.3796	424.3798	0.47	0.005-50.0
N-docosahexaenoyl valine	C ₂₇ H ₄₁ NO ₃	12.7	[M-H] ⁻	426.3014	426.3019	1.17	0.010-50.0
N-linoleoyl phenylalanine	C ₂₇ H ₄₁ NO ₃	13.2	[M-H] ⁻	426.3014	426.3013	0.23	0.005-50.0
N-oleoyl phenylalanine	C ₂₇ H ₄₃ NO ₃	13.9	[M-H] ⁻	428.3170	428.3164	1.40	0.005-50.0

Table 3-1 continued

Compound	Formula	RT	Species	Calculated <i>m/z</i>	Observed <i>m/z</i>	$\Delta m/z$ (ppm)	Linear Range (μM)
<i>N</i> -stearoyl phenylalanine	C ₂₇ H ₄₅ NO ₃	14.7	[M-H] ⁻	430.3327	430.3321	1.39	0.005-50.0
<i>N</i> -docosahexaenoyl leucine	C ₂₈ H ₄₃ NO ₃	13.0	[M-H] ⁻	440.3170	440.3170	0.00	0.005-50.0
<i>N</i> -eicosapentaenoyl phenylalanine	C ₂₉ H ₃₉ NO ₃	12.6	[M-H] ⁻	448.2857	448.2854	0.67	0.005-50.0
<i>N</i> -eicosanoyl phenylalanine	C ₂₉ H ₄₉ NO ₃	15.8	[M-H] ⁻	458.3640	458.3636	0.87	0.005-50.0
<i>N</i> -docosahexaenoyl phenylalanine	C ₃₁ H ₄₁ NO ₃	13.0	[M-H] ⁻	474.3014	474.3012	0.42	0.005-50.0
Bile acids							
Chenodeoxycholic/ Deoxycholic acid	C ₂₄ H ₄₀ O ₄	10.3	[M-H] ⁻	391.2854	391.2881	6.90	0.250-100
Hyodeoxycholic acid	C ₂₄ H ₄₀ O ₄	9.2	[M-H] ⁻	391.2854	391.2879	6.39	0.100-100
Ursodeoxycholic acid	C ₂₄ H ₄₀ O ₄	9.0	[M-H] ⁻	391.2854	391.2879	6.39	0.025-100
α -Muricholic acid	C ₂₄ H ₄₀ O ₅	8.2	[M-H] ⁻	407.2803	407.2828	6.14	0.250-100
β -Muricholic acid	C ₂₄ H ₄₀ O ₅	8.4	[M-H] ⁻	407.2803	407.2831	6.87	0.250-100
Cholic acid	C ₂₄ H ₄₀ O ₅	8.8	[M-H] ⁻	407.2803	407.2826	5.65	0.025-100
Glycolithocholic acid	C ₂₆ H ₄₃ NO ₄	10.4	[M-H] ⁻	432.3119	432.3122	0.69	0.001-100
Glycochenodeoxycholic acid	C ₂₆ H ₄₃ NO ₅	9.0	[M-H] ⁻	448.3068	448.3072	0.89	0.001-100
Glycodeoxycholic acid	C ₂₆ H ₄₃ NO ₅	9.1	[M-H] ⁻	448.3068	448.3075	1.56	0.001-100
Glycocholic acid	C ₂₆ H ₄₃ NO ₆	8.0	[M-H] ⁻	464.3018	464.3017	0.22	0.025-100
Taurolithocholic acid	C ₂₆ H ₄₅ NO ₅ S	13.4	[M-H] ⁻	482.2946	482.2952	1.24	0.025-100
Tauroursdoxycholic acid	C ₂₆ H ₄₅ NO ₆ S	8.7	[M-H] ⁻	498.2895	498.2893	0.40	0.025-100
Taurohyodeoxycholic acid	C ₂₆ H ₄₅ NO ₆ S	8.9	[M-H] ⁻	498.2895	498.2898	0.60	0.005-100
Taurochenodeoxycholic/ Taurodeoxycholic acid	C ₂₆ H ₄₅ NO ₆ S	10.7	[M-H] ⁻	498.2895	498.2893	0.40	0.005-100
Taurocholic acid	C ₂₆ H ₄₅ NO ₇ S	8.9	[M-H] ⁻	514.2844	514.2844	0.00	0.001-100
Tauro- α/β -muricholic acid	C ₂₆ H ₄₅ NO ₇ S	7.9	[M-H] ⁻	514.2844	514.2845	0.19	0.005-100
Tauro- ω -muricholic acid	C ₂₆ H ₄₅ NO ₇ S	8.3	[M-H] ⁻	514.2844	514.2845	0.19	0.025-100

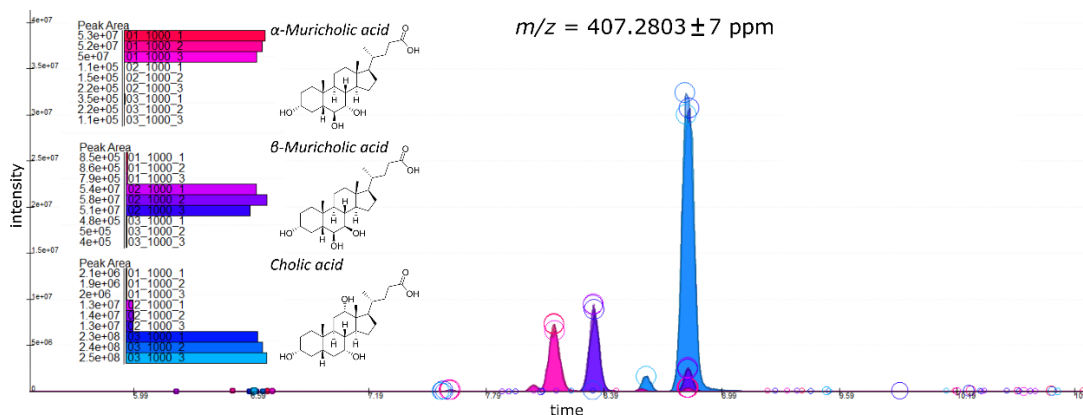


Figure 3-2: Chromatographic resolution of bile acid isomers

Detection

All *N*-acyl amino acids were detected as their respective deprotonated molecular ions $[M-H]^-$ with a mass error of less than 2 ppm. However, spectral peaks with exact m/z correlating to noncovalent dimers $[2M-H]^-$ were also detected for several elmiric acids at identical retention times to the parent molecular ion (**Figure 3-3**).

Bile acids were all detected as $[M-H]^-$ species. The mass accuracy facilitated by Orbitrap mass spectrometry allowed mass errors of less than 5 ppm for most standards. CA, aMCA, bMCA, UDCA, HDCA, and CDCA/DCA had mass errors between 5 and 7 ppm. Calibration curves were generated for all 44 standards to validate a linear relationship between area and concentration (shown in Supporting Information). All standards displayed good linear fits with average slope equal to 0.86, and average coefficients of determination equal to 0.97. Detection information is compiled in Figure 1. The lower limit of detection for the bile acids ranged from 1 nM (the lowest standard analyzed) for many of the glycine and taurine conjugated bile acids to 250 nM for many of the aliphatic bile acids. It is assumed that compounds with structural similarities to GCDCA are more easily ionized via ESI, allowing for the detection at lower concentrations. *N*-acyl amino acids displayed linear ranges down to 5.0 nM, except DHAV, which was not detected at concentrations below 10 nM.

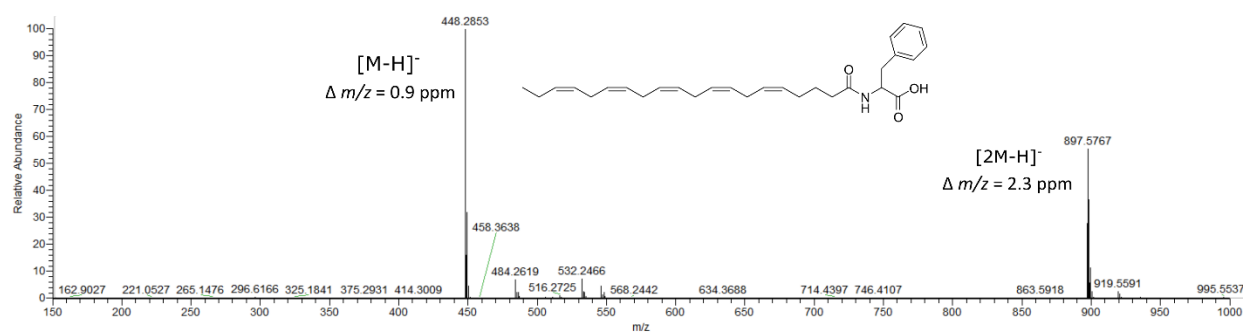
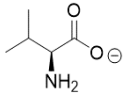
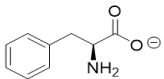
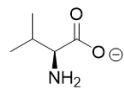
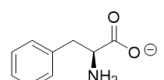


Figure 3-3: Representative illustration of noncovalent *N*-acyl amino acid dimers detected by HRMS. Shown: [M-H]⁻ and [2M-H]⁻ ions for the eicosapentaenoyl phenylalanine standard

Two pairs of lipoamino acids (EPAV and PMOF as well as DHAV and LINF) were isobaric; and, though the standards had slightly differing retention times, resolution is potentially convoluted in biological samples where matrix effects cause subtle changes in chromatography. Although chromatographic resolution of both pairs was achieved in the present work, tandem MS can provide an additional level of resolution if necessary. To illustrate this, PRM was used to isolate the parent ions and fragment them for differentiation in the mass domain. Product ions corresponding to the respective amino acids were observed as the most prevalent fragments in the PRM experiments. Thus, these ions were chosen for quantitation. Surprisingly, the acyl fragments were not observed. A summary of the MS/MS data for the two pairs of isobaric lipoamino acids is presented in **Table 3-2**.

Table 3-2: Differentiation of isobaric *N*-acyl amino acids using Parallel Reaction Monitoring

Compound	Fragment	Formula	RT	Species	Calculated m/z	Observed m/z	Conc. Range (μM)
Eicosapentaenoyl valine	Eicosapenaenoate	C20H29O2-	12.3	[M-H] ⁻	301.2173	not detected	N/A
Eicosapentaenoyl valine		C5H10NO2-	12.3	[M-H] ⁻	116.0717	116.071	0.05-50 μM
Palmitoleoyl phenylalanine	Palmitoleate	C16H29O2-	12.7	[M-H] ⁻	253.2173	not detected	N/A
Palmitoleoyl phenylalanine		C9H10NO2-	12.7	[M-H] ⁻	164.0717	164.072	0.05-50 μM
Docosahexaenoyl valine	Docosahexaenoate	C22H31O2-	13.3	[M-H] ⁻	327.2330	not detected	N/A
Docosahexaenoyl valine		C5H10NO2-	13.3	[M-H] ⁻	116.0717	116.071	0.5-50 μM
Linoleoyl phenylalanine	Linoleate	C18H31O2-	13.0	[M-H] ⁻	279.2330	not detected	N/A
Linoleoyl phenylalanine		C9H10NO2-	13.0	[M-H] ⁻	164.0717	164.071	0.05-50 μM

Relative quantitation of N-acyl amino acids in biological samples

18 spectral features corresponding to *N*-acyl amino acids from the synthetic library were detected and their peaks were integrated. The data was analyzed using Partial Least Squares Differential Analysis (**Figure 3-4**) and visualized using box and whisker plots and stacked bar graphs. Chicks which had been fed lard-supplemented diets displayed the highest levels of total *N*-acyl amino acids in adipose tissue. As expected, the profile of elmiric acids was influenced by the fatty acid composition of the various diets. Supplementation with fish oil, which is rich in docosahexaenoic acid, resulted in elevated levels of *N*-docosahexaenoyl leucine, *N*-docosahexaenoyl valine, and *N*-docosahexaenoyl phenylalanine. In all other diets, these DHA-conjugated amino acids were absent. PLS-DA showed clear separation between the four diets as a function of *N*-acyl amino acid profile. DHAV, DHAL, and DHAF are prominent drivers of the separation in the PLS model, with Variable Importance in Projection (VIP) scores greater than 1. *N*-acyl amino acids were also detected in the serum of chicks from the same study, but the tissue extracts garnered better data. This is not surprising, as the enzyme responsible for the acylation, PM20D1, has been identified in adipocytes.¹¹

Detection of N-AAAs and bile acids in serum

While *N*-acyl amino acids and bile acids were detected best in their respectively ideal sample types, both sets of molecules were detected in serum. As previously described, serum from the broiler chicks was extracted and analyzed via UPLC-HRMS. While the peak intensities of *N*-acyl amino acids from serum were lower than those from tissue, most of the compounds were detected in both sample types. Additionally, six bile acids were detected in the same serum. This illustrates the utility of a combined method.

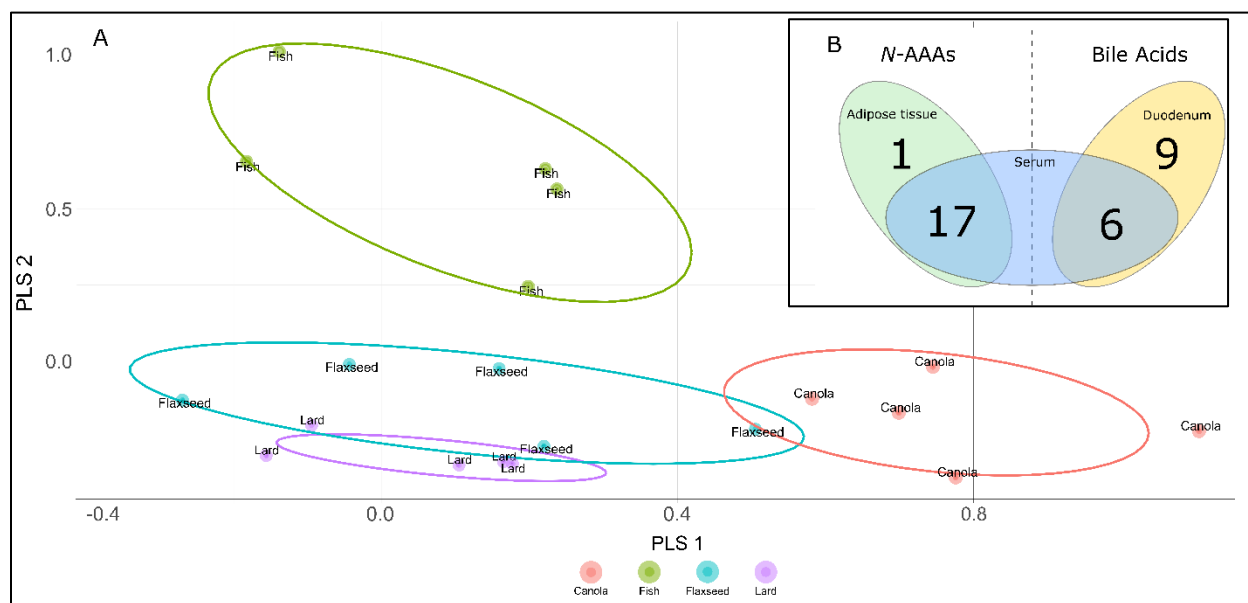


Figure 3-4: Detection of metabolites in biological samples from broiler chicks and mice. A) PLSDA of the *N*-acyl amino acids detected in avian adipocytes. Data is grouped by diet. B) Graphic illustrating the number of metabolites detected in each sample type

Absolute quantitation of bile acids in biological samples

Samples were extracted using the acidic acetonitrile protocol.¹²⁸ After UPLC-HRMS, areas of the extracted-ion chromatograms were calculated for all detected bile acids and deuterated standards. Using the calibration curves, exact concentrations were calculated for 15 compounds. Five of the detected bile acids were above the highest standard analyzed and were not quantifiable. **Table 3-3** shows the 10 bile acids that were quantified with the standard errors for each experimental condition. This biological data shows it is possible to simultaneously quantify several bile acids, including some isomeric compounds (aMCA and bMCA), with low error.

Table 3-3: Calculated concentrations of bile acids extracted from mouse duodenum

Bile Acid	Conc. (μM)
	10% Fat
chenodeoxycholate/deoxycholate(CDCA)/(DCA)	10.01 \pm 4.1
hyodeoxycholate (HDCA)	0.85 \pm 0.58
ursodeoxycholate (UDCA)	0.35 \pm 0.24
beta-muricholate (bMCA)	22.01 \pm 7.4
alpha-muricholate (aMCA)	22.73 \pm 5.31
glycochenodeoxycholate (GCDCA)	0.02 \pm 0.0056
glycodeoxycholic acid (GDCA)	0.04 \pm 0.011
taurolithocholate (TLCA)	0.16 \pm 0.030
taurochenodeoxycholate (TCDCA)	86.67 \pm 20.
taurooursodeoxycholate (TUDCA)/taurodeoxycholate (TDCA)	21.25 \pm 4.5

Conclusion

In conclusion, a single method using reverse-phase UPLC and electrospray high-resolution mass spectrometry was developed for detection of both *N*-acyl amino acids and bile acids. The method provides desirable chromatographic resolution of many bile acid isomers and can provide additional resolution of isobaric *N*-acyl amino acids if a hybrid orbitrap with PRM capability is used. Moreover, the method is suitable for profiling both bile acids and *N*-acyl amino acids in biological systems. In the future, as more biologically active *N*-acyl amino acids are discovered, the synthetic library of standards and the analytical method can be readily expanded to become more comprehensive. Thus, the described method is both robust for current analyses and adaptable for future investigations.

Chapter 3B: Expression of *N*-acyl amino acids is influenced by fasting-induced lipolysis and dietary supplementation with exogenous fatty acids

Introduction

Investigation of adaptive thermogenesis has been recently reinvigorated, since energy expenditure pathways are a promising target for therapy of metabolic diseases.¹⁰ Brown adipose tissue (BAT) and, to some extent, beige adipose tissue are the primary organs responsible for thermogenesis. These adipocytes can promote energy expenditure through a number of molecular mechanisms, which are being revealed by emergent research. Uncoupling protein 1 (UCP1) dependent means of thermogenesis have classically attracted much of the attention, but UCP1-independent mechanisms are gaining awareness. Several classes of small molecule thermogenic stimulators have been studied, and the elucidation of their mechanisms is underway.¹²⁹ One such family of metabolites, *N*-acyl amino acids (*N*-AAAs), recently demonstrated mitochondrial uncoupling activity—promoting thermogenesis and glucose homeostasis in mice.^{11, 130} As endogenously-expressed modulators of energy expenditure, *N*-AAAs possess potential for the safe combat obesity and pathological metabolic disease.⁸ While peptidase M20 domain containing 1 (PM20D1) is known to regulate *N*-AAAs *in vivo* by both hydrolysis and synthesis, the expression of *N*-AAAs is not well understood. For instance, they entail a diverse assortment of amino acid-bound fatty acids, but the selectivity of their synthesis has not been described. Likewise, it is not known whether PM20D1 is solely responsible for the acylation or if enzyme expression is even the most prominent way of regulating *N*-AAA synthesis.

Considering these questions, we hypothesized that lipolysis, through release of free fatty acid substrates, may regulate abundance of *N*-acyl amino acids. To study this, circulating *N*-AAAs can be measured during a period of induced lipolysis. Avians represent a suitable animal model, given the relative ease of inducing lipolysis *via* acute food withdrawal. In addition, they lack BAT, suggesting that UCP1-independent mechanisms of thermogenesis may be particularly active. If, indeed, circulating fatty acids are readily incorporated into *N*-AAAs, diet may have a direct impact on *N*-AAA profile and abundance. To this end, a parallel experiment comparing diets of contrasting fat

compositions can aid in illuminating any selective incorporation of fatty acids into *N*-AAAs. These studies require a method to quickly and generally profile *N*-AAAs from biological samples. Liquid chromatography paired with mass spectrometry provides an apt instrumental platform for accomplishing this goal with the desired sensitivity. The high degree of specificity afforded by mass spectrometry allows discrimination among *N*-AAAs, while the chromatography provides temporal resolution and ease of quantitation. Using such a platform, the *in vivo* expression and regulation of *N*-AAAs can be monitored to provide insight into this UCP1-independent means of promoting thermogenesis. This fundamental research provides invaluable precedence for forthcoming attempts to improve metabolic health by leveraging endogenous thermogenic mechanisms.

Materials and methods

Materials

HR Series NEFA-HR (2) assay kit was purchased from Wako Diagnostics (FUJIFILM Wako Diagnostics, Mountain View, CA). Chloroform, hydrochloric acid, HPLC grade methanol, HPLC grade acetonitrile, HPLC grade water, and formic acid were purchased from Fisher Scientific (Waltham, MA).

Animals and diets

All animal procedures were approved by the University of Tennessee Institutional Animal Care and Use Committee. Mixed-sex Cobb 500 broiler chicks were obtained from a local hatchery and used for both the sequential fasting and diet studies. For the fasting study, chicks were fed a commercial broiler starter diet *ad libitum*, beginning at hatch. At 21 d of age, chicks were randomly allocated to one of five experimental groups: continued *ad libitum* feeding, fasted for four, eight or twelve hours (n=10), or fasted for 12 hrs and then returned to *ad libitum* feeding for two hours (n=5). For the diet study, chicks were fed *ad libitum* a commercial starter diet from hatch until 7 d, then switched to experimental diets containing fat (8% wt:wt) from lard (LA), canola oil (CA), flaxseed oil (FL), or fish oil (FO) into a commercially formulated starter diet base, as previously described. Chicks were fed experimental diets from 7 to 30 d. For both studies, chicks were euthanized by CO₂ asphyxiation. At the time of euthanasia blood was collected by cardiac venipuncture and

transferred to 10 ml serum separator tubes (Fisher Scientific, Pittsburgh, PA), and serum was separated by centrifugation. Samples of abdominal adipose tissue were dissected and immediately snap-frozen. Serum and tissue samples were stored at -80° C until extraction for *N*-AAA analyses.

Extraction of N-acyl amino acids

For each sample, 100 mg of tissue or 0.100 mL of serum was aliquoted into a 2 mL microcentrifuge tube. 800 µL of a mixture containing 1:1 0.1 N aqueous hydrochloric acid: methanol was added, followed by 400 µL of chloroform.¹²⁶ Each sample was vortexed for 1 minute and centrifuged for 5 minutes at 13,000 rpm. The organic (lower) phase was isolated and concentrated to dryness under a stream of dry nitrogen gas. Each sample was resuspended in 300 µL of 10% chloroform in methanol prior to analysis.

Measurement of N-acyl amino acids and free fatty acids

Relative quantitation of *N*-AAAs was performed via a previously-reported Ultra-Performance Liquid Chromatography—Mass Spectrometry (UPLC-HRMS) method which was developed and validated in-house. In short, reverse-phase chromatography using a C-18 column with a solvent system of 0.1% formic acid in water and 0.1% formic acid in acetonitrile was employed for separation. The chromatograph was coupled to an Orbitrap Exactive Plus mass spectrometer (Thermo Fisher Scientific, Waltham, MA) equipped with an electrospray ionization source. Samples were analyzed in negative mode. Using this method, serum and tissue samples were screened for 27 *N*-AAAs featuring hydrophobic amino acids valine, leucine, and phenylalanine by matching exact *m/z* and retention time to an in-house standard library. A commercially available colorimetric assay kit was used to measure serum non-esterified fatty acid (NEFA) levels (Wako Chemicals, Neuss, Germany).

Data analysis

After UPLC-HRMS analysis, data files were converted from *.raw* to *.mzml*.¹²¹ Peaks for all detected *N*-AAAs were integrated using Metabolomic Analysis and Visualization Engine (MAVEN), an open-source software for processing LC-MS data.¹¹⁹ Peak areas of metabolites from adipocytes were normalized to exact weights of frozen tissue. Several

R packages (version 3.4.3) were employed for multivariate analysis, statistical power analysis, and data visualization.^{122-124, 131} ANOVA and correlation analyses were conducted using MetaboAnalyst version 4.0.¹³² Microsoft Excel was used to construct bar graphs and perform additional statistical tests.

PM20D1 expression

Expression of PM20D1 was quantified in abdominal adipose tissue of fed and 12 hr fasted chicks (n=6/group). Total RNA was isolated from approximately 200 mg of abdominal adipose tissue using Invitrogen™ TRIzol™ (Invitrogen, Carlsbad, CA) and reverse transcribed to cDNA using a commercial kit (iScript cDNA Synthesis Kit, Bio-Rad Laboratories, Hercules, CA). Primers specific for PM20D1 and TBC1 domain family, member 8 (*TBC1D8*; used as a housekeeper) were designed using PRIMER3 and based on Gallus_gallus-5.0 (Ensembl release 94).¹³³ QPCR was performed in triplicate for each sample using iQ SYBR Green Master Mix (Bio-Rad Laboratories, Hercules, CA), as previously described.¹³⁴

Results

Fasting-induced lipolysis study

The pool of 27 *N*-AAA reference standards included nine fatty acids, in lengths from 12-22 carbons, combined with three amino acid head groups (valine, leucine, and phenylalanine). These head groups were chosen because of the emphasis on hydrophobic amino acids in the work published by the Spiegelman lab.¹¹ Of these, 18 endogenous *N*-acyl amino acids were detected in serum of broiler chicks that were fed, progressively fasted, or fasted and re-fed. The pool of detected *N*-AAAs included species containing eight different fatty acids and the amino acids valine, phenylalanine, and leucine. Species containing longer-chain fatty acids (20:0, 20:5, and 22:6) were largely absent, with the exception of docosahexaenoyl leucine. Moreover, oleoyl-conjugated amino acids accounting for approximately 35-40% of the detected *N*-AAA pool. This is readily explained by the corn-based diet of the chicks. Effects of progressive fasting on the abundance of circulating *N*-AAAs was evaluated by ANOVA. Fasting significantly affected the total *N*-AAA pool size, which incrementally increased up to 8 hours of fasting.

Partial least squares discriminant analysis (PLS-DA) illustrates the incremental separation between fasting durations (**Figure 3-5**). Total abundance of *N*-AAAs increased (235%) after four hours compared to the fed group ($p=0.0577$). Levels after eight hours of feeding increased by an additional 229% compared to four hours of fasting ($p=0.0085$) and were 464% greater than fed levels ($p=0.00005$). No further increases were found after an additional four hours of fasting (8 vs. 12 levels; $p=0.28$). Serum *N*-AAAs in fasted chicks were rapidly restored to those of *ad libitum* fed chicks by two hours of refeeding (**Figure 3-6**). Levels in the fasted/re-fed group were not different than those from fed controls ($p=0.24$). Total *N*-AAA and serum NEFA levels were compared across time points to assess the relationship free fatty acid availability and *N*-AAA synthesis. Circulating NEFA levels exhibited a similar profile with respect to the fasting state and were correlated with total abundance of *N*-AAAs ($r= 0.976$ from comparison of mean total *N*-AAA to mean total NEFA). These data support the hypothesis that lipolysis is accompanied by acylation of amino acids.

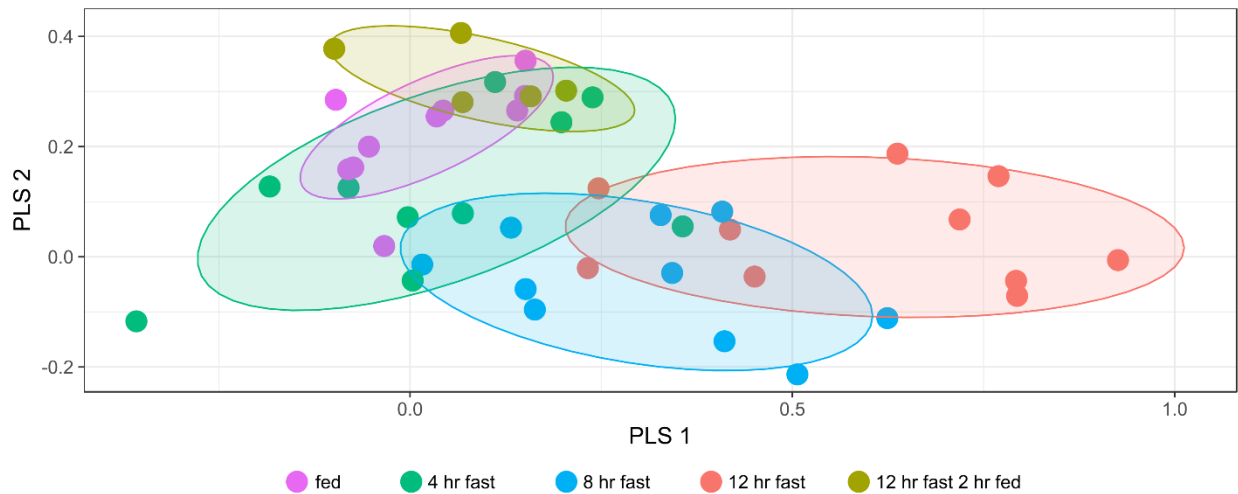


Figure 3-5: PLS-DA of fasting states based on serum *N*-AAAs

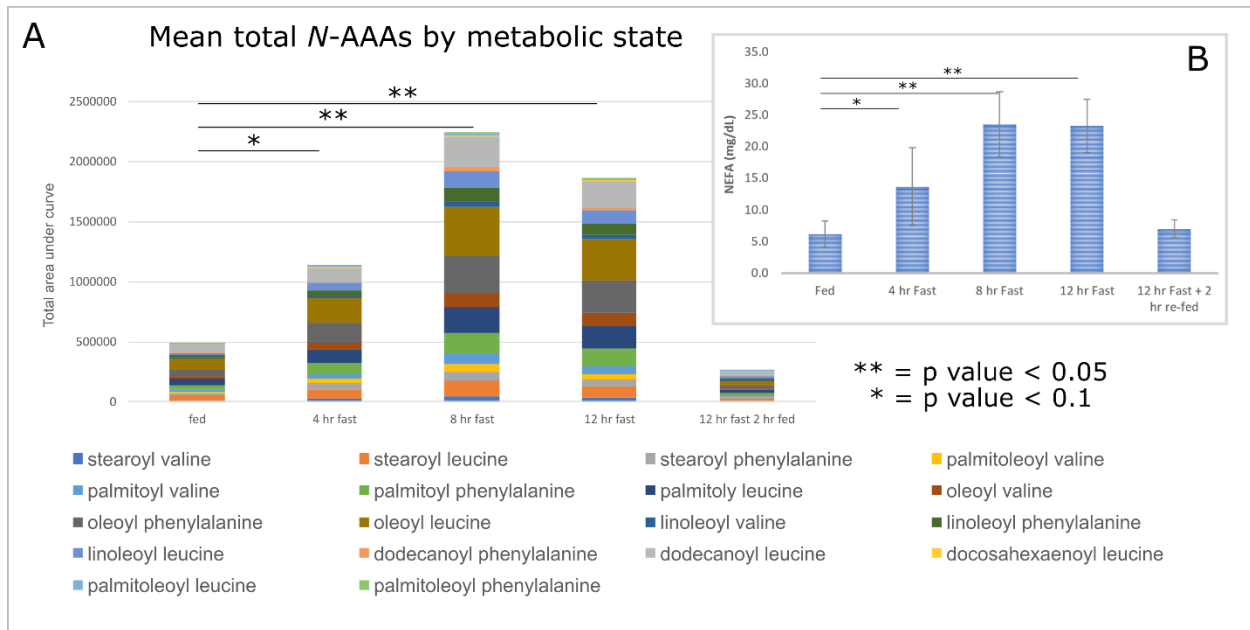


Figure 3-6: Metabolites quantitated in serum of fasted chicks. A) N-AAAs; B) NEFA

In addition to the analysis of bulk N-AAA changes across fasting states, relative percent composition of individual N-AAA species was evaluated (**Figure 3-7**). Abundance of each individual N-AAA was normalized to the total pool level and analyzed by ANOVA. Six N-AAAs varied across the duration of fasting ($p < 0.05$). In general, those containing saturated fatty acids decreased with fasting, while those containing unsaturated fatty acids increased. This can be observed when comparing percentages of stearoyl (18:0) leucine, oleoyl (18:1) valine, and linoleoyl (18:2) phenylalanine among groups. Stearoyl leucine displayed a marked percent-of-total decrease over the fasting period, while percentages of the two N-AAAs that contained unsaturated 18-carbon acyl chains both increased with duration of fast. This may corroborate preexisting literature which suggests that unsaturated fatty acids are more freely mobilized than those which are more highly saturated.¹³⁵ The relationship between saturation of the fatty acid tail and abundance during fasting can be noted in the correlation map below (**Figure 3-8**).

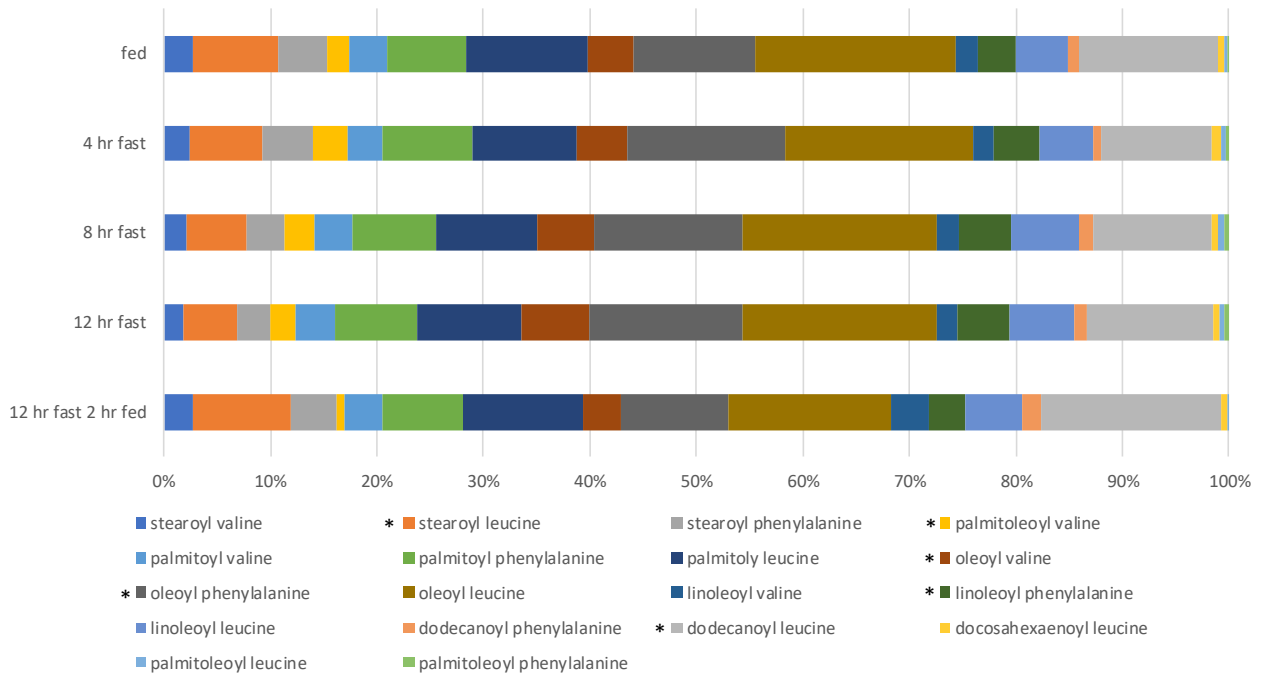


Figure 3- 7: Percent N-AAA composition by fasting state. ANOVA significant N-AAAs are starred

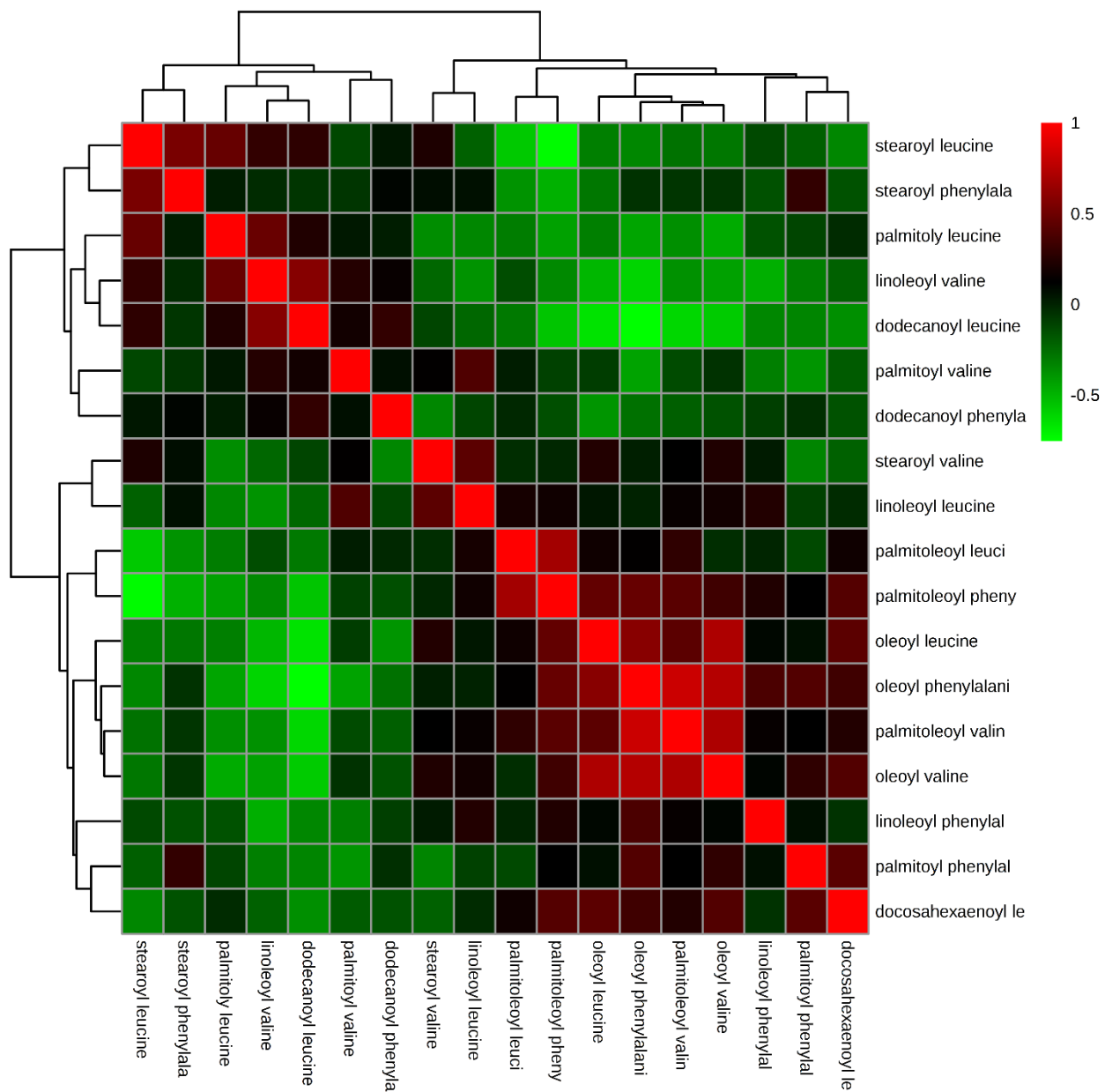


Figure 3-8: Correlation matrix of N-AAAs detected in the serum of fasting chicks

Adipose tissue synthesizes and secretes Peptidase M20 domain-containing enzyme 1 (PM20D1), a bidirectional enzyme that regulates both synthesis and degradation of *N*-AAAs.¹¹ The Align tool in UniProt (uniprot.org) was used to determine that conservation of the amino acid sequence of PM20D1 is relatively high (56.5 %) between chicken and mouse, the species in which PM20D1 was shown to regulate *N*-AAA levels, and that the active sites are perfectly conserved between the two species. Expression of PM20D1 was analyzed by QPCR in adipose tissue to determine if fasting coordinately regulated the dominant enzyme responsible for *N*-AAA abundance. Despite the increase in circulating *N*-AAAs, expression of PM20D1 was decreased in fasted (12 hr) vs. fed controls ($p = 0.03$, **Figure 3-9**). This suggests that the condensation to produce *N*-AAAs may be more dependent on substrate availability than enzyme expression. In addition, because PM20D1 also hydrolyzes *N*-AAAs, it suggests that downregulation of this enzyme may be a mechanism to sustain *N*-AAA levels during periods of energy restriction.

Diet study

Adipose tissue plays a key role in *N*-AAA synthesis through both expression of the PM20D1 enzyme and as the source of fatty acids that are esterified with amino acids. It is also a potential clinically relevant functional target of *N*-AAAs because of their thermogenic effects that are mediated through uncoupling of mitochondrial metabolism. However, local synthesis of *N*-AAAs within adipose tissue has not been reported. Abundance of *N*-AAAs was analyzed in adipose tissue homogenates. Chicks were fed diets enriched in different types of fatty acids to determine if *N*-AAA profiles reflected tissue fatty acid composition. A total of 18 *N*-AAA species, containing 7 different fatty acid tails, were detected in adipose tissue homogenates from one or more diet groups. No species featuring stearoyl or eicosanoyl tails were detected. Composition of the *N*-AAA pool varied markedly in response to dietary fat source. When combined based on fatty acid component, five *N*-AAA subpools (those containing palmitic, oleic, linoleic dodecanoic and docosahexaenoic acids) were significantly affected by diet ($p < 0.05$, **Figure 3-10**).

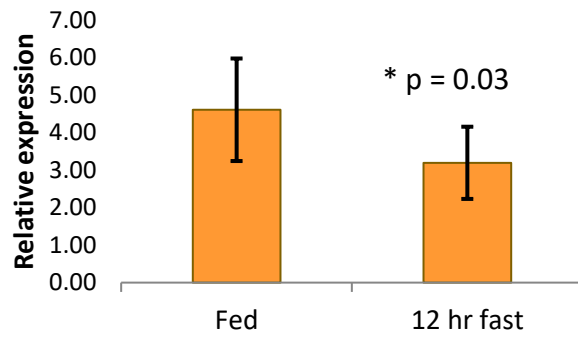


Figure 3-9: Expression of *PM20D1*

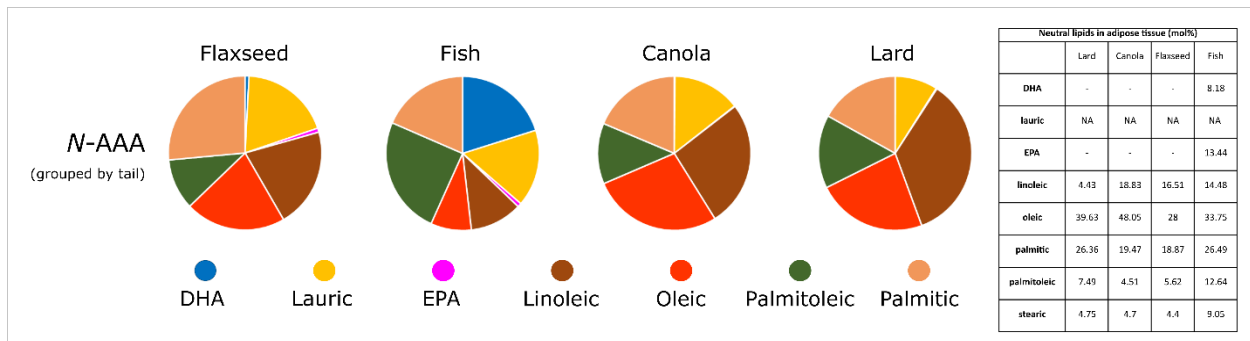


Figure 3-10: *N*-AAAs (grouped by fatty acid) detected in chicks fed 4 distinct diets. The inset shows the levels of fatty acids in neutral lipid from adipose tissue of chicks from the same groups

For example, DHA- and EPA-conjugated amino acids were detected in tissue of chicks fed diets enriched with fish oil, and to a lesser extent flaxseed oil, but absent in the canola and lard groups. However, these data also suggest that some fatty acids, may be preferentially incorporated in *N*-AAAs. For example, across the four diets, tissue levels of linoleic acid were lower in the lard group (~ 4%) than in the other three diets (~ 14-18%). In contrast, *N*-AAAs containing linoleic acid were more abundant in lard-fed chicks (~ 35%) than in the three other diets (~ 10-25%). Differential effects of diet on *N*-AAA composition relative to tissue fatty acid profile were particularly apparent for species containing DHA. This fatty acid comprised ~ 8% of tissue fatty acids in fish oil-fed chicks, yet was present in ~ 26 % of *N*-AAAs.¹²⁷ In contrast, EPA was also enriched in tissue from the fish oil group and in fact represented a greater proportion (~ 13%) of tissue fatty acids than did DHA. However, EPA was markedly less abundant (~ 0.8%) than DHA (~ 26%) in the *N*-AAA pool of fish oil chicks, and only one EPA-containing *N*-AAA species (eicosapentaenoyl valine) was detected in adipose tissue. Similarly, while the oleoyl (18:1) tails accounted for 28-48 % of neutral lipid in tissue and the linoleoyl (18:2) tails accounted for only 4-19%, oleoyl and linoleoyl amino acids were equally represented in each of the 4 groups. In combination, these data indicate that while the dietary fatty acid profile does influence the composition of the tissue *N*-AAA pool, there appear to be additional, selective mechanisms of *N*-AAA synthesis in adipose tissue.

Palmitoleoyl leucine and palmitoleoyl phenylalanine were elevated in the fish-fed group as compared to the canola and flaxseed groups. Since fish oils are a dietary source of palmitoleic acid (up to 7% of the total fatty acids)¹³⁶ enrichment in palmitoleoyl-conjugated amino acids for FS is expected. However, the mean levels of these two metabolites were highest (but highly variable) in LA. As evidenced by the shaded matrix of Pearson correlation constants (**Figure 2-11**), *N*-AAAs were plainly correlated based on their featured fatty acids. For instance, DHA-bound amino acids displayed high positive correlations with one another, as well as a less pronounced positive correlation with eicosapentaenoyl valine. Oleoyl-bound AAs displayed a decidedly negative correlation with DHA-bound AAs, and a positive correlation to linoleoyl, palmitoyl, and dodecanoyl-bound AAs. This illustrates the expected relationship between dietary abundance of fatty acids and acylation of amino acids.

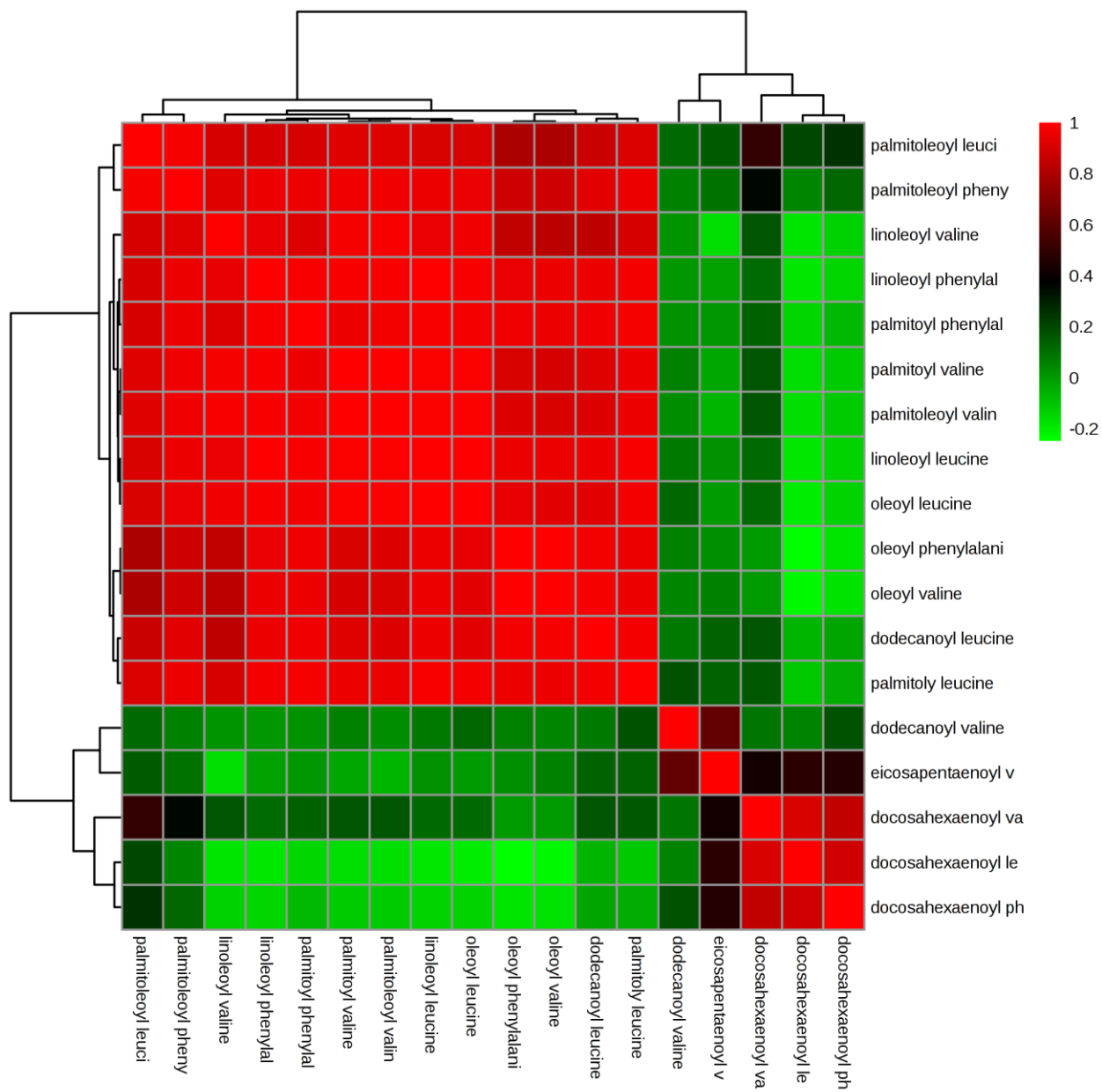


Figure 3-11: Correlation map of *N*-AAAs detected in the tissue of chicks fed 4 distinct diets

Discussion

As evidenced by the inverted ratios of EPA/DHA and oleate/linoleate in neutral lipid and *N*-AAA pool, there appears to be a preferential incorporation of fatty acids with higher degrees of unsaturation. These data suggest that biosynthesis may be selective or indirectly regulated by preferential mobilization of certain fatty acids. This trend is corroborated by the absence of stearyl amino acids detected in either sample type, despite accounting for 4-9% of neutral lipid from the adipose tissue. However, lauryl amino acids were abundant in serum and tissue—a trend which is initially puzzling due to the repeatedly demonstrate preference for incorporation of unsaturated tails. It has been documented that the mobilization of fatty acids into plasma is not proportional to tissue composition. While relative mobilization is positively correlated with unsaturation, it is negatively correlated to chain length.¹³⁵ Previous work showed that polyunsaturated species are most readily mobilized, yet EPA exceeds the relative mobilization of DHA due to the negative contribution of chain length.¹³⁵ However, lauric acid was less easily mobilized than longer fatty acids.¹³⁷ This suggests that—while preferential mobilization appears to indirectly regulate biosynthesis by biasing the pool of substrates—PM20D1 does favor the incorporation of certain fatty acids independent of substrate concentration.

The observation that, in some cases, unsaturated fatty acids are preferentially conjugated amino acids is intriguing, given recent findings by Lin and coworkers.¹³⁰ They observed an incremental increase in uncoupling activity based on unsaturation (C18:0-Phe < C18:1-Phe < C18:2-Phe) which drops off for polyunsaturated species. This begins to reveal a cascade in which diet—by influencing the pool of substrates for *N*-AAA synthesis—can promote differential expression of *N*-AAAs with greater therapeutic impact.

It is noteworthy that DHA-Leu and DHA-Phe also accounted for a small percentage of the detected *N*-AAAs in the flaxseed group. This is explained by the known mechanism through which α -linoleic acid (ALA) is converted to DHA. Linoleic acid (LA) and α -linoleic acid compete for a single set of desaturating and elongating enzymes, which generally limits an appreciable conversion of ALA to DHA.^{138, 139} However, when the relative ratio of ALA to LA is increased, DHA expression is increased.¹³⁹ Flaxseed oil has a high

abundance of ALA (39.0-60.4%), making it one of the richest dietary sources of ALA.^{140,}
¹⁴¹ Thus, the presence of DHA-conjugated amino acids in the flax samples is consistent with the existing literature on conversion of ALA to DHA.

It is also notable that certain *N*-AAAs have demonstrated analgesic and anti-inflammatory activities in previous studies.^{95, 99, 100} Certain *N*-AAAs, especially those featuring polyunsaturated fatty acids, have been shown to increase the ratio of J series prostaglandins to E series both *in vitro* and *in vivo*.⁹⁹ Prostaglandin ratios favoring the J series have been previously shown to promote the resolution of inflammation.⁹⁹ Most recently, the roles of *N*-AAAs in nociception were demonstrated by the antinociceptive behaviors of PM20D1 knockout mice exposed to chemical and inflammatory pain stimuli and the corresponding decrease of some *N*-AAAs in response to chemical pain stimulus.¹⁴² This implicates *N*-AAAs as potential regulators of physiologic nociception. Considering the observation that fasting readily induces *N*-AAA formation, the current work may begin to reveal the molecular mechanism of pain attenuation through fasting. In several recent studies, fasting has been shown to affect nociception in a way that suggests improved pain tolerance relative to postprandial states.¹⁴³⁻¹⁴⁵ This has been demonstrated in both mice and humans. In one case, hunger was shown to selectively inhibit inflammatory pain in comparison to acute pain.¹⁴⁵ Additionally, previous work has implicated caloric restriction as a way to attenuate inflammatory response to reperfusion following ischemia.¹⁴⁶ More broadly, intermittent fasting and caloric restriction have demonstrated an ability to reduce inflammation, especially in metabolically unhealthy patients.¹⁴⁷⁻¹⁵¹

Since *N*-AAAs have demonstrated these abilities and their abundance has been positively correlated with fasting duration, they may be involved in an endogenous process of pain and inflammation regulation based on need state. Similarly, considering the known vulnerability of cancer cells to nutrient deprivation and the documented ability of some *N*-AAAs to inhibit cancer cell proliferation, *N*-AAAs may partially account for the observed benefits of fasting during cancer therapy.^{96, 152} Given these demonstrated relationships, which often link fatty acid composition with therapeutic effect of *N*-AAAs, as well as the

clear influence of diet on their expression, *N*-AAAs may underpin the hypothesis that the benefits of fasting may depend, in part, on the composition of the baseline diet.

Conclusions

Parallel experiments demonstrate that lipolysis rapidly increases the total amount of *N*-AAAs in serum and that the composition of dietary fatty acids notably influences *N*-AAA profiles in tissue. In the lipolysis study, the correspondence between serum NEFA levels and *N*-AAA abundance suggests that availability of free fatty acids may be a primary driver of acylation. In addition, it reveals potential mechanisms for lipolysis induced energy expenditure independent of protein uncouplers such as UCP1. In the case of exogenous fat supplementation, *N*-AAA profiles were distinctly influenced by the dietary fatty acid composition. When comparing the composition of neutral lipid in tissue, enrichment of the *N*-AAA pool with certain fatty acids was observed. This begins to illuminate a probable complex regulation in which preferential mobilization of certain fats manipulates the pool of substrates, which are incorporated into *N*-AAAs with some selectivity. Furthermore, this information about the regulation of *N*-AAAs by fasting and diet may clarify the molecular mechanisms that underly the emerging benefits of calorie restriction.

CHAPTER FOUR: INVESTIGATION OF THE TOXIC ZORO PEPTIDE

Abstract

ZorO (Z-protein often repeated) is a small protein expressed by pathogenic enterohemorrhagic *Escherichia coli* (EHEC *E. coli*) which was recently discovered through bioinformatic searches. While its biological role is still being elucidated, it has been demonstrated to exhibit toxic effects on the cell when overproduced. This puzzling genetic ability for what, at first glance, appears to be self-destruction may be nuanced way for wild strains of bacteria to adapt to host environments or survive stress. This likely entails a much more minimal expression of ZorO, initiating subtle cell behavior that is less apocalyptic. In order to investigate the localization and function of endogenously-expressed ZorO, the protein was be synthesized for biophysical experimentation using solid-phase peptide synthesis. In addition, several analogs of ZorO were synthesized to further enable an investigation of structure-activity relationship. These analogs include versions of ZorO with single amino acid mutations, as well as variants with “tags” for detection, purification, or capturing transient interactions. In the latter case, an unnatural replacement of the amino acid methionine (photo-methionine) was synthesized for incorporation into a synthetic peptide analog. This synthesis work required exploration of literature and reaction optimization, ultimately leading to a novel synthesis of photo-methionine. Much of the synthetic groundwork for studying ZorO has now been laid, and efforts are underway to develop analytical methods for detecting ZorO in cell extracts—a daunting task considering the complex protein composition of the cell and the wide concentration ranges spanned by endogenously-expressed polypeptides.

Preface

The work in this chapter is previously unpublished. It is part of an ongoing collaborative project, in partnership with the Fozo microbiology lab at UTK, with more components than those described herein. The following account is biased towards the chemical component of the project, though ample attention is given to the biological rationale and larger goals. The application of this work is ongoing.

Introduction

The forgotten proteins

Though many proteins have been extensively studied, there remains a class of proteins that has been historically ignored by science. These are small proteins, defined as polypeptides that obtain their slight size—often less than 50 amino acids—by translation of small open reading frames (ORF).¹⁵³ In other words, these polypeptides are synthesized independently instead of expressed by proteolytic cleavage of precursors. Many of these small proteins have been identified either through serendipity or by identification of their respective ORFs with bioinformatics.¹⁵³ While the biological functions of some small proteins have been elucidated, many have unknown activity. Several have been confirmed to play a role in toxin-antitoxin (TA) loci. More specifically, the type 1 toxin-antitoxin pair features a small, hydrophobic protein as the toxin and a strand of antisense sRNA as the antitoxin.¹⁵⁴ These loci have been identified in both gram-positive and gram-negative bacteria.¹⁵⁴ In some cases, toxicity is only induced when the mRNA responsible for encoding the toxin is overexpressed. Toxicity is repressed when the antitoxin base-pairs with the toxin mRNA.¹⁵⁴

ZorO, the type-1 toxin

In recent years, a new type 1 TA family was identified in *Escherichia coli* through bioinformatic searches.¹³ This locus includes two 29 amino acid proteins, differing by a single residue, which are toxic to *E. coli* when overproduced. These have been termed ZorO and ZorP (Z-protein often repeated) and their respective antitoxins, OrzO and OrzP (overexpression reduces Z-protein toxicity).¹³ The focus of this project is the *zorO-orzO* system. Intriguingly, this locus is highly conserved in most pathogenic strains of *E. coli* but absent in common laboratory strains.¹⁴ In true type 1 TA fashion, OrzO binds the 5' end of the mRNA encoding ZorO and represses toxicity in a target-specific manner.¹⁴ While the toxin-antitoxin relationship is understood, there are looming and unanswered questions regarding the biological function of the ZorO protein including whether or not it serves a purpose besides its toxicity.¹⁴ Since domesticated strains of bacteria lack the responsible genes, the *zorO-orzO* locus could be integral to *E. coli* growth in natural habitats.

An astounding number of these small proteins with known biological roles are membrane-localized.¹⁵ In fact, 65% of small proteins produced by *E. coli* are predicted to contain a trans-membrane domain.¹⁵ ZorO follows this trend, as do other type 1 toxins. For instance, TisB, a small protein involved in a similar TA locus, has garnered some attention due to its function within membranes.¹⁵⁵ Experiments show that it aligns antiparallel with itself to create dimers via the formation of 4 salt bridges by electrostatic attraction of oppositely charged residues. This has been termed a “charge zipper”, which has implications in transport (**Figure 4-1**).¹⁵⁶ TisB has been shown to interrupt the proton gradient, which may be explained by the ability of the polar charge zipper interface to transport water into the hydrophobic bilayer core. This could lower the activation barrier for hydroxide ions and protons to pass through the membrane and be neutralized.¹⁵⁶ It is reasonable to suspect that the biological role of ZorO may also be linked to its probable membrane localization. Moreover, preliminary results suggested that ZorO may mimic the function of TisB and form membrane-spanning dimers or oligomers. Certain ZorO derivatives with charged residues replaced do not confer toxicity (unpublished results, Fozo Lab, UTK). Thus, we hypothesized that ZorO may be forming salt bridges and self-dimerizing to create channels that similarly facilitate the breakdown of proton gradients. In this way, it could abolish the link between nutrient synthesis and ATP production by uncoupling respiration.

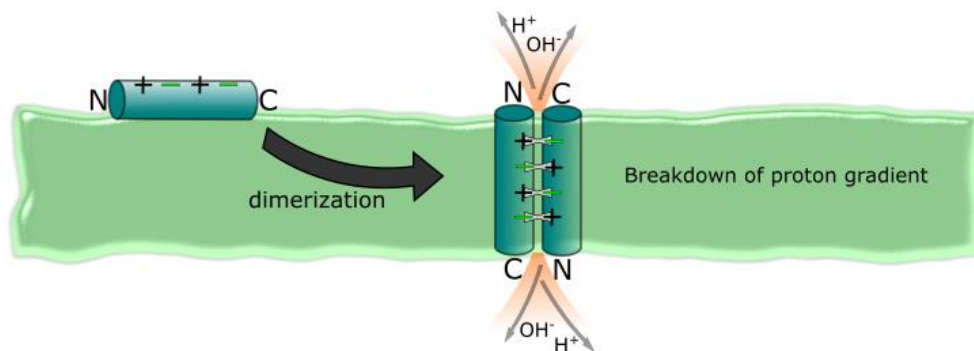


Figure 4-1: Potential mechanism of ZorO toxicity

Project goals

The chosen approach for investigating the biological question of ZorO's mechanism of action involves a substantial amount of synthesis. To elucidate ZorO's putative biological action and localization, appreciable amounts of ZorO are desired. Thus, an initial goal is the synthesis of native ZorO in milligram quantities, allowing the structure and potential oligomerization of ZorO to be probed *in vitro*. Synthesis is an attractive alternative to isolation and enrichment, which can be a significant challenge.

It is also desirable to obtain analogs of ZorO for investigating structure-activity relationship. Analogs with single amino acid mutations can illuminate any influence of sequence on toxicity. Specifically, the charge of the 23rd residue, arginine (R) has been shown to be essential for ZorO toxicity. When R23 is replaced with a lysine *via* mutations to the *zorO* gene, positive charge is retained as is toxicity. When R23 is replaced with a neutral residue like leucine, the resultant peptide loses toxicity (unpublished results, Fozo laboratory, UTK). Thus, it is desirable to synthesize R23K and R23L mutants of ZorO to investigate the way that their structure and/or oligomerization may change in relation to that of native ZorO.

If ZorO is indeed forming oligomers or recruiting other proteins to the membrane, then understanding these associations could provide invaluable data. To this end, photo-cross-linking will be employed as a means of capturing protein interactions for study. In short, photo-cross-linking is a technique which allows strict temporal control over interactions, allowing transient complexes to be covalently captured (**Figure 4-2**).¹⁵⁷ This is accomplished by introducing tags that decompose to highly reactive intermediates (carbenes, nitrenes, or radicals) when exposed to irradiation. These tags can either be external reagents, or modifications to the molecule of interest.¹⁵⁷

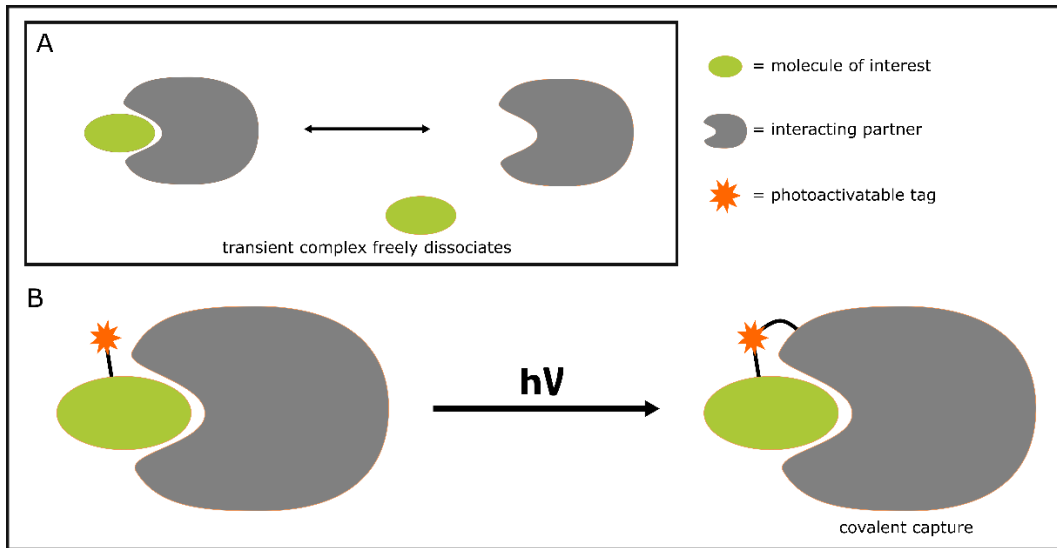


Figure 4-2: Photo-cross-linking

To allow photo-cross-linking, a variant of ZorO, containing a photoactivatable handle which will rapidly react with any molecules in immediate proximity, will be synthesized. The modification must be slight enough to avoid interference with the peptide structure and function, and stable enough to survive in a biological environment until irradiation. For this application, replacement of ZorO's *N*-terminal methionine (M1) with a diazirine-containing unnatural analog should allow the desired activity. This so-called photo-methionine (**Figure 4-3**) has been used previously in photo-cross-linking applications with success.^{158, 159}

Attaching a handle for detection and purification of ZorO is another attractive modification. While many such tags entail attaching the target polypeptide to a second protein, creating a "fusion protein", this has many disadvantages for the present study.¹⁶⁰ These fusions often interfere with the folding of the protein, potentially impacting function.¹⁶⁰ Especially when studying a protein as diminutive as ZorO, care must be taken to alter the molecular weight as little as possible to reduce the impact on function or transport. Therefore, low molecular weight tags which have known reactions with monoclonal antibodies are attractive options.

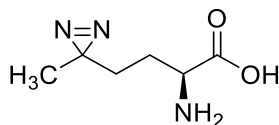


photo-methionine (1)

Figure 4-3: Structure of photo-methionine

The short polypeptide sequence H-D-Y-K-D-D-D-D-K-OH, termed 1X-FLAG, has been successfully used for protein detection and purification. With the complimentary antibody, it can be used for immune-affinity purification or detection by western blotting.¹⁶⁰ As the final synthetic objective, an analog of ZorO bearing this *N*-terminal 1X-FLAG will be made. In addition to these synthesis-heavy aims, a parallel goal includes establishing a method of analytically detecting ZorO extracted from cell cultures. This will be accomplished by UPLC-HRMS with an objective to optimize ZorO detection for eventual measurement of chromosomally-expressed ZorO.

Overview of solid-phase peptide synthesis

To synthesize the peptide, a traditional solid phase peptide synthesis (SPPS) was employed. Pioneered by Robert Bruce Merrifield, who was awarded the Nobel Prize in chemistry in 1984, SPPS has become the method of choice, as it boasts many improvements over solution phase approaches.¹⁶¹ Most notably, solid-phase synthesis allows minimal purification between successive reactions. There are many components necessary for SPPS, including a solid support, activating groups, linkers, and protecting groups.¹⁶¹ The general workflow is illustrated in **Figure 4-4**.¹⁶¹ The process begins with the C-terminal residue which has a protected *N*-terminus. It is then anchored to a solid-support resin containing either by amidation or esterification. Once the C-terminal residue is linked to the resin, the *N*-terminus is deprotected, leaving the free N^α amine. The next residue, with an activated C-terminus and protected *N*-terminus, is added, creating the amide bond between residues. Successive deprotections and additions of prepared residues results in the elongation of the biopolymer.

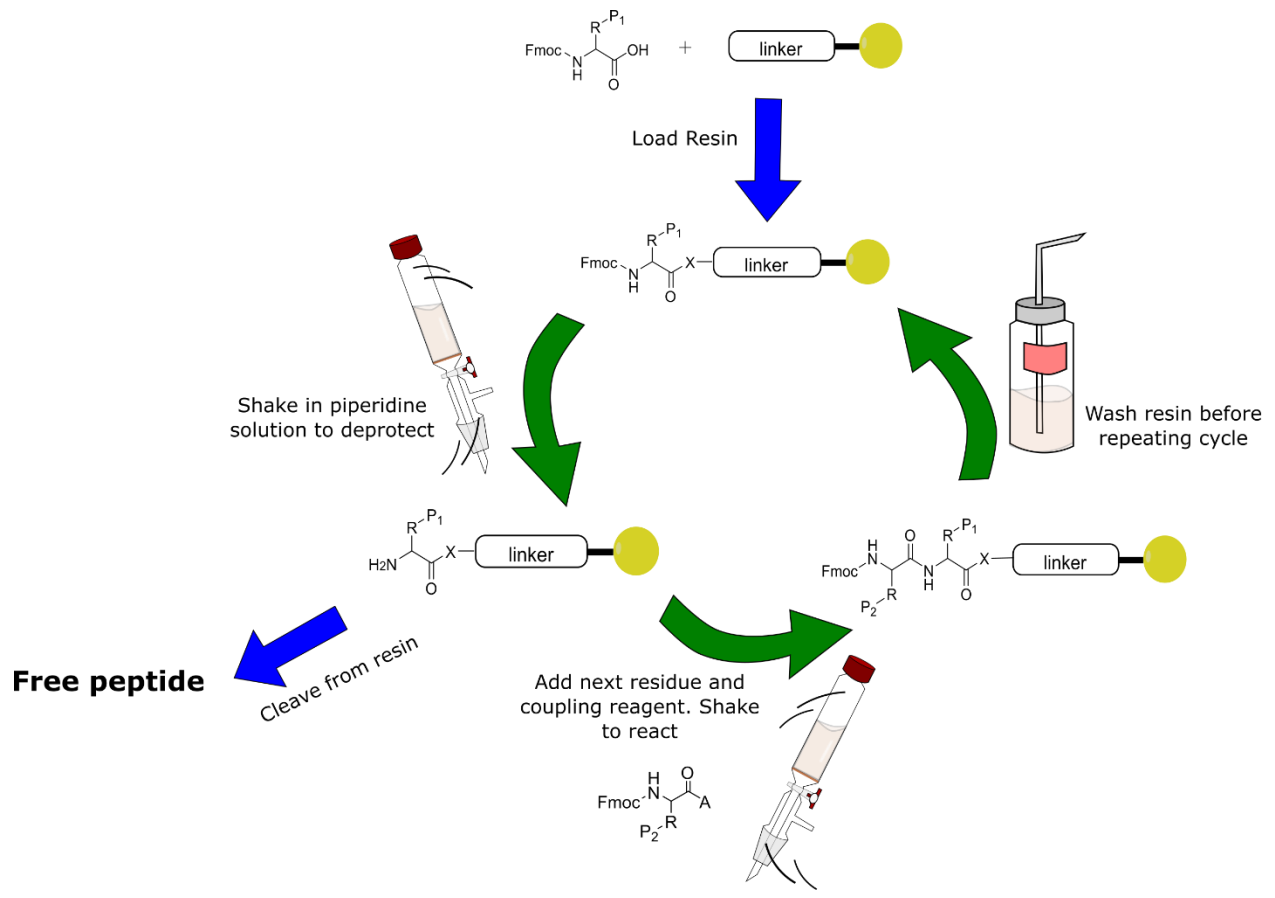


Figure 4-4: Overview of solid-phase peptide synthesis (SPPS)

The most common of these strategies relies on fluorenylmethoxy-carbonyl (Fmoc) and tert-butyl (tBu) protecting groups.¹⁶¹ Crosslinked polystyrene (PS) beads are widely used as the solid support, but require a low loading to assemble difficult sequences.¹⁶¹ Liberation of the Fmoc-protected *N*-terminus is achieved by a brief treatment with piperidine and subsequent washing.¹⁶¹ Coupling reactions are carried out by addition of excess activated amino acid. Vigorous stirring is unnecessary and potentially detrimental, as diffusion phenomena dictate the reaction kinetics.¹⁶¹ After the entire sequence is assembled, a global deprotection liberates the peptide from the solid support and simultaneously deprotects the side chains. This is accomplished by the use of an acidic cleavage cocktail containing silane scavengers to quench carbocations that are generated.¹⁶¹ The cocktail consists of trifluoroacetic acid, water, and triisopropyl silane.¹⁶¹ This process affords intact peptides which can be isolated by precipitation, centrifugation, and lyophilization. This entire process can be accomplished manually, with potential intervention at each step of the process, or automatically, with pre-programmed synthetic sequences. Each approach has its unique advantages and disadvantages.

Precedence for diazirines in photo-cross-linking applications

Diazirines were first synthesized and characterized in the 1960's.¹⁶² Their notable attributes include a remarkable stability despite the strained appearance of the three-membered ring, along with their ability to liberate nitrogen gas and form highly reactive carbene species upon photolysis, pyrolysis, or ultrasonication.¹⁶² Thus, diazirines make ideal functional groups for photo-cross-linking or photoaffinity labelling applications. Diazirines have a number of advantages over other commonly used photo-cross-linking groups. For example, benzophenone, a classic photoactivatable moiety, requires longer irradiation times to produce its reactive triplet state carbonyl.¹⁶³ This introduces a risk for nonspecific labelling¹⁶³ and could also present more opportunity to negatively affect cell viability. Perhaps the greatest advantage of the diazirine for photo-cross-linking applications is its compact size. When probing interactions, it is necessary to create a suitable mimic of the desired partner¹⁶³ by choosing a tag that minimally changes its steric properties. Diazirines are nearly isosteric with methyl groups¹⁶⁴ and among the smallest groups used for photo-cross-linking. Thus, they excel in this regard. Another advantage

of diazirines is their short half-life, lending them high specificity.¹⁵⁹ Their carbene intermediates undergo rapid quenching when not poised for crosslinking.¹⁶⁴ In the absence of nearby molecules, carbenes are quickly quenched either by solvents or by internal rearrangement to form alkene species.¹⁶⁵

While aryl azides had been previously employed in crosslinking^{166, 167}, diazirines were not used in this application until the early 90's. In 1992, 3-[3-(3-(trifluoromethyl)diazirin-3-yl)phenyl]-2,3-dihydroxypropionic acid was developed as a crosslinking reagent by Bochkariov and Kogon.¹⁶⁸ In 1993, Khorana and coworkers designed and employed a diazirine-containing radiolabeled crosslinking reagent to understand the contact between Meta II rhodopsin and the retinal GTP binding protein transducin.¹⁶⁹ A disulfide in the cross-linking reagent allowed it to make disulfide bonds with the cysteine residues in rhodopsin, followed by irradiation and cross-linking of the diazirine to transducin.^{162, 169} This was followed by cleavage of the disulfide bond to transfer the radiolabel to transducin.¹⁶⁹ In the years following, several diazirine-containing photo-labile reagents were used for cross-linking. For example, In 1998, Girshovich and associates used the diazirine reagent reported by Bochkariov and Kogon to investigate the interactions of heat-shock chaperone protein GroEL with membrane protein SecA, an integral component of translation machinery.¹⁷⁰

Diazirine derivatives of amino acids—specifically photo-methionine and photo-leucine—for probing protein-protein interactions were introduced in 2005 by Theile and coworkers.¹⁵⁹ These novel and unnatural amino acids were subjected to tests for toxicity and function. Addition of photo-leucine or photo methionine to cell cultures did not affect viability (via trypan blue exclusion) or morphology.¹⁵⁹ In addition, the brief irradiation required for photoactivation of the diazirines did not affect cell viability. Furthermore, it was demonstrated that photo-methionine can, at least partially, functionally replace methionine. This was determined by incorporating photo-amino acids into biomolecules *in vivo*, and observing that the fluorescence intensities and localization of 3 separate GFP fused GTPases were unchanged and the function of β -galactosidase was not impaired.¹⁵⁹ In the same publication, Theile and coworkers proved the ability of photo-methionine to capture protein-protein interactions. They elucidated the interactions of membrane

proteins PGRMC1 and SREBP as well as PGRMC1 and Insig-1 using photo-cross-linking followed by immunoprecipitation and western blotting.¹⁵⁹

Since this original proof of concept was reported, photo-methionine has been employed in several other studies. In 2007, Muir and associates used the site-specific incorporation of photo-methionine to capture protein oligomerization of the Smad2 signaling protein.¹⁵⁸ They synthesized Fmoc photo-methionine and used it in a classic solid phase peptide synthesis to give a synthetic peptide that was added to a complimentary protein thioester, yielding a semi-synthetic protein bearing the diazirine functionality.¹⁵⁸ Along with a control sample containing the unmodified protein, protein isolates were irradiated and subjected to western blotting. These experiments revealed cross-linked species with apparent molecular weights corresponding to dimers and trimers of Smad2. These species were barely detectable in the irradiated control and non-irradiated samples.¹⁵⁸ Moreover, their strategic placement of the cross-linker unveiled a transient interaction between MH2 regions of Smad2 that is dependent on post-translational modification.¹⁵⁸ In 2012, Kölbl and coworkers employed photo-methionine and photo-leucine to investigate secondary peptide structure by interrogating intramolecular interactions.¹⁷¹ In this work, peptides containing photo-amino acids were irradiated and digested prior to mass spectrometric analysis. This process produced data that corroborated the proposed β -strand structure of the peptides in question.¹⁷¹ Recently, Lössl and associates employed photo-methionine and photo-leucine as a compliment to other techniques to study the interactions of nidogen-1 and laminin γ 1.¹⁷² Their results were reported in 2014.¹⁷²

Methods

Synthesis and characterization of photo-methionine

Initially, the synthesis of photo-methionine was attempted *via* a modification of the method published by Thiele and coworkers (**Figure 4-5**).¹⁵⁹ 4-acetylbutyric acid (**2**) is allowed to stir in ammonium hydroxide at 0°C before the slow addition of hydroxylamine-O-sulfonic acid (HOSA).¹⁷³ This yielded crude diaziridine **3**, which is concentrated *in vacuo* and resuspended in methanol. Introduction of catalytic triethylamine, followed by slow addition of molecular iodine oxidizes the diaziridine functionality to a diazirine, affording **4**.¹⁷⁴

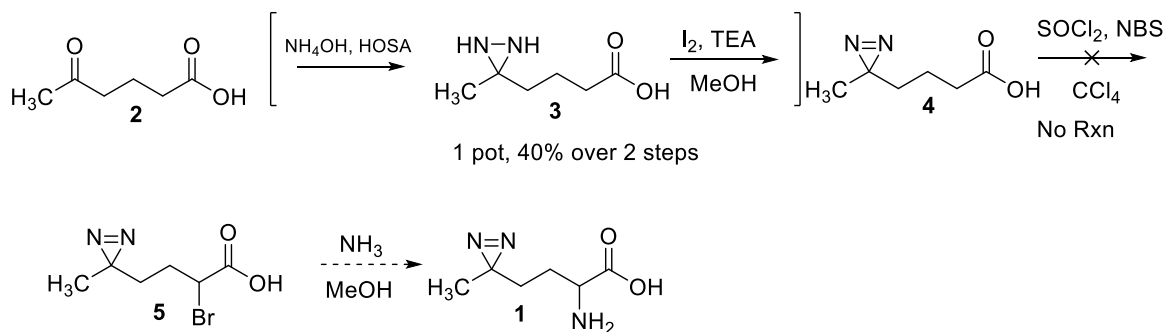
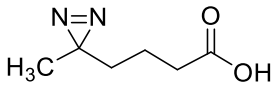
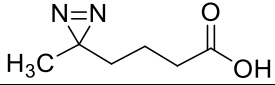
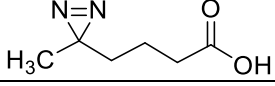
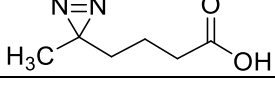
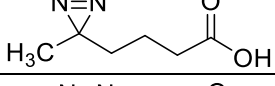
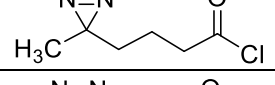
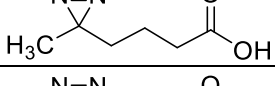
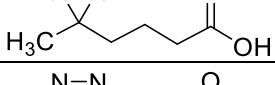
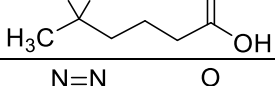
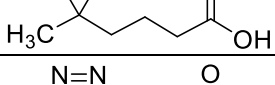
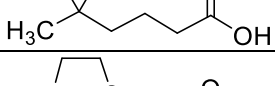
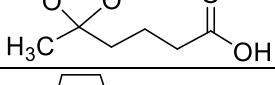
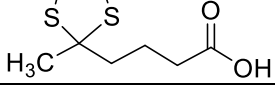
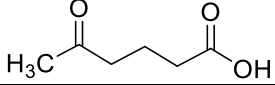
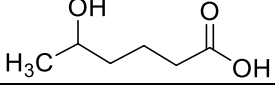


Figure 4-5: Synthesis of photo-methionine by Thiele et. al.

Chromatographic purification of **4** is followed by an attempted α -bromination of the carboxylic acid. This step proved problematic, and work to afford the α -halo acid in acceptable yield was a lengthy but fruitless endeavor. Despite many attempts at reaction optimization (**Table 4-1**), the pure, halogenated product was never achieved. Prolonged stirring of **5** in methanolic ammonia should give photo-methionine in 4 steps, yet this was not completed due to issues with the previous step.

A second possible synthesis for photo-methionine was attempted as a means of circumventing the issues encountered during α -halogenation (**Figure 4-6**). Instead of introducing the diazirine at the beginning, this alternate route installs the labile functionality at the end. This could clearly be advantageous for protection of the labile functional group. This method begins with protected glutamic acid **6**, which is subjected to Weinreb amidation to afford **7**.¹⁵⁸ Methylmagnesium iodide is added to **7**, giving the glutamate-derived ketone **8**.¹⁵⁸ The ketone can then be subjected to ammonium hydroxide and HOSA¹⁷³ or liquid ammonia and HOSA^{158, 174} to yield the diaziridine, which is subsequently oxidized to the corresponding diazirine to give protected photo-methionine (**9**). Deprotection with trifluoroacetic acid (TFA) and HCl was reported to yield **1**.¹⁵⁸ Issues with the diaziridination led to the investigation of an amended route based on this method (**Figure 4-7**). This did not lead to the desired product, as discussed later.

Table 4-1: Attempted alpha halogenation conditions

Starting material	Reagent	Solvent	Temp	Time
	SOCl ₂ , NBS cat. HBr	CCl ₄	50 °C	2 h
	SOCl ₂ , NBS cat. HBr	CCl ₄	65 °C	30 min
	SOCl ₂ , NBS cat. HBr	CCl ₄	70 °C	1.5 h
	SOCl ₂ , NBS cat. HBr	CCl ₄	70 °C (pressurized)	1.5 h
	SOCl ₂ , NBS cat. HBr	CCl ₄	85 °C	1.5 h
	NBS, cat. HBr	CCL ₄	70 °C	1.5 h
	SOCl ₂ , ICl	CCl ₄	70 °C	1.5 h
	1) PBr ₃ , Br ₂ 2) H ₂ O	neat	70-80 °C	10-20 h
	SOCl ₂ , Br ₂	SOCl ₂	50 °C, then RT	ON
	SOCl ₂ , I ₂	SOCl ₂	70 °C	2+ h
	NBS, H ₂ SO ₄	TFA	70-85 °C	16 h
	SOCl ₂ , Br ₂	SOCl ₂	50 °C, then RT	ON
	SOCl ₂ , Br ₂	SOCl ₂	50 °C, then RT	ON
	SOCl ₂ , Br ₂	SOCl ₂	50 °C, then RT	ON
	SOCl ₂ , Br ₂	SOCl ₂	50 °C, then RT	ON

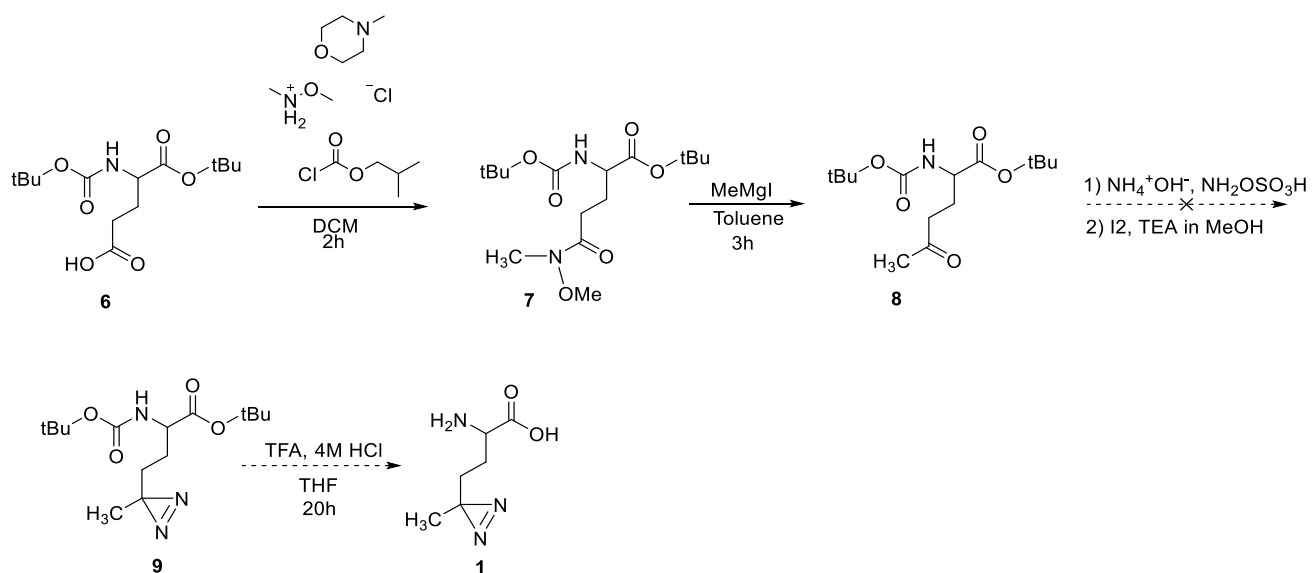


Figure 4-6: Synthesis of photo-methionine by Muir et. al.

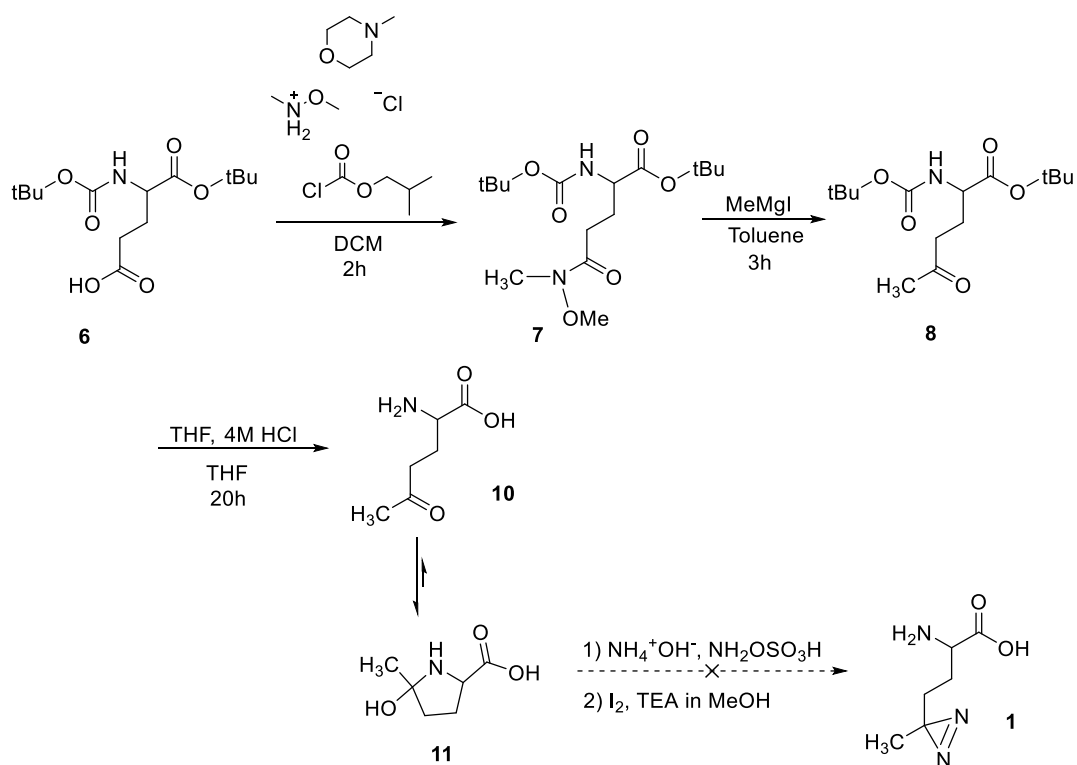


Figure 4-7: Synthesis of photo-methionine adapted from Muir et. al.

1-(*tert*-butyl) 2-methyl 5-oxopyrrolidine-1,2-dicarboxylate (14): The novel synthesis of photo-methionine (**Figure 4-8**) begins with pyroglutamic acid, which was protected as described by Huy-Dinh et al.¹⁷⁵ In short, 50 μL of concentrated hydrochloric acid was added to a solution of pyroglutamic acid (5g) in dry methanol (50 mL) and the mixture was stirred for 24 hours. Concentration *in vacuo* yielded an oil which was sufficiently pure in quantitative yield. Boc anhydride (12.0g), triethylamine (3.62g, 500 μL), and DMAP (0.426g) were added to the resultant methyl ester (5g) in DCM. This mixture was permitted to react for 24 hours at room temperature and extracted with ether. The organic layer was washed with aqueous NaHSO_4 , followed by 1M Na_2CO_3 , and finally brine. The ethereal solution was dried with MgSO_4 , filtered, and concentrated *in vacuo*. The product was purified on silica gel with 80:20 ether/petroleum ether to yield *N*-Boc methyl pyroglutamate (98.8%). ^1H NMR (300 MHz, CDCl_3): δ 1.48 ppm (s,9H), 1.99-2.05 (m, 1H), 2.27-2.35 (m, 1H), 3.77 (s, 3H), 4.60 (dd, 1H).

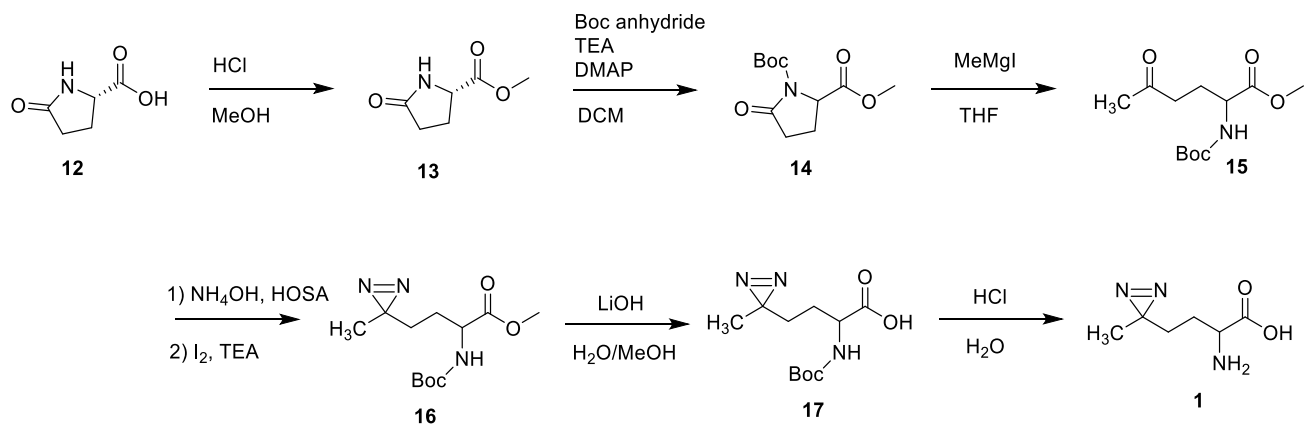


Figure 4-8: Synthesis of photo-methionine by Gibson et. al.

Methyl 2-((*tert*-butoxycarbonyl)amino)-5-oxohexanoate (15): The ketone was formed *via* a Grignard reaction following published conditions.¹⁷⁶ A solution of protected pyroglutamate (14) in anhydrous THF was stirred under an inert atmosphere in an acetonitrile/dry ice bath. MeMgI (1.2eq) in ether was added dropwise. After 2 hours of stirring, the mixture was transferred to the freezer overnight. The reaction was then quenched with ammonium chloride, acidified to pH 2-3, and extracted into EtOAc. The product was purified by flash chromatography with 7:3 hexanes to EtOAc with 1% acetic acid (*r_f* = 0.4, vanillin) to give 66.8% yield of the ketone. ¹H NMR (300 MHz, CDCl₃): δ 1.42 ppm (s, 9H), 1.82-1.90 (m, 1H), 2.07-2.12 (m, 1H), 2.13 (s, 3H), 2.46-2.60 (m, 2H), 3.72 (s, 3H), 4.25 (bs, 1H), 5.09 (bs, 1H).

Methyl 2-((*tert*-butoxycarbonyl)amino)-4-(3-methyl-3H-diazirin-3-yl)butanoate (16): Ketone 15 was dissolved in excess 7N methanolic ammonia and stirred in an ice/salt bath. Hydroxylamine-O-sulfonic acid (1 eq) was added portion wise. The mixture was allowed to stir in the cooling bath for 1hr, then at room temperature for 1 hr. The resulting solution and white precipitate were stored in the refrigerator overnight, after which the mixture was concentrated *in vacuo*, redissolved in MeOH, and filtered. The filtrate was stirred at 0° C with triethylamine (0.25 eq). Iodine crystals were added portion wise until a deep red hue persisted. The product was concentrated *in vacuo* and purified *via* flash chromatography with 85:15 hexanes/EtOAc (*r_f* = 0.3). ¹H NMR (300 MHz, CDCl₃): δ 1.01 ppm (s, 3H), 1.34-1.40 (m, 1H), 1.44 (bs, 12H), 1.68-1.74 (m, 1H), 3.74 (s, 3H), 4.27 (bs, 1H), 4.97 (bs, 1H).

Photo-methionine (1): Molecule 16 (1 eq) was dissolved in MeOH, and LiOH•H₂O (1.5eq) in an equivalent volume of water was added. The mixture was refluxed overnight in the dark. The reaction was neutralized with ammonium chloride, and the product was extracted in EtOAc. Free acid 17 was obtained in high purity, and no further purification was needed. ¹H NMR (300 MHz, CDCl₃): δ 1.03 ppm (s, 3H), 1.44 (bs, 12H), 1.80 (m, 2H), 4.28 (bs, 1H), 4.95 (bs, 1H). The free amino acid was obtained by stirring Boc-photo-methionine in a 50:50 mixture of 4N aqueous HCl:THF at room temperature for 20 hours (in the dark). Evaporation of the solvent gave pure photo-methionine, as determined by

NMR. ¹H NMR (300 MHz, D₂O): δ 0.91 ppm (s, 3H), 1.33-1.51 (m, 2H), 1.64-1.84 (m, 2H), 3.93 (t, 1H).

A successful route to photo-methionine allows synthesis of a diazirine-functionalized ZorO analog. However, this requires Fmoc protection of photo-methionine, which is achieved according to the methods of Muir and coworkers.¹⁵⁸ Specifically, Photo-methionine (246 mg) and NaHCO₃ (207 mg) are dissolved in a minimal amount of water. 12 mL dioxane is added to the mixture. Fmoc-OSu (517 mg) in dioxane was then added over 15 minutes, and the reaction is allowed to proceed in the dark at room temperature for 20 hours. The reaction is then acidified to pH 3 with HCl and extracted with EtOAc. Organic layers were dried with MgSO₄, filtered, and concentrated *in vacuo*. Flash chromatography over silica with 2.5% MeOH in DCM with 1% acetic acid yields Fmoc-photo-methionine.

Synthesis and characterization of ZorO and analogs

Peptide synthesis was achieved manually with a wrist-action shaker and 25 mL vessels with integral filter frit and 3-way stopcock. Initial difficulties were encountered when coupling with 3 equivalents of each successive amino acid. Given the trans-membrane domain of ZorO that features a high content of residues with hydrocarbon side-chains, this difficulty was not startling. Existing literature suggested that unusually hydrophobic peptide sequences are difficult to assemble.¹⁶¹ This is usually due to aggregation of resin-bound chains leading to insufficient solvation. Sometimes these issues with solvation require addition of chaotropic salts to change the polarity of the environment or, in more severe situations, incorporation of amino acids with modified protecting groups every 6-7 residues to prevent aggregation.¹⁶¹ After several failed attempts to construct the desired peptide, success was achieved by modification of standard coupling parameters without chaotropic agents or modifications.

The successful protocol for peptide synthesis entailed purchasing a polystyrene-based Wang resin with the C-terminal Fmoc-protected lysine residue bound. This resin was swollen in minimal amounts of methylene chloride for 10-15 minutes, after which the vessel was drained, and the resin was washed with DMF. Next, excess 20% piperidine in

DMF was added to the resin, the vessel was shaken for 1 minute, and the liquid was drained under aspirator vacuum. This process was repeated with a 15-minute shaking duration. After this two-stage deprotection, the resin was washed 3 times with DMF. Couplings were accomplished by solvating 6 equivalents of the subsequent amino acid 5.9 equivalents of HBTU as a peptide coupling reagent, and 12 equivalents of diisopropylethylamine (DIPEA).¹⁷⁷ After couplings, the resin was washed 3x with DMF. The Kaiser test was used to monitor the success of each coupling. To accomplish this, a few beads of the clean resin were placed in a 1-dram glass vial with 3-4 drops of each of the following 3 solutions: A) 5% ninhydrin in ethanol; B) 80% phenol in ethanol; C) KCN in pyridine (2mL 0.001M KCN in 98 mL pyridine) and heated for 5 minutes at approximately 110° C.¹⁶¹ Reaction of the ninhydrin with free amines produces a blue color (+), indicating incomplete coupling. In the event of a positive Kaiser test, the coupling was repeated until the beads were colorless upon testing. Once the full sequence was assembled via cycles of deprotection and coupling, a final Fmoc deprotection was accomplished to liberate the *N*-terminus of the resin-bound peptide. Finally, the resin was allowed to stir in excess trifluoroacetic acid/water/triisopropyl silane 95:2.5:2.5 for 90-180 minutes.¹⁶¹ The cleavage cocktail was filtered, and the filtrate was concentrated under a stream of nitrogen. Peptide was achieved by dropping the free peptide solution into cold ether (diethyl ether or MTBE), centrifuging, decanting, and lyophilizing from an ACN/H₂O solution.^{161, 177}

The synthetic targets of SPPS are the previously-described ZorO analogs (**Figure 4-9**).

Detection of ZorO by UPLC-HRMS

Measuring expression of ZorO from cell cultures would provide valuable information, especially if the measurement is sensitive enough to monitor chromosomal expression. To this end, a method was developed to quantitate ZorO in fractions of *E. coli* cell lysate. Initially, the synthetic standards of native ZorO were used to validate the method and confirm experimental mass-to-charge ratios and chromatographic retention time. Synthetic ZorO was suspended in a mixture of ACN/H₂O with TFA and diluted to the desired concentrations with PBS.

Native ZorO

H—M—D—T—L—T—Q—K—L—T—V—L—I—A—V—L—E—L—L—V—A—L—L—R—L—I—D—L—L—K—OH

Chemical Formula: C₁₅₂H₂₇₃N₃₅O₄₀S

R23 > K ZorO

H—M—D—T—L—T—Q—K—L—T—V—L—I—A—V—L—E—L—L—V—A—L—L—K—L—I—D—L—L—K—OH


Chemical Formula: C₁₅₂H₂₇₃N₃₃O₄₀S

1X-FLAG ZorO

H—D—Y—K—D—D—D—K—M—D—T—L—T—Q—K—L—T—V—L—I—A—V—L—E—L—L—V—A—L—L—R—L—I—D—L—L—K—OH

Chemical Formula: C₁₉₃H₃₃₁N₄₅O₅₉S

Photo-met ZorO

H——D—T—L—T—Q—K—L—T—V—L—I—A—V—L—E—L—L—V—A—L—L—R—L—I—D—L—L—K—OH

Chemical Formula: C₁₅₃H₂₇₃N₃₇O₄₀

Figure 4-9: ZorO and analogs synthesized with SPPS

Samples were analyzed on a Q Exactive Plus Hybrid Orbitrap equipped with a Dionex UltiMate 3000 with nano-flow capabilities (Thermo Fisher Scientific, Waltham, MA). An Easy-Spray source interfaced the chromatograph with the mass spectrometer. The method featured a 10-minute load onto a C18 PepMap 100 μ -Precolumn (5 μ m, 100 \AA , 300 μ m x 5 mm) with an isocratic flow of 2% ACN in H₂O with 0.1% formic acid at 10.0 μ L per minute. After loading on the precolumn, valve switching allowed the analytes to be back-flushed onto the analytical column. The samples were separated on a PepMap RSLC C18 Easy-Spray column (3 μ M, 100 \AA , 75 μ m x 15 cm) with a solvent system of A) 0.1% formic acid in H₂O; and B) 0.1% formic acid in ACN. The separation gradient was as follows: $t = 0$ min, 100% solvent A, 0% solvent B; $t = 10$ min, 0% solvent A, 100% solvent B; $t = 35$ min, 0% solvent A, 100% solvent B; $t = 45$ min, 0% solvent A, 100% solvent B; $t = 45.1$ min, 100% solvent A, 0% solvent B; $t = 60$ min, 100% solvent A, 0% solvent B. The flow rate was a constant 0.200 μ L per minute.

Spectra were collected in full-scan mode with positive polarity, a scan range of 266.7 - 4000 m/z , and a resolution of 70,000. The AGC target was set to 3e6, and the maximum

IT was 200 ms. For ionization, the sheath gas (N₂) was set to 25, and the aux gas flow was 8. A spray voltage of 4.00 kV with a capillary temp of 320 °C.

Before the validated method was applied to biological extracts, sample preparation was carefully considered. Cell lysates, even when separated into cytoplasmic and membrane fractions, are incredibly complex. Therefore, target proteins are typically enriched or isolated prior to analysis by mass spectrometry. This can be achieved through immunoprecipitation, electrophoresis and excision, affinity chromatography, or other emerging technologies. Many of these approaches are costly and may require specific antibodies or covalent modification of the target protein with a tag. These options are not desirable for studying ZorO, which has unknown function and is diminutive at 3.2 kDa. Covalent modification could drastically influence the properties of this small protein, and the absence of existing ZorO research has not made immunological isolation techniques available. Thus, gel excision was chosen as a method of protein enrichment prior to analysis.

In brief, ZorO was overexpressed in a variant of *E. coli* MG1655 using a ZorO-encoding plasmid and P_{BAD} promoter. Transcription of *zorO* was initiated by addition of arabinose (0.2% maximum) to growing cultures.¹⁷⁸ Cell lysates were separated into cytoplasmic and membrane fractions, and the two protein fractions were subjected to gel electrophoresis using a Tricine gel (containing tris, glycerol, TEMED, ammonium persulfate) designed for optimal resolution of low molecular weight proteins. After electrophoresis, bands were cut from the gel to target the 3 kDa range. Proteins were recovered from the gel using an elution buffer (Tris HCl, NaCl and EDTA).

Results and discussion

Small molecule synthesis

While 2 routes to photo-methionine had been previously described, they entailed their unique difficulties when reproduction of the methods were attempted. The synthetic scheme afforded by the work of Thiele and coworkers was direct but included an alpha-bromination that was unable to be duplicated in any notable yield. This is potentially due to the propensity of the reagents to create radicals, which could initiate a cascade of

reactions to interfere with the diazirine and create a confounding matrix of unwanted products. **Figure 4-10** shows stacked NMR spectra displaying the difficulty achieving the α -bromination. Even in a pressurized vessel, the intended transformation did not appear to occur. Variations of this reaction were attempted many times, and results were mixed. The most promising reaction outcomes (**Figure 4-10, row 5**) were subjected to column chromatography in an attempt to achieve purified product. However, none were fruitful.

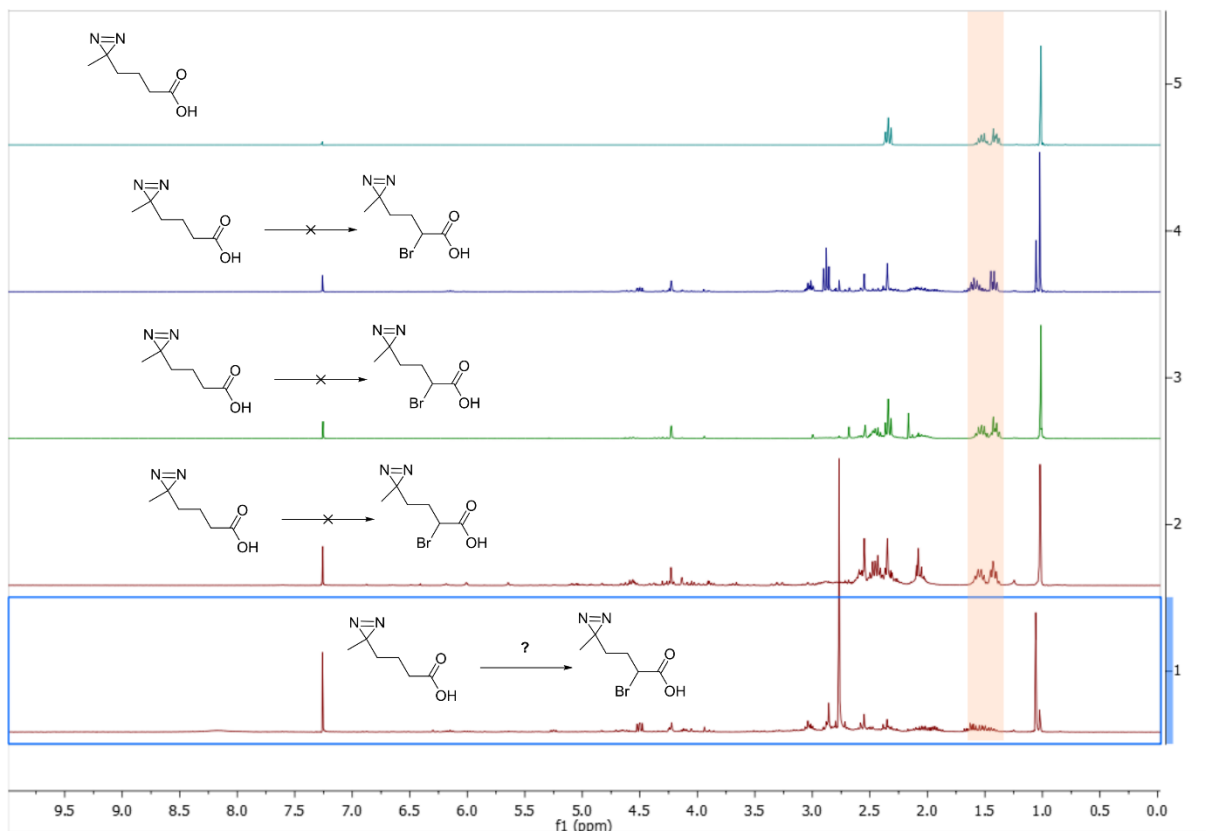


Figure 4-10: ¹H NMR characterization of alpha-bromination attempts

The scheme by Muir and associates was adopted to avoid the halogenation difficulties. This approach was less direct, requiring the synthesis of a Weinreb amide intermediate before installing the ketone handle as a precursor to the diazirine. In addition, the bulky protecting groups selected for the reaction appeared to hinder a detectable conversion of ketone to diazirine. The problem was not remedied by deprotection before diaziridination, likely due to the hemiaminal form of the ketone that is favored by the equilibrium.

Thus, to circumvent all the above issues, a novel synthesis of photo-methionine was conducted. Because it begins with a protected pyroglutamate, the newest route avoids the Weinreb amide intermediate and is more direct than the one reported by Muir et. al. In addition, the use of smaller C-terminal protecting groups enables the diaziridination to occur without steric hindrance. The novel scheme eliminates the need for the halogenation reaction, avoiding the most dubious reaction in the sequence from the work of Thiele et. al. For all these reasons, photo-methionine was afforded in ample yield with technical simplicity using the original method described herein. The merits of this method were realized when the product was achieved with minimal time lost to investigating or optimizing reaction parameters. The success of this novel method led to abandonment of the two previous schemes. With a reliable route to photo-methionine established, current efforts are underway to afford larger amounts of the Fmoc-protected product for use in SPPS.

Peptide synthesis

Solid-phase peptide synthesis is commonly used to assemble products with fewer than 50 residues. Thus, it should pair nicely with the study of small proteins in bacteria, which conventionally have the same size restraint. There is some difficulty; however, with practical synthesis of these small proteins. Most small, independently-encoded proteins act at the membrane, suggesting an enrichment of hydrophobic residues. This is certainly true in the case of ZorO, which has a sequence that is 62.0% composed of residues with hydrocarbon side-chains (Leu, Ile, Ala, Val). These highly hydrophobic peptides are more difficult to synthesize by SPPS, and issues are intensified as chain length increases. While modern automated peptide synthesizers are suitable for high-throughput, many of these longer, hydrophobic sequences are more reliably achieved with manual synthesis.

The advantage of manual synthesis is its modularity. In contrast to automated protocols, manual synthesis includes monitoring the yield of each individual coupling for completeness, providing information about the necessity of recoupling and the location of problematic residues.¹⁷⁷ This both enables the specialized treatment of problematic couplings and drastically decreases the economic and environmental impact as compared to automated syntheses which may include second couplings by default instead of as needed.¹⁷⁷

Native ZorO, bearing the sequence shown in **Figure 4-9**, was achieved using the previously-described protocol and characterized by mass spectrometry. The experimental mass spectrum of synthetic ZorO, in the +2, +3, and +4 charge states, perfectly matched the spectra predicted Yergey Algorithm using IsoPro 3.1.¹⁷⁹ Of these peak clusters, the +3 charge state was the most abundant (**Figure 4-11**).

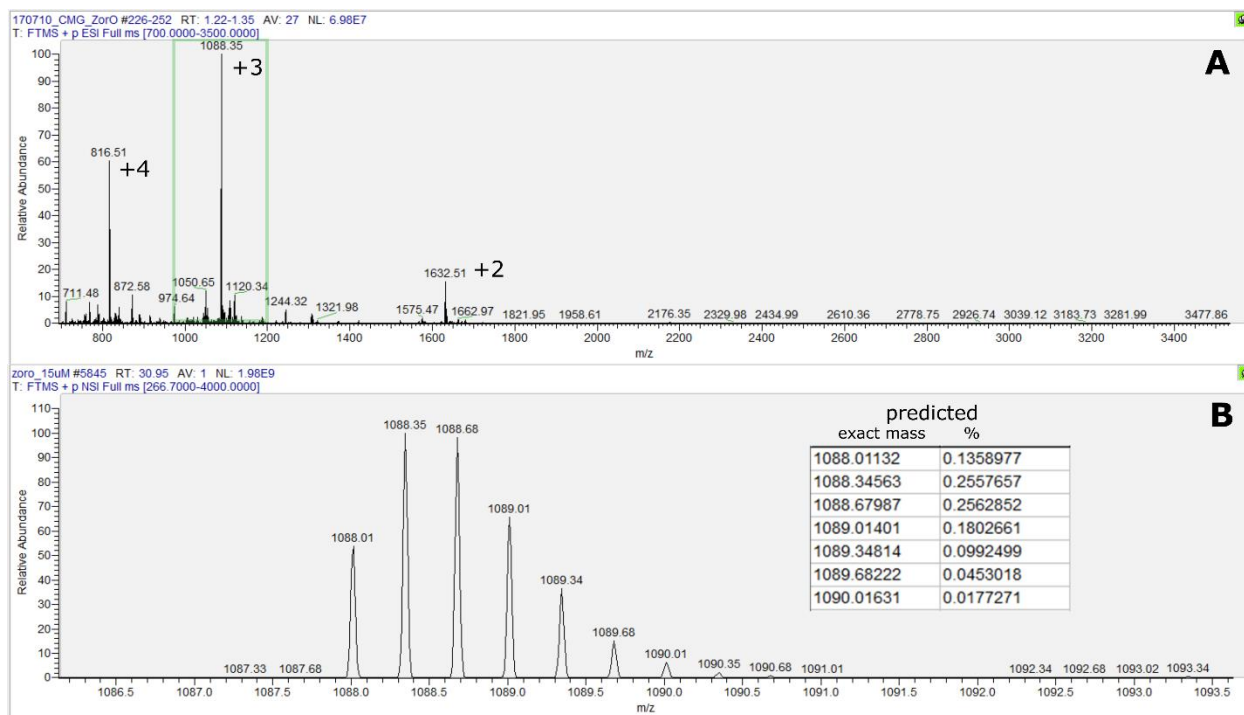


Figure 4-11: MS Characterization of synthetic ZorO

Likewise, the R23K analog was successfully synthesized. However, the benefits of manual synthesis were especially noted when synthesizing the R23K analog of ZorO. Repeated issues were encountered when attempting to couple the L8 residue to the resin-bound chain. Extended coupling times were employed, yet 4 coupling cycles were required to achieve a negative Kaiser test. This care would not have been afforded by automated synthesis, which could not have enabled identification of problematic residues until the completion of the synthesis, wasting resources and opportunity. The mass spectra of synthetic R23K ZorO matched the predicted data (**Figure 4-12**). Again, the +3 charge state was observed to be the most abundant ion. Unlike that of the other synthetic peptides, characterization MS data for 1X-FLAG ZorO did not perfectly match the predicted peaks. This could be independent of synthetic success, and simply related to the altered ionization parameters caused by the addition of the FLAG. As evidenced in the mass spectrum of 1X-FLAG ZorO, a unique mass error is noticed.

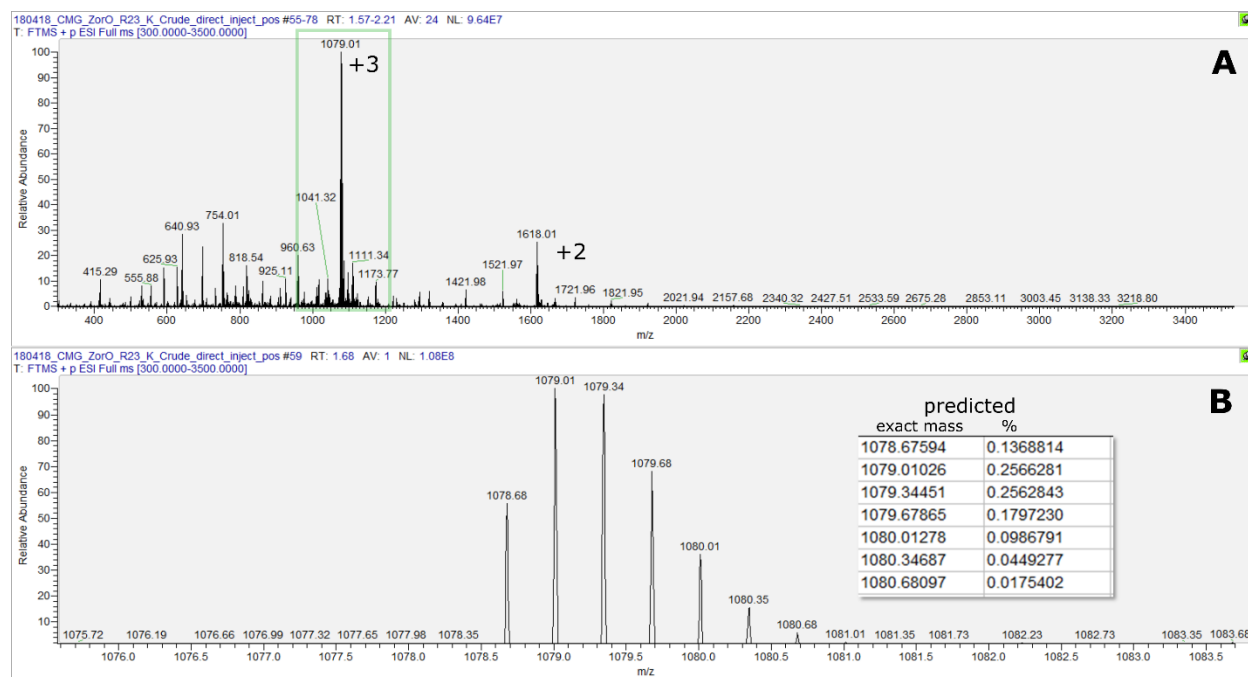


Figure 4-12: MS characterization of synthetic R23K ZorO

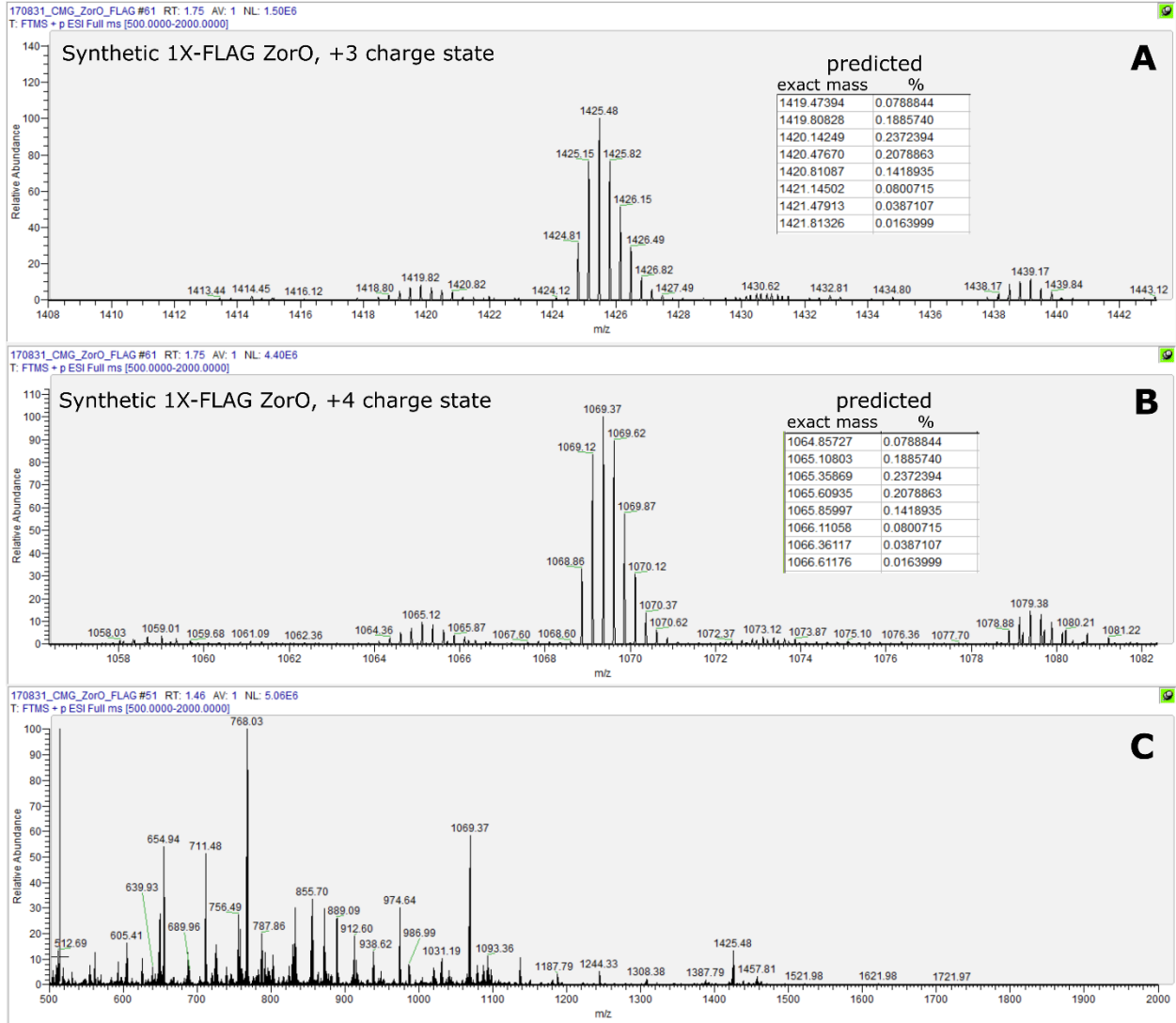


Figure 4-13: MS characterization of 1X-FLAG ZorO

Based on the mass discrepancy of the most abundant ions, the +3 charge state of 1X-FLAG ZorO differs from the expected mass by $17/3$ m/z , while the +4 charge state differs by an error of $17/4$ m/z . This seem to indicate that 1X-FLAG ZorO is forming ammonium adducts in the ionization source.

Alternatively, Issues in SPSS may have arisen with the 8 additional residues either because of a straightforward relationship between chain length and synthetic ease or because of the drastic change in polarity between the hydrophobic ZorO structure and the hydrophilic FLAG sequence. Because of the unclear MS characterization data, the 1X-FLAG ZorO was subjected to detection by electrophoresis and western blotting with FLAG-specific antibodies. As seen in **Figure 4-14**, synthetic FLAG-tagged ZorO runs and stains in a similar fashion to the plasmid-expressed 1X-FLAG ZorO, except the obvious smearing which could be due overloading the gel with synthetic protein. Detection by western blotting indicates that the FLAG head-group of the ZorO analog is intact. However, direct infusion mass spectrometry shows potential byproducts of the peptide synthesis which may be due to sequence deletions for certain subsets of the peptide population. From these data, it becomes evident that 1X-FLAG ZorO must undergo further purification and characterization.

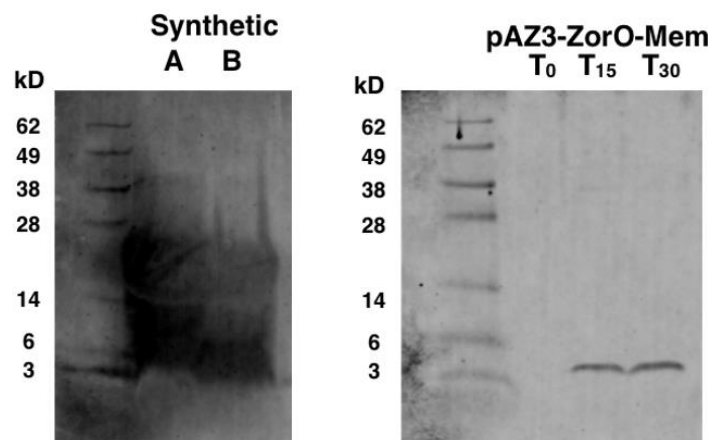


Figure 4-14: Western blot of synthetic 1X-FLAG ZorO and 1X-FLAG ZorO expressed from plasmids

UPLC-HRMS detection

Analysis of the synthetic standard showed that ZorO elutes at 30.0 minutes under the previously-described conditions and that it is detected predominantly in the +3 and +4 charge states (**Figure 4-15**). As a method of confirmation, the isotopologue distributions of both charge states were matched to the simulated isotope clusters predicted by the Yergey Algorithm using IsoPro 3.1.¹⁷⁹

Detection of ZorO from biological sources proved to be a significant undertaking. This is likely related to the innate difficulty of studying individual small proteins from bacteria. Genome annotation of lab-strain *E. coli* reveals over 4000 genes that encode for proteins.¹⁸⁰ Some of these proteins, such as acyl carrier protein, are highly abundant and account for a significant percentage of the cellular protein.¹⁸¹ Other classes of proteins, including many of these small proteins with emerging research focus, regulate their respective biological functions at low concentrations. Given the fact that ZorO is toxic when highly expressed, the natural abundance is surely low.

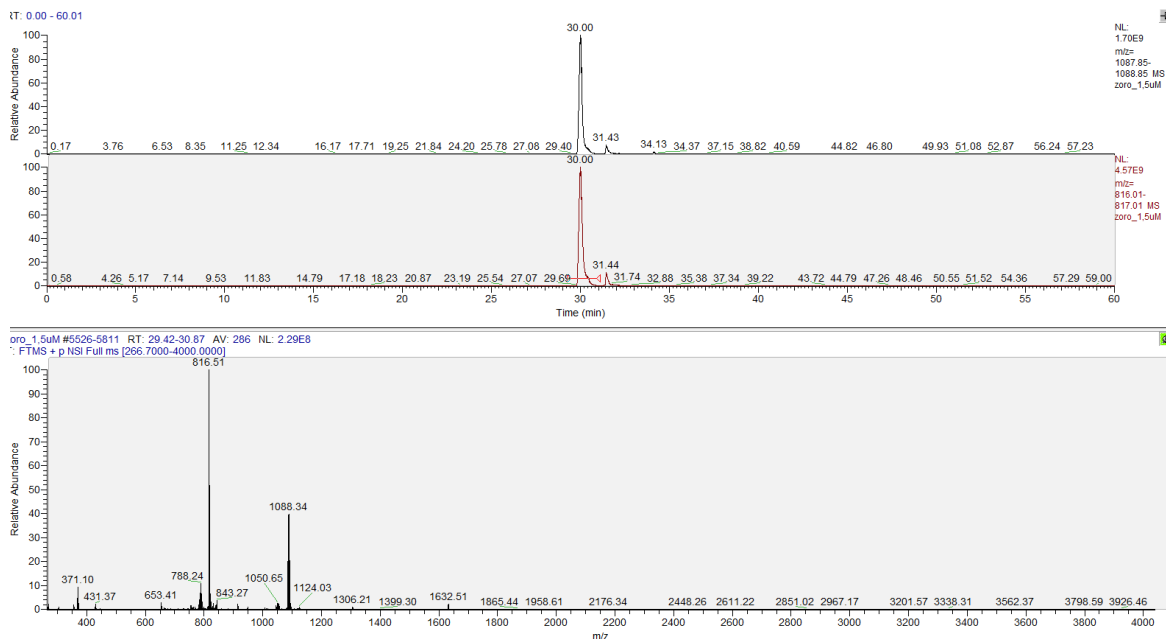


Figure 4-15: ZorO standard analyzed via UPLC-HRMS as a standard

Detecting a single lowly-abundant protein from cell extracts is akin to finding the proverbial needle in the haystack. In addition to the inherent concentration concerns raised by sample complexity, issues of ion suppression, dynamic range limitations, and instrument sensitivity can hinder identification of individual species.

While the nominal mass of the +3 charge state of ZorO was detected in cellular subfractions, the peak was low in intensity and lacked the characteristic isotopic spread of a polypeptide. It is possible that ZorO may be forming adducts in the electrospray source with counterions that are specific to the sample preparation process. Thus, the data is currently being mined to deconvolute the myriad of peaks observed in the chromatograph of gel-excised *E. coli* lysate to identify any possible species of ZorO that differ from the m/z of the synthetic standard.

Conclusions and future work

There are other issues that are specific to an analysis of this nature, as most current biochemical protocols and assays are biased against small proteins. For instance, it is difficult to raise antibodies against a hydrophobic polypeptide the size of ZorO.¹² Moreover, most experiments targeted at understanding interactions of small proteins have included tagging small proteins with tags that surpass their molecular weight.¹² While analogs such as Photo-met ZorO and 1X-FLAG ZorO are promising proxies for studying ZorO's structure and function, there is a significant challenge posed by future biophysical characterization of ZorO's interactions.

These limitations of the field also impact mass spectrometric detection. Enrichment of target proteins prior to analysis is a preferred way to isolate the desired protein analytes. However, for reasons described above, small polypeptides lack a precedence for immunoprecipitation compared to larger proteins. Gel excision was employed in the current study to enrich ZorO prior to UPLC-HRMS analysis. However, it lacks the specificity of immunoaffinity based methods. An added layer of complication comes from the probable low relative abundance of ZorO in cell lysates. Even overexpression of ZorO yielded spectra that were unclear. Only once ZorO detection is achieved at maximal expression can the arabinose can be titrated back to create a biological calibration curve to find an effective limit of detection of ZorO in biological matrices.

While the investigation of ZorO is only beginning, a significant groundwork of synthetic and analytical methods has been laid. A reliable and reproducible route to photo-methionine has been established. Likewise, a method for the manual synthesis of ZorO and analogs has been optimized and employed to create a small library of modified ZorO derivatives for future study. With this optimization, the possibility of synthesizing new analogs arises. For instance, the D26 position of native ZorO may also feature a negative charge that is necessary for the putative oligomerization. In order to investigate this, D26E and D26N analogs of ZorO will be made. Concerning other potential analogs with unnatural amino acids, there remains opportunity. Incorporation of an alkyne-bearing amino acid (for which work is underway) will allow copper-catalyzed “Click” cycloadditions with azides. This provides the opportunity to attach tags for affinity purification or fluorescent labelling using bio-orthogonal chemistry. In addition to continued synthesis, means of probing structure and function must be optimized. This includes biophysical tests for quantifying oligomerization of ZorO mutants relative to the native peptide and characterizing any oligomers that can be captured with the photo-cross-linking of diazirine-functionalized ZorO. In parallel, efforts to allow UPLC-HRMS detection of ZorO from cell matrices will continue.

The field of small protein analysis is evolving, yet still presents challenges to overcome. The techniques and approaches described in this chapter are building blocks for both the hopeful characterization of ZorO’s function and interactions and the advancement of small protein analysis as a discipline.

**CHAPTER FIVE: DESIGN AND INCEPTION OF INTERDISCIPLINARY
COURSEWORK IN UNDERGRADUATE CHEMICAL BIOLOGY
EDUCATION**

Introduction

A final component of this interdisciplinary work focuses on efforts to incorporate interdisciplinarity into higher education. While research endeavors are becoming increasingly collaborative across disciplines, typical undergraduate classes maintain marked boundaries between departments and, therefore, disciplines. While students of science are briefly exposed to the interplay of chemistry and biology in molecular biology courses, there remains important areas that are untouched; namely, the immediate applications of classic chemical techniques towards the understanding of biological systems. With this void in mind, a new laboratory course was designed in which undergraduate chemistry majors would exert the tools of a chemist to explore biological systems. In the design phase, this course had several primary objectives: 1) expose students to every major class of biomolecules; 2) incorporate classical techniques applied directly towards biological questions; 3) feature both qualitative (presence/absence) analyses as well as quantitative; 4) encourage engagement with the scientific literature; and 5) introduce students to the interface of disciplines in a way that highlights the contributions of both fields. To accomplish these goals, a novel 300-level chemical biology laboratory course was designed and implemented. This course relies on organic chemistry (a suggested prerequisite) and analytical chemistry (a suggested corequisite) and is intended to become a core requirement of a chemistry major. By design, this provide a unique experience that is distinguished from molecular biology courses by the heavy emphasis on applied chemistry. In each experiment of the semester-long sequence, continued emphasis is placed on the tools and rationales that are rooted in pure chemistry, and the answers they facilitate when applied to biological systems (**Figure 5-1**).

The central dogma expanded

Most molecular biology courses emphasize the central dogma of molecular biology: genes are transcribed to RNA, which is translated to proteins. This cascade of biological macromolecules is certainly foundational; however, it excludes the diverse array of small molecules (<1500 Da) that are also key components to life at the molecular level.

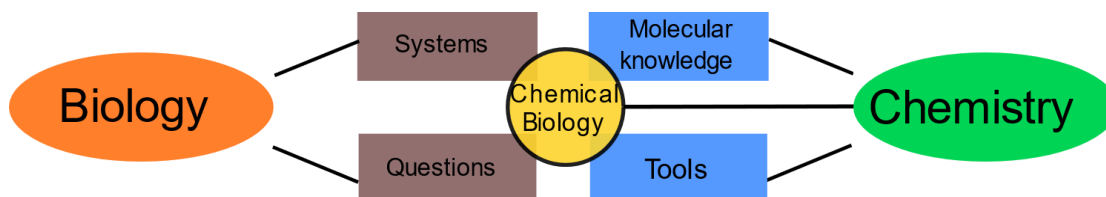


Figure 5-1: Concept map of the course scope: firmly rooted in chemistry to answer biological inquiries

This is the first crossroads at which this novel course diverges from a standard biochemistry lab. While ample attention is given to deoxyribonucleic acids and polypeptides, additional emphasis will bring small molecules into the spotlight. Saccharides, important players in energy storage and structural composition of cells, will be emphasized. Lastly, small lipophilic signaling molecules will be investigated in the final experimental sequence of the semester. This provides the four major classes of biomolecules which will be examined (**Table 5-1**).

A chemical arsenal

This course is designed to include methodologies which are mainstays of chemistry. Techniques for synthesis, separation, and instrumentation are some of the primary contributions that chemists have made to the scientific community. Thus, they are amply represented in the following materials. In addition to metabolite synthesis, solid-phase peptide synthesis (SPPS) will be introduced and employed. Similarly, the essential synthesis-based technique of polymerase chain reaction (PCR) is introduced as an avenue for exploiting natural enzymes to assist with desired syntheses. Both techniques are Nobel-winning innovations that are prime examples of using chemistry to facilitate biology. On the separations front, gel electrophoresis will be taught and implemented for separation of charged biopolymers. Other techniques such as peptide precipitation and liquid-liquid extractions will be incorporated into later experiments. Concerning instrumentation, fluorescence and luminescence spectroscopy will be employed. In addition, students will be introduced to means of sequencing polypeptides using mass spectrometry. These techniques are intentionally paired with the four previously-described classes of molecules to create the basic matrix of course content in **Table 5-1**.

Course outline and schedule

Table 5- 1: Composition of course, highlighting the featured techniques and molecular classes

Biomolecules → Techniques ↓	DNA	Polypeptides	Saccharides	Other small molecules (metabolites)
Reactions and Synthesis	PCR amplification	Solid-phase synthesis	Reactions to probe structure	Biomimetic synthesis of metabolites
Separation and purification	Agarose gel electrophoresis	Precipitation		Liquid-liquid extraction
Spectroscopy / Spectrometry	UV imaging of gels using EtBr	Fluorescence spectroscopy and mass spectrometry	Visual interpretation of changes	NMR characterization, Luminescence spectroscopy
Data analysis and interpretation	Using reference standards	Analyzing MS spectra, plotting spectroscopic data	Connecting observations to molecule structure	Interpreting NMR spectra, plotting spectroscopic data
Applications	Amplifying and detecting genes for antibiotic resistance, including design of specific primers for PCR	Sequencing peptides with tandem mass spectrometry, synthesizing fluorogenic peptides and using them to determine protease specificity	Assessing structural features of sugars by leveraging reactivity, then determining the identity of unknown saccharides	Synthesis of pheromones, using them to modulate luminescence in bacterial colonies, and monitoring effects of concentration

Table 5- 2: Semester schedule. Weeks marked with an asterisk are part of a multi-week sequence. Weeks marked with a double dagger are dry labs. Applicable references are cited for each week

Lab week	Experiment	Agenda
0	Orientation	Complete chemical and biological safety training, introduce chemical biology <i>via</i> instructor, establish goals and expectations for the semester, practice micropipette techniques.
1*†	Genes part 1: PCR design	Design PCR primers given the sequence of target (ampicillin resistance gene). Use internet-based tools to explore the gene sequence, identify possible primers, and evaluate their properties. Extract DNA for use the following week.
2*	Genes part 2: amplification ¹⁸²	Using primers very similar to the ones designed in week one and the DNA extracted previously, the target gene is amplified using polymerase chain reaction.
3*	Genes part 3: detection ¹⁸³	The amplified DNA from week 2 is subjected to gel electrophoresis and visualization with EtBr for observing presence/absence of the ampicillin resistance gene.
4†	Polypeptides: sequencing ¹⁸⁴	In this dry lab, MS/MS spectra of polypeptides are provided. They are interpreted and discussed as a basis for sequencing peptides and identifying peptide modifications.
5*	Polypeptides: synthesis ^{161, 185}	SPPS is introduced, with C-terminal coumarin-modified tripeptides as the synthetic target. Solid-phase synthesis commences.
6*	Polypeptides: Synthesis cont'd	SPPS resumes, allowing the completion of tripeptide synthesis, culminating in cleavage from the resin and isolation of the peptide products.
7*	Polypeptides: proteases ¹⁸⁵	The synthetic fluorogenic peptides are combined with various proteases, and the changes in fluorescence are recorded with a microplate spectrometer to elucidate the specificity of proteolytic activity.
8	Saccharide structure ¹⁸⁶	Known saccharide solutions are assessed with colorimetric reaction-based assays to distinguish sugars based on molecular features. Using these same tests, unknown sugars will be identified.
9*	Chemical signaling part 1 ¹⁸⁷	Acyl homoserine lactones (AHLs), common bacterial signaling molecules, are synthesized using a microwave reactor.
10*	Chemical signaling part 2 ¹⁸⁸	AHL synthesis is completed, and the resulting AHLs are used to modulate luminescence in <i>V. fischeri</i> cultures using luminescence spectroscopy.
11	Lab final/ practical	Using an inquisition-based approach, experiments are proposed (or executed) as an exercise in adapting technologies learned earlier in the semester to new situations.

Methods of evaluation and material reinforcement

As appropriate for a mid-to-upper level science course, this novel chemical biology lab encourages scientific writing. Thus, student evaluation is primarily comprised of formal, typed lab reports for each experiment. These reports consist of the following structure:

Introduction

Background, objectives, biological questions/systems, and chemical tools/understanding

Materials and methods

List of instrumentation, chemicals, glassware, and supplies
Experimental procedure using actual amounts and values

Observations and results

Raw data (graphs, pictures of gels, yields, spectra, etc...)

Discussion and conclusion

Re-introduce the objectives and describe the way they were addressed

Interpret the data and results

Conclude the success of the experiment

In addition to the typed lab reports, concept maps are completed and turned in for every experiment. This serves as a simple way to reinforce the interplay between the chemical tools employed and the driving biological motivation. These concept maps are a single page chart similar in structure to **Figure 5-1** but containing information specific to the present experiment. These will be graded, but ultimately serve to evaluate student understanding and train students to produce a condensed overview of each experiment that preserves the important experimental rationales. Lastly, intermittent questions or in-lab quizzes will be interspersed into the course to continually engage students in between the due dates of formal lab reports.

For each sequence of experiments students will be required to read peer-reviewed publications that establish precedence for the current techniques. These publications are chosen because of their foundational content, historical significance, or scientific precedence. By combining high-level reading comprehension with active learning in the

laboratory, students will mature in their abilities to both digest and produce scientific writing.

Week 11 ends the semester with a culminating assignment. To promote inquiry-based experimental design, students are presented with a biological question that can be answered using techniques and protocols like the ones employed earlier in the semester. In its initial iteration, this experimental design prompt took the form of a written examination, requiring students to propose experiments on paper. While this encouraged students to identify the purpose and explanatory power of the various data generated during the semester, it fell short of its ultimate potential. Thus, the assignment transitioned to presentation-based experiment proposal format. This allows students to work collaboratively with their lab partners to create thorough experimental plans, and propose specific workflows, and defend their rationales. By requiring students to engage their inquiry and creativity, these mock research proposals help immerse upper-level undergraduates in the methods of the scientific community. While a presentation-based assignment serves these goals, an ultimate objective is to facilitate a lab practical in which students can design their own experiments and be supplied to resources to perform them.

Objectives: Learning outcomes

Over the course of the semester, students will apply a myriad of techniques. For the course to function as designed, these experiences are framed in a way that reinforces three central learning outcomes: A) Students will digest interdisciplinary work, learning to identify the predominant contributions of chemistry in each application; B) Guided by principles of metacognition, students will develop a mature comprehension of experimental design rationales; and C) Drawing from specific model cases, students will learn to creatively propose biological applications of chemical techniques. The final of these three objectives is a culminating outcome that draws from the former two.

To keep these outcomes central, instruction for this course should continually reiterate the ways that chemistry has facilitated explorations in molecular biology. For instance, while PCR is widely considered to be a tool of molecular biologists, it is fundamentally based in chemistry. Emphasizing that PCR is ultimately a means of chemical synthesis

will encourage appreciation for the historical contributions of chemistry and highlighting the ways it has enabled answers will foster an appreciation for interdisciplinarity. These themes are reinforced throughout the semester with the introduction of each new technique.

Course material

A comprehensive collection of course material for this semester-long sequence was produced *de novo*, with significant inspiration from the scientific literature. Lab handouts were written for each experiment and assembled to yield an in-house lab manual. In addition, pre-lab lecture materials were conceived and constructed. All of this material is contained in **Appendix D**.

Discussion

Selected experiments from this sequence warrant further discussion to understand their feasibility, cost, benefit, and novelty. Most notable are the two multi-week sequences that use solid-phase synthesis and culminate with microplate assays.

The peptide synthesis and protease specificity sequence relies on the ability to synthesize ACC-modified peptide substrates beginning with an ACC-functionalized resin. This resin, while commercially available from limited suppliers, is quite expensive. It can be synthesized in larger quantities for reduced cost. However, this requires extensive time cost and considerable synthetic skill. This should be noted, as maintaining an internally-supplied source of the ACC-functionalized Rinke Amide resin presents opportunity costs that should be evaluated. Either cost is outweighed by the benefit. By performing this multi-week experiment, students are introduced to SPPS, which has both historical and technical significance. Also, it allows for a single polypeptide experiment which explores structure and function of both small tripeptides and entire enzymes

Initially, the chemical signaling sequence culminated with observing the quorum sensing response to AHLs using biosensor strains of *Chromobacterium violaceum* and/or *Agrobacterium tumefaciens* in petri dish-based assays. These approaches relied on the observable presence or absence of color in solid agar. While these assays have been

previously incorporated into an undergraduate educational laboratory, they failed to give reliable results in the pilot semesters of the currently-described course. More importantly, they yield only qualitative data from which presence or absence of AHLs can be inferred. To streamline the technical workflow and allow for the generation of numerical data, these petri dish-based assays are being replaced with a quantitative quorum-induced luminescence assay in *Vibrio fischeri* DC22. The preparation of the DC22 biosensor assay is more straightforward, necessitating only liquid cultures. In addition, it replaces visual observation with chemical instrumentation and result in quantitative data which can be accessed via statistical analyses. For these three reasons, it is preferred over the petri dish-based assays.

Most of the experiments in this sequence were originally conceived, drawing from collaborative experience and adapting protocols from research literature. This facilitates the desired innovation level of the course and can be especially recognized in the case of the tandem synthesis/assay experiments. However, the carbohydrate analysis laboratory is more straightforward. It relies on well-established colorimetric tests that are commonly used in undergraduate laboratories. For instance, Florida Institute of Technology (Melbourne, Florida), Boğaziçi University (Istanbul, Turkey), Amrita Vishwa Vidyapeetham (Coimbatore, India), and King Saud University (Riyadh, Saudi Arabia) all employ very similar experiments in their curriculum. These institutions have made their lab procedures available online, which provided the template for the carbohydrate lab for the present sequence. While this unit may lack the innovation and adaption of research literature that typifies the course as a whole, it affords a well-rounded sequence by the insertion of these straightforward qualitative tests between synthesis and instrumentation-heavy modules. Nonetheless, this presents an avenue for continued course evolution to incorporate higher-level problem solving into the carbohydrate analysis unit.

Dry labs were conceived with intentionality, exposing students to experimental design (in the case of week 1) and data analysis (in the case of week 4). These are vital components of successful laboratory science, which can be easily overlooked in technique-driven courses. By incorporating two dry labs into a sequence that otherwise relies on bench work, students are familiarized with all phases of experimentation. This provides a more

holistic experience, breeding a familiarity that should equate to improved research preparedness.

Outlook

In summary, a semester-long laboratory course was designed with the intent to explore applications of chemical methodologies while simultaneously introducing undergraduates to the biological rationales that often facilitate interdisciplinary science. The conception of curriculum for this course was deliberate, intended to create a cohesive set of laboratory sessions that introduces students to 4 major classes of biomolecules, encourages them to understand biological rationales, and reinforces the value of prominent chemical techniques.

Many undergraduate chemistry majors gravitate either towards mathematics, with limited exposure to the life sciences, or towards the health sciences, with plans to attend health professional schools following graduation. This course serves the chemistry program in two regards, by the intentional incorporation of biology that makes the mathematically-oriented chemistry majors increasingly well-rounded and simultaneously bridges the gaps between the physical and life sciences for those who intend to pursue health professions with a background in chemistry.

This course was piloted for 3 semesters and is undergoing optimization to accommodate larger class sizes and become a core component of the chemistry program at the University of Tennessee. With the completion of this foundational work, opportunities for course evolution are facilitated. Efforts will continue to make this sequence cohesive and highlight the complementarity. Eventually, a multi-faceted lab practical will be incorporated into this curriculum, allowing the semester to culminate in a way that challenges students to engage in creative and wholistic thinking to investigate environmental samples using the techniques fortified by the existing modules.

CHAPTER SIX: CONCLUSIONS

In conclusion, a breadth of experience in chemical methodology was applied to facilitate interdisciplinary science and improve interdisciplinary education. Bridging analytical chemistry and environmental science, mass spectrometry was used in conjunction with biological metadata to demonstrate that microbial communities without previous exposure to hydraulic fracturing removed the fracturing biocide glutaraldehyde from the environment more rapidly than previously unexposed communities. In a separate vignette, synthesis and mass spectrometry were used to institute and validate a method of measuring *N*-acyl amino acids (*N*-AAA) from biological samples. Subsequently, it was demonstrated that *N*-AAA expression is highly correlated to lipolysis *via* fasting and regulated by dietary fat composition. This allowed the generation of new hypotheses about molecular health benefits of fasting and diet. Later, in a synthesis-heavy endeavor, the toxic small protein ZorO was investigated. These efforts were slowed by methodological setbacks, but ultimately yielded several tools to enable the future study of ZorO and other small proteins, including a novel and improved synthesis of photo-methionine. Lastly, an original and innovative laboratory course was designed and implemented with the intent to encourage interdisciplinarity in undergraduate education.

WORKS CITED

1. Wagner, C. S.; Roessner, J. D.; Bobb, K.; Klein, J. T.; Boyack, K. W.; Keyton, J.; Rafols, I.; Börner, K., Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics* **2011**, *5*, 14-26.
2. Van Noorden, R., Interdisciplinary research by the numbers. *Nature* 17 September 2015, 2015, pp 306-307.
3. Montgomery, C. T.; Smith, M. B., Hydraulic Fracturing: History of an Enduring Technology. *SPE-1210-0026-JPT* **2010**, *62*, 26-40.
4. Kahrilas, G. A.; Blotevogel, J.; Stewart, P. S.; Borch, T., Biocides in Hydraulic Fracturing Fluids: A Critical Review of Their Usage, Mobility, Degradation, and Toxicity. *Environmental Science & Technology* **2015**, *49*, 16-32.
5. McDonnell, G.; Russell, A. D., Antiseptics and disinfectants: activity, action, and resistance. *Clinical microbiology reviews* **1999**, *12*, 147-179.
6. Childress, E. S.; Alexopoulos, S. J.; Hoehn, K. L.; Santos, W. L., Small Molecule Mitochondrial Uncouplers and Their Therapeutic Potential. *Journal of Medicinal Chemistry* **2018**, *61*, 4641-4655.
7. Terada, H., Uncouplers of oxidative phosphorylation. *Environmental health perspectives* **1990**, *87*, 213-218.
8. Fan, W.; Evans, R., The quest to burn fat, effortlessly and safely. *Science* **2016**, *353*, 749-750.
9. Rousset, S.; Alves-Guerra, M.-C.; Mozo, J.; Miroux, B.; Cassard-Doulcier, A.-M.; Bouillaud, F.; Ricquier, D., The Biology of Mitochondrial Uncoupling Proteins. *Diabetes* **2004**, *53*, S130-S135.
10. Wu, J.; Cohen, P.; Spiegelman, B. M., Adaptive thermogenesis in adipocytes: Is beige the new brown? *Genes & Development* **2013**, *27*, 234-250.
11. Long, J. Z.; Svensson, K. J.; Bateman, L. A.; Lin, H.; Kamenecka, T.; Lokurkar, I. A.; Lou, J.; Rao, R. R.; Chang, M. R.; Jedrychowski, M. P.; Paulo, J. A.; Gygi, S. P.; Griffin, P. R.; Nomura, D. K.; Spiegelman, B. M., The Secreted Enzyme PM20D1 Regulates Lipidated Amino Acid Uncouplers of Mitochondria. *Cell* **2016**, *166*, 424-435.
12. Storz, G.; Wolf, Y. I.; Ramamurthi, K. S., Small proteins can no longer be ignored. *Annual review of biochemistry* **2014**, *83*, 753-777.

13. Fozo, E. M., New type I toxin-antitoxin families from "wild" and laboratory strains of E-coli Ibs-Sib, ShoB-OhsC and Zor-Orz. *RNA Bio.* **2012**, *9*, 1504-1512.
14. Wen, J.; Won, D.; Fozo, E. M., The ZorO-OrzO type I toxin-antitoxin locus: repression by the OrzO antitoxin. *Nucleic Acids Res.* **2014**, *42*, 1930-1946.
15. Fontaine, F.; Fuchs, R. T.; Storz, G., Membrane Localization of Small Proteins in Escherichia coli. *J. Biol. Chem.* **2011**, *286*, 32464-32474.
16. Malakoff, D., The gas surge. *Science* **2014**, *344*, 1464-1467.
17. EIA, U. S. *Drilling Productivity Report For Key Tight Oil and Shale Gas REgions*; 2017.
18. Stringfellow, W. T.; Domen, J. K.; Camarillo, M. K.; Sandelin, W. L.; Borglin, S., Physical, chemical, and biological characteristics of compounds used in hydraulic fracturing. *Journal of hazardous materials* **2014**, *275*, 37-54.
19. Kahrilas, G. A.; Blotevogel, J.; Stewart, P. S.; Borch, T., Biocides in hydraulic fracturing fluids: A critical review of their usage, mobility, degradation, and toxicity. *Environmental science & technology* **2014**, *49*, 16-32.
20. Cluff, M. A.; Hartsock, A.; MacRae, J. D.; Carter, K.; Mouser, P. J., Temporal changes in microbial ecology and geochemistry in produced water from hydraulically fractured Marcellus Shale gas wells. *Environmental science & technology* **2014**, *48*, 6508-6517.
21. Struchtemeyer, C. G.; Elshahed, M. S., Bacterial communities associated with hydraulic fracturing fluids in thermogenic natural gas wells in North Central Texas, USA. *FEMS Microbiology Ecology* **2012**, *81*, 13-25.
22. Liang, R.; Davidova, I. A.; Marks, C. R.; Stamps, B. W.; Harriman, B. H.; Stevenson, B. S.; Duncan, K. E.; Suflita, J. M., Metabolic Capability of a Predominant Halanaerobium sp. in Hydraulically Fractured Gas Wells and Its Implication in Pipeline Corrosion. *Frontiers in microbiology* **2016**, *7*.
23. Mouser, P. J.; Borton, M.; Darrah, T. H.; Hartsock, A.; Wrighton, K. C., Hydraulic fracturing offers view of microbial life in the deep terrestrial subsurface. *FEMS Microbiology Ecology* **2016**.

24. Murali Mohan, A.; Hartsock, A.; Hammack, R. W.; Vidic, R. D.; Gregory, K. B., Microbial communities in flowback water impoundments from hydraulic fracturing for recovery of shale gas. *FEMS Microbiology Ecology* **2013**, *86*, 567-580.
25. Vikram, A.; Lipus, D.; Bibby, K., Produced water exposure alters bacterial response to biocides. *Environmental science & technology* **2014**, *48*, 13001-13009.
26. Akob, D. M.; Cozzarelli, I. M.; Dunlap, D. S.; Rowan, E. L.; Lorah, M. M., Organic and inorganic composition and microbiology of produced waters from Pennsylvania shale gas wells. *Appl Geochem* **2015**, *60*, 116-125.
27. Leung, H. W., Aerobic and anaerobic metabolism of glutaraldehyde in a river water-sediment system. *Arch Environ Contam Toxicol* **2001**, *41*, 267-73.
28. Leung, H.-W., Ecotoxicology of glutaraldehyde: review of environmental fate and effects studies. *Ecotoxicology and environmental safety* **2001**, *49*, 26-39.
29. McLaughlin, M. C.; Borch, T.; Blotvogel, J., Spills of Hydraulic Fracturing Chemicals on Agricultural Topsoil: Biodegradation, Sorption, and Co-contaminant Interactions. *Environ Sci Technol* **2016**, *50*, 6071-8.
30. Rogers, J. D.; Ferrer, I.; Tummings, S. S.; Bielefeldt, A. R.; Ryan, J. N., Inhibition of Biodegradation of Hydraulic Fracturing Compounds by Glutaraldehyde: Groundwater Column and Microcosm Experiments. *Environmental Science & Technology* **2017**, *51*, 10251-10261.
31. Kahrilas, G. A.; Blotvogel, J.; Corrin, E. R.; Borch, T., Downhole Transformation of the Hydraulic Fracturing Fluid Biocide Glutaraldehyde: Implications for Flowback and Produced Water Quality. *Environmental Science & Technology* **2016**, *50*, 11414-11423.
32. Sano, L. L.; Krueger, A. M.; Landrum, P. F., Chronic toxicity of glutaraldehyde: differential sensitivity of three freshwater organisms. *Aquatic toxicology* **2005**, *71*, 283-96.
33. Levis, E., Texas Company Pays \$93,710 settlement for polluting Clearfield County Creek. Pennsylvania Fish and Boat Commission. *Pennsylvania Fish and Boat Commission* 2016.
34. PADEP *Oil and Gas Annual Report*, Pennsylvania Department of Environmental Protection Office of Oil and Gas Management 2013.

35. Commission, G. P. C. I. O. G. C. FracFocus Chemical Disclosure Registry. <https://fracfocusdata.org/DisclosureSearch/Search.aspx> (March 16),
36. Trexler, R.; Solomon, C.; Brislawn, C. J.; Wright, J. R.; Rosenberger, A.; McClure, E. E.; Grube, A. M.; Peterson, M. P.; Keddache, M.; Mason, O. U.; Hazen, T. C.; Grant, C. J.; Lamendella, R., Assessing impacts of unconventional natural gas extraction on microbial communities in headwater stream ecosystems in Northwestern Pennsylvania. *Frontiers in microbiology* **2014**, *5*, 522.
37. Grant, C. J.; Lutz, A. K.; Kulig, A. D.; Stanton, M. R., Fracked ecology: Response of aquatic trophic structure and mercury biomagnification dynamics in the Marcellus Shale Formation. *Ecotoxicology* **2016**, *25*, 1739-1750.
38. Lutz, A. K.; Grant, C. J., Impacts of hydraulic fracturing development on macroinvertebrate biodiversity and gill morphology of net-spinning caddisfly (Hydropsychidae, Diplectrona) in northwestern Pennsylvania, USA. *Journal of Freshwater Ecology* **2016**, *31*, 211-217.
39. Ulrich, N. K., Veronica; Drucker, Rebecca; Wright, Justin; McLimas, Christopher; Hazen, Terry; Campa, Maria F.; Grant, Christopher; Lamendella, Regina, Response of Aquatic Bacterial Communities to Hydraulic Fracturing in Northwestern Pennsylvania: A Five-Year Study. *Scientific Reports* **In Review**.
40. Lauer, N. E.; Harkness, J. S.; Vengosh, A., Brine Spills Associated with Unconventional Oil Development in North Dakota. *Environ Sci Technol* **2016**, *50*, 5389-97.
41. Grant, C. J.; Weimer, A. B.; Marks, N. K.; Perow, E. S.; Oster, J. M.; Brubaker, K. M.; Trexler, R. V.; Solomon, C. M.; Lamendella, R., Marcellus and mercury: Assessing potential impacts of unconventional natural gas extraction on aquatic ecosystems in northwestern Pennsylvania. *J Environ Sci Health A Tox Hazard Subst Environ Eng* **2015**, *50*, 482-500.
42. Dow Delivering Value with an Optimized Microbial Control Program in Oil and Gas Operations. http://msdssearch.dow.com/PublishedLiteratureDOWCOM/dh_097e/0901b8038097eef2.pdf?filepath=microbial/pdfs/noreg/253-02697.pdf&fromPage=GetDoc (April 30),

43. Stough, J. M.; Dearth, S. P.; Denny, J. E.; LeCleir, G. R.; Schmidt, N. W.; Campagna, S. R.; Wilhelm, S. W., Functional Characteristics of the Gut Microbiome in C57BL/6 Mice Differentially Susceptible to *Plasmodium yoelii*. *Frontiers in microbiology* **2016**, *7*, 1520.
44. Lu, W.; Clasquin, M. F.; Melamud, E.; Amador-Noguez, D.; Caudy, A. A.; Rabinowitz, J. D., Metabolomic analysis via reversed-phase ion-pairing liquid chromatography coupled to a stand alone orbitrap mass spectrometer. *Analytical chemistry* **2010**, *82*, 3212-21.
45. Smith, M. B.; Rocha, A. M.; Smillie, C. S.; Olesen, S. W.; Paradis, C.; Wu, L. Y.; Campbell, J. H.; Fortney, J. L.; Mehlhorn, T. L.; Lowe, K. A.; Earles, J. E.; Phillips, J.; Techtmann, S. M.; Joyner, D. C.; Elias, D. A.; Bailey, K. L.; Hurt, R. A.; Preheim, S. P.; Sanders, M. C.; Yang, J.; Mueller, M. A.; Brooks, S.; Watson, D. B.; Zhang, P.; He, Z. L.; Dubinsky, E. A.; Adams, P. D.; Arkin, A. P.; Fields, M. W.; Zhou, J. Z.; Alm, E. J.; Hazen, T. C., Natural Bacterial Communities Serve as Quantitative Geochemical Biosensors. *Mbio* **2015**, *6*.
46. Caporaso, J. G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F. D.; Costello, E. K.; Fierer, N.; Pena, A. G.; Goodrich, J. K.; Gordon, J. I.; Huttley, G. A.; Kelley, S. T.; Knights, D.; Koenig, J. E.; Ley, R. E.; Lozupone, C. A.; McDonald, D.; Muegge, B. D.; Pirrung, M.; Reeder, J.; Sevinsky, J. R.; Turnbaugh, P. J.; Walters, W. A.; Widmann, J.; Yatsunencko, T.; Zaneveld, J.; Knight, R., QIIME allows analysis of high-throughput community sequencing data. *Nature methods* **2010**, *7*, 335-6.
47. Ritalahti, K. M.; Amos, B. K.; Sung, Y.; Wu, Q.; Koenigsberg, S. S.; Loffler, F. E., Quantitative PCR targeting 16S rRNA and reductive dehalogenase genes simultaneously monitors multiple Dehalococcoides strains. *Applied and environmental microbiology* **2006**, *72*, 2765-74.
48. Harms, G.; Layton, A. C.; Dionisi, H. M.; Gregory, I. R.; Garrett, V. M.; Hawkins, S. A.; Robinson, K. G.; Saylor, G. S., Real-time PCR quantification of nitrifying bacteria in a municipal wastewater treatment plant. *Environ Sci Technol* **2003**, *37*, 343-51.
49. McMurdie, P. J.; Holmes, S., phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one* **2013**, *8*, e61217.

50. Team, R. C. R: A language and environment for statistical computing. .
<http://www.R-project.org/>
51. Aronesty, E. *ea-utils: "Command-line tools for processing biological sequencing data"*, 2011.
52. Edgar, R. C., Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **2010**, *26*, 2460-1.
53. Edgar, R. C.; Haas, B. J.; Clemente, J. C.; Quince, C.; Knight, R., UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **2011**, *27*, 2194-200.
54. DeSantis, T. Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E. L.; Keller, K.; Huber, T.; Dalevi, D.; Hu, P.; Andersen, G. L., Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology* **2006**, *72*, 5069-72.
55. Caporaso, J. G.; Bittinger, K.; Bushman, F. D.; DeSantis, T. Z.; Andersen, G. L.; Knight, R., PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **2010**, *26*, 266-7.
56. Wang, Q.; Garrity, G. M.; Tiedje, J. M.; Cole, J. R., Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and environmental microbiology* **2007**, *73*, 5261-7.
57. McDonald, D.; Price, M. N.; Goodrich, J.; Nawrocki, E. P.; DeSantis, T. Z.; Probst, A.; Andersen, G. L.; Knight, R.; Hugenholtz, P., An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME journal* **2012**, *6*, 610-8.
58. Werner, J. J.; Koren, O.; Hugenholtz, P.; DeSantis, T. Z.; Walters, W. A.; Caporaso, J. G.; Angenent, L. T.; Knight, R.; Ley, R. E., Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys. *The ISME journal* **2012**, *6*, 94-103.
59. Pinheiro, J. B., D.; DebRoy, S.; Sarkar, D.; and R Core Team nlme: Linear and Nonlinear Mixed Effects Models. *R package version 3.1-131.1* **2018**.
60. Lenth, R. L., J.; Herve, M., Estimated Marginal Means, aka Least-Square Means. *R package version 1.1.2* **2018**.

61. Love, M. I.; Huber, W.; Anders, S., Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **2014**, *15*, 550.
62. Lozupone, C.; Lladser, M. E.; Knights, D.; Stombaugh, J.; Knight, R., UniFrac: an effective distance metric for microbial community comparison. *The ISME journal* **2011**, *5*, 169-172.
63. Dixon, P., VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* **2003**, *14*, 927-930.
64. Grant, C. J.; Weimer, A. B.; Marks, N. K.; Perow, E. S.; Oster, J. M.; Brubaker, K. M.; Trexler, R. V.; Solomon, C. M.; Lamendella, R., Marcellus and mercury: assessing potential impacts of unconventional natural gas extraction on aquatic ecosystems in northwestern Pennsylvania. *Journal of Environmental Science and Health, Part A* **2015**, *50*, 482-500.
65. Akob, D. M.; Mumford, A. C.; Orem, W.; Engle, M. A.; Klinges, J. G.; Kent, D. B.; Cozzarelli, I. M., Wastewater Disposal from Unconventional Oil and Gas Development Degrades Stream Quality at a West Virginia Injection Facility. *Environmental Science & Technology* **2016**, *50*, 5517-5525.
66. Ferrer, I.; Thurman, E. M., Analysis of hydraulic fracturing additives by LC/Q-TOF-MS. *Analytical and bioanalytical chemistry* **2015**, *407*, 6417-28.
67. Shade, A.; Peter, H.; Allison, S. D.; Baho, D. L.; Berga, M.; Burgmann, H.; Huber, D. H.; Langenheder, S.; Lennon, J. T.; Martiny, J. B.; Matulich, K. L.; Schmidt, T. M.; Handelsman, J., Fundamentals of microbial community resistance and resilience. *Front Microbiol* **2012**, *3*, 417.
68. Techtmann, S. M.; Zhuang, M.; Campo, P.; Holder, E.; Elk, M.; Hazen, T. C.; Conmy, R.; Santo Domingo, J. W., Corexit 9500 Enhances Oil Biodegradation and Changes Active Bacterial Community Structure of Oil-Enriched Microcosms. *Applied and Environmental Microbiology* **2017**, *83*, e03462-16.
69. Rojo, F., Degradation of alkanes by bacteria. *Environmental microbiology* **2009**, *11*, 2477-90.
70. Schneiker, S.; Martins dos Santos, V. A.; Bartels, D.; Bekel, T.; Brecht, M.; Buhrmester, J.; Chernikova, T. N.; Denaro, R.; Ferrer, M.; Gertler, C.; Goesmann, A.; Golyshina, O. V.; Kaminski, F.; Khachane, A. N.; Lang, S.; Linke, B.; McHardy, A. C.;

Meyer, F.; Nechitaylo, T.; Puhler, A.; Regenhardt, D.; Rupp, O.; Sabirova, J. S.; Selbitschka, W.; Yakimov, M. M.; Timmis, K. N.; Vorholter, F. J.; Weidner, S.; Kaiser, O.; Golyshin, P. N., Genome sequence of the ubiquitous hydrocarbon-degrading marine bacterium *Alcanivorax borkumensis*. *Nature biotechnology* **2006**, *24*, 997-1004.

71. Bolla, J. R.; Su, C. C.; Delmar, J. A.; Radhakrishnan, A.; Kumar, N.; Chou, T. H.; Long, F.; Rajashankar, K. R.; Yu, E. W., Crystal structure of the *Alcanivorax borkumensis* YdaH transporter reveals an unusual topology. *Nat Commun* **2015**, *6*, 6874.

72. Nogales, B.; Lanfranconi, M. P.; Pina-Villalonga, J. M.; Bosch, R., Anthropogenic perturbations in marine microbial communities. *FEMS microbiology reviews* **2011**, *35*, 275-98.

73. López-Pérez, M.; Rodríguez-Valera, F., The Family Alteromonadaceae. In *The Prokaryotes: Gammaproteobacteria*, Rosenberg, E.; DeLong, E. F.; Lory, S.; Stackebrandt, E.; Thompson, F., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 2014; pp 69-92.

74. Ivanova, E. P.; Ng, H. J.; Webb, H. K., The Family Pseudoalteromonadaceae. In *The Prokaryotes: Gammaproteobacteria*, Rosenberg, E.; DeLong, E. F.; Lory, S.; Stackebrandt, E.; Thompson, F., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 2014; pp 575-582.

75. Silveira, C. B.; Thompson, F., The Family Alcanivoraceae. In *The Prokaryotes: Gammaproteobacteria*, Rosenberg, E.; DeLong, E. F.; Lory, S.; Stackebrandt, E.; Thompson, F., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 2014; pp 59-67.

76. Albuquerque, L.; da Costa, M. S., The Family Idiomarinaceae. In *The Prokaryotes: Gammaproteobacteria*, Rosenberg, E.; DeLong, E. F.; Lory, S.; Stackebrandt, E.; Thompson, F., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 2014; pp 361-385.

77. Arahal, D. R.; Ventosa, A., The Family Halomonadaceae. In *The Prokaryotes: Volume 6: Proteobacteria: Gamma Subclass*, Dworkin, M.; Falkow, S.; Rosenberg, E.; Schleifer, K.-H.; Stackebrandt, E., Eds. Springer New York: New York, NY, 2006; pp 811-835.

78. Vikram, A.; Bomberger, J. M.; Bibby, K. J., Efflux as a Glutaraldehyde Resistance Mechanism in *Pseudomonas fluorescens* and *Pseudomonas aeruginosa* Biofilms. *Antimicrobial Agents and Chemotherapy* **2015**, *59*, 3433-3440.

79. Fahrenfeld, N. L.; Delos Reyes, H.; Eramo, A.; Akob, D. M.; Mumford, A. C.; Cozzarelli, I. M., Shifts in microbial community structure and function in surface waters impacted by unconventional oil and gas wastewater revealed by metagenomics. *Sci Total Environ* **2017**, *580*, 1205-1213.
80. Le, P. T.; Pontarotti, P.; Raoult, D., Alphaproteobacteria species as a source and target of lateral sequence transfers. *Trends in Microbiology* **2014**, *22*, 147-156.
81. Leff, L. G.; Vaun McArthur, J.; Shimkets, L. J., Information spiraling: Movement of bacteria and their genes in streams. *Microbial Ecology* **1992**, *24*, 11-24.
82. Epand, R. M.; Walker, C.; Epand, R. F.; Magarvey, N. A., Molecular mechanisms of membrane targeting antibiotics. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2016**, *1858*, 980-987.
83. Kelly, D. P.; McDonald, I. R.; Wood, A. P., The Family Methylobacteriaceae. In *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria*, Rosenberg, E.; DeLong, E. F.; Lory, S.; Stackebrandt, E.; Thompson, F., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 2014; pp 313-340.
84. Drury, B.; Scott, J.; Rosi-Marshall, E. J.; Kelly, J. J., Triclosan exposure increases triclosan resistance and influences taxonomic composition of benthic bacterial communities. *Environ Sci Technol* **2013**, *47*, 8923-30.
85. Oh, S.; Tandukar, M.; Pavlostathis, S. G.; Chain, P. S. G.; Konstantinidis, K. T., Microbial community adaptation to quaternary ammonium biocides as revealed by metagenomics. *Environmental Microbiology* **2013**, *15*, 2850-2864.
86. Mumford, A. C.; Akob, D. M.; Klinges, J. G.; Cozzarelli, I. M., Common Hydraulic Fracturing Fluid Additives Alter the Structure and Function of Anaerobic Microbial Communities. *Appl Environ Microbiol* **2018**, *84*.
87. Tamas, I.; Smirnova, A. V.; He, Z.; Dunfield, P. F., The (d)evolution of methanotrophy in the Beijerinckiaceae--a comparative genomics analysis. *ISME J* **2014**, *8*, 369-82.
88. Koskella, B.; Vos, M., Adaptation in Natural Microbial Populations. *Annual Review of Ecology, Evolution, and Systematics* **2015**, *46*, 503-522.
89. Leahy, J. G.; Colwell, R. R., Microbial-Degradation of Hydrocarbons in the Environment. *Microbiological Reviews* **1990**, *54*, 305-315.

90. Brereton, R. G.; Lloyd, G. R., Partial least squares discriminant analysis: taking the magic away. *Journal of Chemometrics* **2014**, *28*, 213-225.
91. Barker, M.; Rayens, W., Partial least squares for discrimination. *Journal of Chemometrics* **2003**, *17*, 166-173.
92. Lee, L. C.; Liong, C.-Y.; Jemain, A. A., Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. *Analyst* **2018**, *143*, 3526-3539.
93. Flegal, C. L. O. M. D. C. C. D. F. a. K. M. *Prevalence of Obesity Among Adults and Youth: United States, 2011–2014*; U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES, Centers for Disease Control and Prevention, National Center for Health Statistics 2015.
94. Bray, G. A., Medical Consequences of Obesity. *The Journal of Clinical Endocrinology & Metabolism* **2004**, *89*, 2583-2589.
95. Burstein, S. H., N-Acyl Amino Acids (Elmiric Acids): Endogenous Signaling Molecules with Therapeutic Potential. *Molecular Pharmacology* **2018**, *93*, 228-238.
96. Burstein, S.; Salmonsén, R., Acylamido analogs of endocannabinoids selectively inhibit cancer cell proliferation. *Bioorganic & Medicinal Chemistry* **2008**, *16*, 9644-9651.
97. Neelam, P.; Vanessa, H. W. S., N-arachidonoyl glycine, an endogenous lipid that acts as a vasorelaxant via nitric oxide and large conductance calcium-activated potassium channels. *British Journal of Pharmacology* **2010**, *160*, 594-603.
98. McHugh, D.; Hu, S. S. J.; Rimmerman, N.; Juknat, A.; Vogel, Z.; Walker, J. M.; Bradshaw, H. B., N-arachidonoyl glycine, an abundant endogenous lipid, potently drives directed cellular migration through GPR18, the putative abnormal cannabidiol receptor. *BMC Neuroscience* **2010**, *11*, 44.
99. Burstein, S. H.; Adams, J. K.; Bradshaw, H. B.; Fraioli, C.; Rossetti, R. G.; Salmonsén, R. A.; Shaw, J. W.; Walker, J. M.; Zipkin, R. E.; Zurier, R. B., Potential anti-inflammatory actions of the elmiric (lipoamino) acids. *Bioorganic & Medicinal Chemistry* **2007**, *15*, 3345-3355.
100. Huang, S. M.; Bisogno, T.; Petros, T. J.; Chang, S. Y.; Zavitsanos, P. A.; Zipkin, R. E.; Sivakumar, R.; Coop, A.; Maeda, D. Y.; De Petrocellis, L.; Burstein, S.; Di Marzo, V.; Walker, J. M., Identification of a New Class of Molecules, the Arachidonoyl Amino Acids,

and Characterization of One Member That Inhibits Pain. *Journal of Biological Chemistry* **2001**, 276, 42639-42644.

101. Hofmann, A. F.; Hagey, L. R.; Krasowski, M. D., Bile salts of vertebrates: structural variation and possible evolutionary significance. *Journal of Lipid Research* **2010**, 51, 226-246.

102. Danielsson, H., Influence of Bile Acids on Digestion and Absorption of Lipids. *The American Journal of Clinical Nutrition* **1963**, 12, 214-219.

103. Hofmann, A. F.; Borgström, B., The Intraluminal Phase of Fat Digestion in Man: The Lipid Content of the Micellar and Oil Phases of Intestinal Content Obtained during Fat Digestion and Absorption. *Journal of Clinical Investigation* **1964**, 43, 247-257.

104. Kuipers, F.; Bloks, V. W.; Groen, A. K., Beyond intestinal soap—bile acids in metabolic control. *Nature Reviews Endocrinology* **2014**, 10, 488.

105. de Boer, J. F.; Bloks, V. W.; Verkade, E.; Heiner-Fokkema, M. R.; Kuipers, F., New insights in the multiple roles of bile acids and their signaling pathways in metabolic control. *Current Opinion in Lipidology* **2018**, 29, 194-202.

106. Albaugh, V. L.; Flynn, C. R.; Cai, S.; Xiao, Y.; Tamboli, R. A.; Abumrad, N. N., Early Increases in Bile Acids Post Roux-en-Y Gastric Bypass Are Driven by Insulin-Sensitizing, Secondary Bile Acids. *The Journal of Clinical Endocrinology & Metabolism* **2015**, 100, E1225-E1233.

107. Bhutta, H. Y.; Rajpal, N.; White, W.; Freudenberg, J. M.; Liu, Y.; Way, J.; Rajpal, D.; Cooper, D. C.; Young, A.; Tavakkoli, A.; Chen, L., Effect of Roux-en-Y Gastric Bypass Surgery on Bile Acid Metabolism in Normal and Obese Diabetic Rats. *PLOS ONE* **2015**, 10, e0122273.

108. Dutia, R.; Embrey, M.; O'Brien, S.; Haeusler, R. A.; Agénor, K. K.; Homel, P.; McGinty, J.; Vincent, R. P.; Alaghband-Zadeh, J.; Staels, B.; le Roux, C. W.; Yu, J.; Laferrère, B., Temporal changes in bile acid levels and 12 α -hydroxylation after Roux-en-Y gastric bypass surgery in type 2 diabetes. *International Journal Of Obesity* **2015**, 39, 806.

109. Flynn, C. R.; Albaugh, V. L.; Cai, S.; Cheung-Flynn, J.; Williams, P. E.; Brucker, R. M.; Bordenstein, S. R.; Guo, Y.; Wasserman, D. H.; Abumrad, N. N., Bile diversion to the

distal small intestine has comparable metabolic benefits to bariatric surgery. *Nature Communications* **2015**, *6*, 7715.

110. Palau-Rodriguez, M.; Tulipani, S.; Marco-Ramell, A.; Miñarro, A.; Jauregui, O.; Gonzalez-Dominguez, R.; Sanchez-Pla, A.; Ramos-Molina, B.; Tinahones, F. J.; Andres-Lacueva, C., Characterization of Metabolomic Profile Associated with Metabolic Improvement after Bariatric Surgery in Subjects with Morbid Obesity. *Journal of Proteome Research* **2018**.

111. Tan, B.; O'Dell, D. K.; Yu, Y. W.; Monn, M. F.; Hughes, H. V.; Burstein, S.; Walker, J. M., Identification of endogenous acyl amino acids based on a targeted lipidomics approach. *Journal of Lipid Research* **2010**, *51*, 112-119.

112. Scherer, M.; Gnewuch, C.; Schmitz, G.; Liebisch, G., Rapid quantification of bile acids and their conjugates in serum by liquid chromatography–tandem mass spectrometry. *Journal of Chromatography B* **2009**, *877*, 3920-3925.

113. Alnouti, Y.; Csanaky, I. L.; Klaassen, C. D., Quantitative-profiling of bile acids and their conjugates in mouse liver, bile, plasma, and urine using LC–MS/MS. *Journal of Chromatography B* **2008**, *873*, 209-217.

114. Bentayeb, K.; Batlle, R.; Sánchez, C.; Nerín, C.; Domeño, C., Determination of bile acids in human serum by on-line restricted access material–ultra high-performance liquid chromatography–mass spectrometry. *Journal of Chromatography B* **2008**, *869*, 1-8.

115. García-Cañaveras, J. C.; Donato, M. T.; Lahoz, A., Ultra-Performance Liquid Chromatography-Mass Spectrometry Targeted Profiling of Bile Acids: Application to Serum, Liver Tissue, and Cultured Cells of Different Species. In *Mass Spectrometry in Metabolomics: Methods and Protocols*, Raftery, D., Ed. Springer New York: New York, NY, 2014; pp 233-247.

116. Hagio, M.; Matsumoto, M.; Fukushima, M.; Hara, H.; Ishizuka, S., Improved analysis of bile acids in tissues and intestinal contents of rats using LC/ESI-MS. *Journal of Lipid Research* **2009**, *50*, 173-180.

117. Krautbauer, S.; Liebisch, G., LC-MS/MS Analysis of Bile Acids. In *Clinical Metabolomics: Methods and Protocols*, Giera, M., Ed. Springer New York: New York, NY, 2018; pp 103-110.

118. Perwaiz, S.; Tuchweber, B.; Mignault, D.; Gilat, T.; Yousef, I. M., Determination of bile acids in biological fluids by liquid chromatography-electrospray tandem mass spectrometry. *Journal of Lipid Research* **2001**, *42*, 114-119.
119. F., C. M.; Eugene, M.; D., R. J., LC-MS Data Processing with MAVEN: A Metabolomic Analysis and Visualization Engine. *Current Protocols in Bioinformatics* **2012**, *37*, 14.11.1-14.11.23.
120. Melamud, E.; Vastag, L.; Rabinowitz, J. D., Metabolomic Analysis and Visualization Engine for LC-MS Data. *Analytical Chemistry* **2010**, *82*, 9818-9826.
121. Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P., ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **2008**, *24*, 2534-2536.
122. Wickham, H., *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag: New York, 2009.
123. Vu, V. Q. *ggbiplot: A ggplot2 based biplot*, 2011.
124. Sanchez, C. *Discriminer: Tools of the Trade for Discriminant Analysis.*, 2013.
125. Rabinowitz, J. D.; Kimball, E., Acidic Acetonitrile for Cellular Metabolome Extraction from Escherichia coli. *Analytical Chemistry* **2007**, *79*, 6167-6173.
126. Milne, S.; Ivanova, P.; Forrester, J.; Alex Brown, H., Lipidomics: An analysis of cellular lipids by ESI-MS. *Methods* **2006**, *39*, 92-103.
127. Torchon, E. T.; Das, S.; Beckford, R. C.; Voy, B. H., Enriching the Starter Diet in n-3 Polyunsaturated Fatty Acids Reduces Adipocyte Size in Broiler Chicks. *Current Developments in Nutrition* **2017**, *1*, e001644-e001644.
128. Stough, J. M. A.; Dearth, S. P.; Denny, J. E.; LeCleur, G. R.; Schmidt, N. W.; Campagna, S. R.; Wilhelm, S. W., Functional characteristics of the gut microbiome in C57BL/6 mice differentially susceptible to Plasmodium yoelii. *Frontiers in Microbiology* **2016**, *7*.
129. Chang, S.-H.; Song, N.-J.; Choi, J. H.; Yun, U. J.; Park, K. W., Mechanisms underlying Ucp1 dependent and independent adipocyte thermogenesis. *Obesity Reviews* *0*.
130. Lin, H.; Long, J. Z.; Roche, A. M.; Svensson, K. J.; Dou, F. Y.; Chang, M. R.; Strutzenberg, T.; Ruiz, C.; Cameron, M. D.; Novick, S. J.; Berdan, C. A.; Louie, S. M.;

- Nomura, D. K.; Spiegelman, B. M.; Griffin, P. R.; Kamenecka, T. M., Discovery of Hydrolysis-Resistant Isoindoline N-Acyl Amino Acid Analogues that Stimulate Mitochondrial Respiration. *Journal of Medicinal Chemistry* **2018**, *61*, 3224-3230.
131. Champely, S. *pwr: Basic Functions for Power Analysis*, 2017.
132. Chong, J.; Soufan, O.; Li, C.; Caraus, I.; Li, S.; Bourque, G.; Wishart, D. S.; Xia, J., MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Research* **2018**, *46*, W486-W494.
133. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B. C.; Remm, M.; Rozen, S. G., Primer3—new capabilities and interfaces. *Nucleic Acids Research* **2012**, *40*, e115-e115.
134. Ji, B.; Middleton, J. L.; Ernest, B.; Saxton, A. M.; Lamont, S. J.; Campagna, S. R.; Voy, B. H., Molecular and metabolic profiles suggest that increased lipid catabolism in adipose tissue contributes to leanness in domestic chickens. *Physiological Genomics* **2014**, *46*, 315-327.
135. Conner, W. E.; Lin, D. S.; Colvis, C., Differential mobilization of fatty acids from adipose tissue. *Journal of Lipid Research* **1996**, *37*, 290-8.
136. Frigolet, M. E.; Gutiérrez-Aguilar, R., The Role of the Novel Lipokine Palmitoleic Acid in Health and Disease. *Advances in Nutrition* **2017**, *8*, 173S-181S.
137. Hunter, J. D.; Buchanan, H.; Nye, E. R., The mobilization of free fatty acids in relation to adipose tissue triglyceride fatty acids in the rat. *Journal of Lipid Research* **1970**, *11*, 259-265.
138. Gibson, R. A.; Neumann, M. A.; Lien, E. L.; Boyd, K. A.; Tu, W. C., Docosahexaenoic acid synthesis from alpha-linolenic acid is inhibited by diets high in polyunsaturated fatty acids. *Prostaglandins, Leukotrienes and Essential Fatty Acids (PLEFA)* **2013**, *88*, 139-146.
139. Blank, C.; Neumann, M. A.; Makrides, M.; Gibson, R. A., Optimizing DHA levels in piglets by lowering the linoleic acid to α -linolenic acid ratio. *Journal of Lipid Research* **2002**, *43*, 1537-1543.
140. Gebauer, S. K.; Psota, T. L.; Harris, W. S.; Kris-Etherton, P. M., n-3 Fatty acid dietary recommendations and food sources to achieve essentiality and cardiovascular benefits. *The American Journal of Clinical Nutrition* **2006**, *83*, 1526S-1535S.

141. Goyal, A.; Sharma, V.; Upadhyay, N.; Gill, S.; Sihag, M., Flax and flaxseed oil: an ancient medicine & modern functional food. *Journal of Food Science and Technology* **2014**, *51*, 1633-1653.
142. Long, J. Z.; Roche, A. M.; Berdan, C. A.; Louie, S. M.; Roberts, A. J.; Svensson, K. J.; Dou, F. Y.; Bateman, L. A.; Mina, A. I.; Deng, Z.; Jedrychowski, M. P.; Lin, H.; Kamenecka, T. M.; Asara, J. M.; Griffin, P. R.; Banks, A. S.; Nomura, D. K.; Spiegelman, B. M., Ablation of PM20D1 reveals *N*-acyl amino acid control of metabolism and nociception. *Proceedings of the National Academy of Sciences* **2018**, *115*, E6937-E6945.
143. Jang, S.-P.; Park, S.-H.; Jung, J.-S.; Lee, H.-J.; Hong, J.-W.; Lee, J.-Y.; Suh, H.-W., Characterization of changes of pain behavior and signal transduction system in food-deprived mice. *Animal cells and systems* **2018**, *22*, 227-233.
144. Saxena, I.; Kumar, M.; Verma, A., Pain response during fasting and postprandial conditions in healthy young Indian males. *International Journal of Clinical and Experimental Physiology* **2014**, *1*, 262-265.
145. Alhadeff, A. L.; Su, Z.; Hernandez, E.; Klima, M. L.; Phillips, S. Z.; Holland, R. A.; Guo, C.; Hantman, A. W.; De Jonghe, B. C.; Betley, J. N., A Neural Circuit for the Suppression of Pain by a Competing Need State. *Cell* **2018**, *173*, 140-152.e15.
146. Chandrasekar, B.; Nelson, J. F.; Colston, J. T.; Freeman, G. L., Calorie restriction attenuates inflammatory responses to myocardial ischemia-reperfusion injury. *American Journal of Physiology-Heart and Circulatory Physiology* **2001**, *280*, H2094-H2102.
147. Mattson, M. P.; Wan, R., Beneficial effects of intermittent fasting and caloric restriction on the cardiovascular and cerebrovascular systems. *The Journal of Nutritional Biochemistry* **2005**, *16*, 129-137.
148. Longo, Valter D.; Mattson, Mark P., Fasting: Molecular Mechanisms and Clinical Applications. *Cell Metabolism* **2014**, *19*, 181-192.
149. WAN, R.; CAMANDOLA, S.; MATTSON, M. P., Intermittent fasting and dietary supplementation with 2-deoxy-d-glucose improve functional and metabolic cardiovascular risk factors in rats. *The FASEB Journal* **2003**, *17*, 1133-1134.
150. Castello, L.; Froio, T.; Maina, M.; Cavallini, G.; Biasi, F.; Leonarduzzi, G.; Donati, A.; Bergamini, E.; Poli, G.; Chiarpotto, E., Alternate-day fasting protects the rat heart

against age-induced inflammation and fibrosis by inhibiting oxidative damage and NF- κ B activation. *Free Radical Biology and Medicine* **2010**, *48*, 47-54.

151. Mattson, Mark P., Energy Intake and Exercise as Determinants of Brain Health and Vulnerability to Injury and Disease. *Cell Metabolism* **2012**, *16*, 706-722.

152. Nencioni, A.; Caffa, I.; Cortellino, S.; Longo, V. D., Fasting and cancer: molecular mechanisms and clinical application. *Nature Reviews Cancer* **2018**, *18*, 707-719.

153. Storz, G.; Wolf, Y. I.; Ramamurthi, K. S., Small Proteins Can No Longer Be Ignored. In *Annual Review of Biochemistry, Vol 83*, Kornberg, R. D., Ed. Annual Reviews: Palo Alto, 2014; Vol. 83, pp 753-+.

154. Brantl, S., Bacterial type I toxin-antitoxin systems. *RNA Bio.* **2012**, *9*, 1488-1490.

155. Unoson, C.; Wagner, E. G. H., A small SOS-induced toxin is targeted against the inner membrane in Escherichia coli. *Mol. Microbio.* **2008**, *70*, 258-270.

156. Walther, T. H.; Gottselig, C.; Grage, S. L.; Wolf, M.; Vargiu, A. V.; Klein, M. J.; Vollmer, S.; Prock, S.; Hartmann, M.; Afonin, S.; Stockwald, E.; Heinzmann, H.; Nolandt, O. V.; Wenzel, W.; Ruggerone, P.; Ulrich, A. S., Folding and Self-Assembly of the TatA Translocation Pore Based on a Charge Zipper Mechanism. *Cell* **2013**, *152*, 316-326.

157. Preston, G. W.; Wilson, A. J., Photo-induced covalent cross-linking for the analysis of biomolecular interactions. *Chemical Society Reviews* **2013**, *42*, 3289-3301.

158. Vila-Perelló, M.; Pratt, M. R.; Tulin, F.; Muir, T. W., Covalent Capture of Phospho-Dependent Protein Oligomerization by Site-Specific Incorporation of a Diazirine Photo-Cross-Linker. *J. Am. Chem. Soc.* **2007**, *129*, 8068-8069.

159. Suchanek, M.; Radzikowska, A.; Thiele, C., Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nat Meth* **2005**, *2*, 261-268.

160. Hopp, T. P.; Prickett, K. S.; Price, V. L.; Libby, R. T.; March, C. J.; Pat Cerretti, D.; Urdal, D. L.; Conlon, P. J., A Short Polypeptide Marker Sequence Useful for Recombinant Protein Identification and Purification. *Bio/Technology* **1988**, *6*, 1204-1210.

161. Amblard, M.; Fehrentz, J.-A.; Martinez, J.; Subra, G., Methods and protocols of modern solid phase peptide synthesis. *Mol Biotechnol* **2006**, *33*, 239-254.

162. Blencowe, A.; Hayes, W., Development and application of diazirines in biological and synthetic macromolecular systems. *Soft Matter* **2005**, *1*, 178-205.

163. Dubinsky, L.; Krom, B. P.; Meijler, M. M., Diazirine based photoaffinity labeling. *Bioorg. Med. Chem.* **2012**, *20*, 554-570.
164. MacKinnon, A. L.; Taunton, J., Target Identification by Diazirine Photo-Cross-linking and Click Chemistry. *Current protocols in chemical biology* **2009**, *1*, 55-73.
165. Ford, F.; Yuzawa, T.; Platz, M. S.; Matzinger, S.; Fülcher, M., Rearrangement of Dimethylcarbene to Propene: Study by Laser Flash Photolysis and ab Initio Molecular Orbital Theory. *J. Am. Chem. Soc.* **1998**, *120*, 4430-4438.
166. Dhanasekaran, N.; Wessling-Resnick, M.; Kelleher, D. J.; Johnson, G. L.; Ruoho, A. E., Mapping of the carboxyl terminus within the tertiary structure of transducin's alpha subunit using the heterobifunctional cross-linking reagent, 125I-N-(3-iodo-4-azidophenylpropionamido-S-(2-thiopyridyl) cysteine. *J. Biol. Chem.* **1988**, *263*, 17942-50.
167. Chong, P. C.; Hodges, R. S., A new heterobifunctional cross-linking reagent for the study of biological interactions between proteins. I. Design, synthesis, and characterization. *J. Biol. Chem.* **1981**, *256*, 5064-5070.
168. Bochkariov, D. E.; Kogon, A. A., Application of 3-[3-(3-(trifluoromethyl)diazirin-3-yl)phenyl]-2,3-dihydroxypropionic acid, carbene-generating, cleavable cross-linking reagent for photoaffinity labeling. *Analytical Biochemistry* **1992**, *204*, 90-95.
169. Resek, J. F.; Bhattacharya, S.; Khorana, H. G., A new photo-crosslinking reagent for the study of protein-protein interactions. *J. Org. Chem.* **1993**, *58*, 7598-7601.
170. Bochkareva, E. S.; Solovieva, M. E.; Girshovich, A. S., Targeting of GroEL to SecA on the cytoplasmic membrane of Escherichia coli. *Proceedings of the National Academy of Sciences* **1998**, *95*, 478-483.
171. Kölbel, K.; Ihling, C. H.; Sinz, A., Analysis of Peptide Secondary Structures by Photoactivatable Amino Acid Analogues. *Angewandte Chemie International Edition* **2012**, *51*, 12602-12605.
172. Lössl, P.; Kölbel, K.; Tänzler, D.; Nannemann, D.; Ihling, C. H.; Keller, M. V.; Schneider, M.; Zaucke, F.; Meiler, J.; Sinz, A., Analysis of Nidogen-1/Laminin γ 1 Interaction by Cross-Linking, Mass Spectrometry, and Computational Modeling Reveals Multiple Binding Modes. *PLoS ONE* **2014**, *9*, e112886.
173. Schmitz, E.; Ohme, R., 3,3-Pentamethylenediaziridine. *Org. Syn.* **1965**, *45*.

174. Church, R. F. R.; Weiss, M. J., Diazirines. II. Synthesis and properties of small functionalized diazirine molecules. Observations on the reaction of a diaziridine with the iodine-iodide ion system. *J. Org. Chem.* **1970**, *35*, 2465-2471.
175. Vu, H.-D.; Renault, J.; Roisnel, T.; Gouault, N.; Uriac, P., Methanesulfonic Acid Mediated Cyclization and Meyer–Schuster Rearrangement of γ -Amino-ynones: Access to Enantiopure Pyrrolidine Exocyclic Vinylogous Amides. *European Journal of Organic Chemistry* **2014**, *2014*, 4506-4514.
176. Geoffrey M. Bilcer, T. D., Sudha V. Ankala, John C. Lilly, Chunfeng Liu, Hui Lei, Arun K. Ghosh, Makoto Inoue Pyrrolidine compounds which inhibit beta-secretase activity and methods of use thereof. 2013.
177. Coin, I.; Beyermann, M.; Bienert, M., Solid-phase peptide synthesis: from standard procedures to the synthesis of difficult sequences. *Nature Protocols* **2007**, *2*, 3247.
178. Wen, J.; Harp, J. R.; Fozo, E. M., The 5' UTR of the type I toxin ZorO can both inhibit and enhance translation. *Nucleic Acids Research* **2016**, *45*, 4006-4020.
179. Olson, M. T.; Yergey, A. L., Calculation of the isotope cluster for polypeptides by probability grouping. *Journal of the American Society for Mass Spectrometry* **2009**, *20*, 295-302.
180. Blattner, F. R.; Plunkett, G.; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J. D.; Rode, C. K.; Mayhew, G. F.; Gregor, J.; Davis, N. W.; Kirkpatrick, H. A.; Goeden, M. A.; Rose, D. J.; Mau, B.; Shao, Y., The Complete Genome Sequence of *Escherichia coli* K-12. *Science* **1997**, *277*, 1453-1462.
181. Boom, T. V.; John E. Cronan, J., GENETICS AND REGULATION OF BACTERIAL LIPID METABOLISM. *Annual Review of Microbiology* **1989**, *43*, 317-343.
182. Saiki, R.; Gelfand, D.; Stoffel, S.; Scharf, S.; Higuchi, R.; Horn, G.; Mullis, K.; Erlich, H., Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **1988**, *239*, 487-491.
183. Westermeier, R., Gel Electrophoresis. In *eLS*, 2013.
184. Wysocki, V. H.; Resing, K. A.; Zhang, Q.; Cheng, G., Mass spectrometry of peptides and proteins. *Methods* **2005**, *35*, 211-222.
185. Harris, J. L.; Backes, B. J.; Leonetti, F.; Mahrus, S.; Ellman, J. A.; Craik, C. S., Rapid and general profiling of protease specificity by using combinatorial fluorogenic

substrate libraries. *Proceedings of the National Academy of Sciences* **2000**, *97*, 7754-7759.

186. Elzagheid, M. I., Laboratory Activities to Introduce Carbohydrates Qualitative Analysis to College Students. *World Journal of Chemical Education* **2018**, *6*, 82-86.

187. Geske, G. D.; Wezeman, R. J.; Siegel, A. P.; Blackwell, H. E., Small Molecule Inhibitors of Bacterial Quorum Sensing and Biofilm Formation. *Journal of the American Chemical Society* **2005**, *127*, 12762-12763.

188. Colton, D. M.; Stabb, E. V.; Hagen, S. J., Modeling Analysis of Signal Sensitivity and Specificity by *Vibrio fischeri* LuxR Variants. *PLOS ONE* **2015**, *10*, e0126474.

189. Miyashiro, T.; Ruby, E. G., Shedding light on bioluminescence regulation in *Vibrio fischeri*. *Mol Microbiol* **2012**, *84*, 795-806.

APPENDICES

Appendix A: Supporting information for chapter 2

A1: DESeq2 Results HF- vs HF+ Enrichment at Day 0 Prior to GA Addition

OTU	base Mean	log2FoldChange	lfcSE	stat	pvalue	padj	Rank1	Rank2	Rank3	Rank4	Rank5	Rank6
11162	47.34 198	-2.193913	0.689 598	3.181 436	0.001 465	0.007 65	k__Bacteria	p__Proteobacteria	c__Epsilonproteobacteria	o__Campylobacteriales	f__Campylobacteraceae	g__Arcobacter
New.ReferenceOTU6484	3.039 282	-2.830132	0.860 831	3.287 674	0.001 01	0.005 926	k__Bacteria	p__Proteobacteria	c__Alphaproteobacteria	o__Rhizobiales	f__Rhizobiaceae	g__Rhizobium
837424	62.88 603	3.089221	0.453 887	6.806 141	1.00 E-11	1.21 E-09	k__Bacteria	p__Proteobacteria	c__Alphaproteobacteria	o__Rhodospirillales	f__Hyphomicrobiaceae	g__
1845162	22.89 659	-2.902606	0.913 128	3.178 751	0.001 479	0.007 65	k__Bacteria	p__Proteobacteria	c__Alphaproteobacteria	o__Rhodospirillales	f__Rhodospirillaceae	g__Novispirillum
4614	3.013 996	4.048275	1.125 613	3.596 507	0.000 323	0.002 35	k__Bacteria	p__Proteobacteria	c__Alphaproteobacteria	o__Rhizobiales	f__Hyphomicrobiaceae	g__Pedomicrobium
4300564	1182. 266	-2.343628	0.394 975	5.933 615	2.96 E-09	1.40 E-07	k__Bacteria	p__Proteobacteria	c__Alphaproteobacteria	o__Caulobacteriales	f__Caulobacteraceae	g__Asticcacaulis
New.ReferenceOTU5190	229.7 425	-2.095606	0.570 181	3.675 338	0.000 238	0.001 904	k__Bacteria	p__Chlamydiae	c__Chlamydiae	o__Chlamydiales	f__Rhabdochlamydiaceae	g__Candidatus Rhabdochlamydia
279220	3.953 876	-3.02733	0.881 746	3.433 337	0.000 596	0.004 028	k__Bacteria	p__Proteobacteria	c__Gammaproteobacteria	o__Oceanospirillales	f__Oceanospirillaceae	g__Amphritea
646549	719.4 334	-2.327894	0.660 348	3.525 251	0.000 423	0.003 037	k__Bacteria	p__Proteobacteria	c__Gammaproteobacteria	o__Pseudomonadales	f__Pseudomonadaceae	g__Pseudomonas
210236	73.59 37	2.119667	0.551 446	3.843 834	0.000 121	0.001 092	k__Bacteria	p__OP3	c__koll11	o__	f__	g__
1136154	26.38 491	2.126885	0.400 933	5.304 84	1.13 E-07	3.39 E-06	k__Bacteria	p__Acidobacteria	c__Acidobacteria5	o__	f__	g__
846667	31.18 204	-3.704207	0.870 947	4.253 082	2.11 E-05	0.000 241	k__Bacteria	p__Acidobacteria	c__Acidobacteria	o__Acidobacteriales	f__Acidobacteriaceae	g__Acidicapsa
239787	108.2 123	-2.384851	0.556 648	4.284 308	1.83 E-05	0.000 215	k__Bacteria	p__Acidobacteria	c__Acidobacteria	o__Acidobacteriales	f__Acidobacteriaceae	g__
4456187	6.810 858	2.382094	0.705 376	3.377 056	0.000 733	0.004 762	k__Bacteria	p__BHI80-139	c__	o__	f__	g__
687206	43.69 541	4.308538	0.832 137	5.177 68	2.25 E-06	5.69 E-06	k__Bacteria	p__Acidobacteria	c__[Chloracidobacteria]	o__RB41	f__Ellin6075	g__
876170	18.61 959	3.044209	0.616 636	4.936 801	7.94 E-05	1.59 E-05	k__Bacteria	p__Actinobacteria	c__Actinobacteria	o__Actinomycetales	f__Microbacteriaceae	g__Salinibacterium
813668	845.4 841	-2.272306	0.664 591	3.419 106	0.000 628	0.004 14	k__Bacteria	p__Actinobacteria	c__Actinobacteria	o__Actinomycetales	f__ACK-M1	g__
New.ReferenceOTU15756	6.124 113	3.363531	1.005 712	3.344 429	0.000 825	0.005 151	k__Bacteria	p__Armatimonadetes	c__0319-6E2	o__	f__	g__
837885	4.733 167	-6.32344	1.725 266	3.665 196	0.000 247	0.001 938	k__Bacteria	p__Firmicutes	c__Clostridia	o__Clostridiales	f__[Tissierellaceae]	g__Sedimentibacter
143418	21.03 482	6.457385	1.447 522	4.460 991	8.16 E-06	0.000 11	k__Bacteria	p__Verrucomicrobia	c__[Pedosphaerae]	o__[Pedosphaerales]	f__Ellin517	g__
801258	63.93 828	-2.372174	0.610 026	3.888 642	0.000 101	0.000 97	k__Bacteria	p__Verrucomicrobia	c__[Methylacidiphilae]	o__Methylacidiphilales	f__	g__
516569	1042. 129	-3.832002	0.683 3	5.608 083	2.05 E-08	7.57 E-07	k__Bacteria	p__Verrucomicrobia	c__Verrucomicrobiae	o__Verrucomicrobiales	f__Verrucomicrobiaceae	g__Prostheco bacter
557467	6.110 787	3.886557	0.853 781	4.552 169	5.31 E-06	7.74 E-05	k__Bacteria	p__Gemmatimonadetes	c__Gemmatimonadetes	o__Gemmatimonadales	f__Ellin5301	g__
New.ReferenceOTU26980	10.23 665	3.905893	0.900 864	4.335 72	1.45 E-05	0.000 179	k__Bacteria	p__Cyanobacteria	c__4C0d-2	o__SM1D11	f__	g__
New.ReferenceOTU38440	9.917 951	-5.361393	1.267 866	4.228 676	2.35 E-05	0.000 257	k__Bacteria	p__Cyanobacteria	c__Chloroplast	o__Cryptophyta	f__	g__
New.ReferenceOTU1728	231.1 528	-2.724327	0.593 826	4.587 756	4.48 E-06	6.73 E-05	k__Bacteria	p__Cyanobacteria	c__Chloroplast	o__Stramenopiles	f__	g__
New.ReferenceOTU56767	7.184 534	4.633563	1.253 721	3.695 849	0.000 219	0.001 787	k__Bacteria	p__ZB3	c__BS119	o__	f__	g__
New.ReferenceOTU1161	5.285 508	2.766702	0.755 418	3.662 477	0.000 25	0.001 938	k__Bacteria	p__Chloroflexi	c__TK17	o__mle1-48	f__	g__
204899	21.93 396	2.470338	0.488 134	5.060 777	4.18 E-07	9.13 E-06	k__Bacteria	p__Chloroflexi	c__Anaerolineae	o__H39	f__	g__
1039699	286.9 018	-2.604784	0.692 806	3.759 758	0.000 17	0.001 41	k__Bacteria	p__Armatimonadetes	c__Chthonomonadetes	o__Chthonomonadales	f__Chthonomonadaceae	g__
New.ReferenceOTU30447	380.2 165	-2.243244	0.465 503	4.818 969	1.44 E-06	2.67 E-05	k__Bacteria	p__Armatimonadetes	c__[Fimbrimonadia]	o__[Fimbrimonadales]	f__[Fimbrimonadaceae]	g__Fimbrimonas
New.ReferenceOTU53964	3.928 368	4.424498	1.287 984	3.435 211	0.000 592	0.004 028	k__Bacteria	p__Armatimonadetes	c__[Fimbrimonadia]	o__[Fimbrimonadales]	f__[Fimbrimonadaceae]	g__
141815	30.30 255	3.828168	0.851 369	4.496 485	6.91 E-05	9.77 E-05	k__Bacteria	p__Armatimonadetes	c__Armatimonadia	o__Armatimonadales	f__Armatimonadaceae	g__Armatimonas
New.ReferenceOTU51180	15.02 636	3.30893	0.645 341	5.127 412	2.94 E-07	7.06 E-06	k__Bacteria	p__Planctomycetes	c__Phycisphaerae	o__Phycisphaerales	f__	g__
518864	6.055 709	3.744301	0.974 865	3.840 841	0.000 123	0.001 092	k__Bacteria	p__Planctomycetes	c__OM190	o__agg27	f__	g__
New.ReferenceOTU30401	4.085 789	3.754071	0.976 273	3.845 307	0.000 12	0.001 092	k__Bacteria	p__Planctomycetes	c__OM190	o__CL500-15	f__	g__
New.ReferenceOTU48953	3.393 883	3.708252	0.950 243	3.902 423	9.52 E-05	0.000 935	k__Bacteria	p__WS3	c__PRR-12	o__Sediment-1	f__	g__
209782	10.91 541	2.581815	0.712 946	3.621 331	0.000 293	0.002 203	k__Bacteria	p__Chlorobi	c__BSV26	o__A89	f__	g__
New.ReferenceOTU18116	69.46 244	2.4089	0.742 947	3.242 359	0.001 185	0.006 708	k__Bacteria	p__Bacteroidetes	c__[Saprosiriales]	o__[Saprosiriales]	f__Saprosiraceae	g__

575486	14.29 116	-2.955491	0.674 937	- 4.378 914	1.19 E-05	0.000 151	k_Ba cteria	p__Bacteroides	c__Bacteroidia	o__Bacteroidales	f__Porphyromonadaceae	g__Paludibacter
2614	6.428 6	-4.80527	1.555 66	3.088 895	0.002 009	0.009 974	k_Ba cteria	p__Bacteroides	c__Cytophagia	o__Cytophagales	f__Cyclobacteriaceae	g__
1068470	18.77 054	-2.502451	0.773 249	3.236 281	0.001 211	0.006 773	k_Ba cteria	p__Bacteroides	c__Sphingobacteria	o__Sphingobacteriales	f__Sphingobacteriaceae	g__Pedobacter
1038865	664.3 906	-3.483155	0.491 445	7.087 576	1.36 E-12	3.28 E-10	k_Ba cteria	p__Bacteroides	c__Sphingobacteria	o__Sphingobacteriales	f__Sphingobacteriaceae	g__
3025055	263.1 595	-2.133306	0.640 053	3.333 015	0.000 859	0.005 219	k_Ba cteria	p__Bacteroides	c__[Saprosiriae]	o__[Saprosiriales]	f__	g__
712582	22.40 42	3.284905	0.700 438	4.689 79	2.73 E-06	4.30 E-05	k_Ba cteria	p__Fibrobacteres	c__Fibrobacteria	o__258ds10	f__	g__
1001333	12.75 491	-2.077087	0.635 571	3.268 066	0.001 083	0.006 201	k_Ba cteria	p__Acidobacteria	c__Holophagae	o__Holophagales	f__Holophagaceae	g__
177975	3968. 259	-4.289763	0.550 561	7.791 626	6.62 E-15	3.18 E-12	k_Ba cteria	p__Proteobacteria	c__Alphaproteobacteria	o__Sphingomonadales	f__Sphingomonadaceae	g__Novosphingobium
564411	2146. 885	-3.115014	0.627 728	4.962 364	6.96 E-07	1.46 E-05	k_Ba cteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__Oxalobacteraceae	g__Polynucleobacter
952163	4.007 059	-4.834986	1.504 231	3.214 258	0.001 308	0.007 131	k_Ba cteria	p__Proteobacteria	c__Betaproteobacteria	o__Neisseriales	f__Neisseriaceae	g__Silvimonas
1104923	29.16 899	3.834611	0.613 346	6.251 955	4.05 E-10	2.44 E-08	k_Ba cteria	p__Proteobacteria	c__Betaproteobacteria	o__A21b	f__EB1003	g__
583489	92.12 283	2.255639	0.355 346	6.347 726	2.19 E-10	1.50 E-08	k_Ba cteria	p__Proteobacteria	c__Betaproteobacteria	o__	f__	g__
836783	133.3 941	-3.089256	0.739 749	4.176 085	2.97 E-05	0.000 317	k_Ba cteria	p__Proteobacteria	c__Gammaproteobacteria	o__Alteromonadales	f__Shewanellaceae	g__Shewanella
529644	37.44 613	2.569163	0.475 812	5.399 534	6.88 E-08	2.18 E-06	k_Ba cteria	p__Proteobacteria	c__Betaproteobacteria	o__SC-I-84	f__	g__
566578	628.5 373	-4.823223	0.893 763	5.396 536	6.79 E-08	2.18 E-06	k_Ba cteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__Comamonadaceae	g__Limnobacter
944197	194.0 79	-2.660165	0.386 521	6.882 333	5.89 E-12	9.44 E-10	k_Ba cteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__Oxalobacteraceae	g__Herminimonas
1060517	28540 .05	-5.561951	0.987 092	5.634 681	1.75 E-08	7.03 E-07	k_Ba cteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__Oxalobacteraceae	g__Janthinobacterium
352419	100.2 517	-2.052368	0.653 117	3.142 421	0.001 676	0.008 574	k_Ba cteria	p__Proteobacteria	c__Betaproteobacteria	o__SBla14	f__	g__
590601	4325. 567	-2.788918	0.676 977	4.119 665	3.79 E-05	0.000 388	k_Ba cteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__Oxalobacteraceae	g__
330902	34.18 851	3.320655	0.638 127	5.203 755	1.95 E-07	5.22 E-06	k_Ba cteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__	g__
587098	88.90 706	-2.673014	0.567 227	4.712 421	2.45 E-06	4.06 E-05	k_Ba cteria	p__Proteobacteria	c__Betaproteobacteria	o__Gallionellales	f__Gallionellaceae	g__Gallionella
New.CleanUp.Referen ceOTU174689	1249. 865	-4.665306	0.697 794	6.685 788	2.30 E-11	1.84 E-09	k_Ba cteria	p__Verrucomicrobia	c__Verrucomicrobiae	o__Verrucomicrobiales	f__Verrucomicrobiaceae	g__
1028297	90.00 013	-3.047859	0.957 59	3.182 844	0.001 458	0.007 65	k_Ba cteria	p__Verrucomicrobia	c__[Spartobacteria]	o__[Chthoniobacteriales]	f__[Chthoniobacteraceae]	g__Ellin506
937848	54.54 269	-3.084718	0.960 454	3.211 729	0.001 319	0.007 131	k_Ba cteria	p__Verrucomicrobia	c__[Spartobacteria]	o__[Chthoniobacteriales]	f__[Chthoniobacteraceae]	g__heteroC45_4W
335484	37.99 237	3.837284	0.818 691	4.687 095	2.77 E-06	4.30 E-05	k_Ba cteria	p__Proteobacteria	c__Betaproteobacteria	o__Rhodocyclales	f__Rhodocyclaceae	g__Uliginosibacterium
831179	18.51 828	3.212604	0.777 303	4.133 016	3.58 E-05	0.000 374	k_Ba cteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__Comamonadaceae	g__Rubrivivax
556648	9.012 52	5.616373	1.168 017	4.808 47	1.52 E-06	2.71 E-05	k_Ba cteria	p__Acidobacteria	c__[Chloracidobacteria]	o__PK29	f__	g__
37406	167.2 015	-3.096749	0.585 646	5.287 745	1.24 E-07	3.50 E-06	k_Ba cteria	p__Proteobacteria	c__Alphaproteobacteria	o__Rhodospirillales	f__Rhodospirillaceae	g__Telmatospirillum
228223	20.35 519	-2.824689	0.750 62	3.763 14	0.000 168	0.001 41	k_Ba cteria	p__Proteobacteria	c__Alphaproteobacteria	o__Rhodospirillales	f__Rhodospirillaceae	g__Magnetospirillum
831824	1208. 293	-2.044954	0.538 616	3.796 681	0.000 147	0.001 282	k_Ba cteria	p__Proteobacteria	c__Alphaproteobacteria	o__Rhodospirillales	f__Acetobacteraceae	g__
267354	83.58 104	-5.378316	0.908 369	5.920 85	3.20 E-09	1.40 E-07	k_Ba cteria	p__Proteobacteria	c__Alphaproteobacteria	o__Rhodospirillales	f__Acetobacteraceae	g__Acidocella
581747	6.360 133	-4.004986	1.063 503	3.765 844	0.000 166	0.001 41	k_Ba cteria	p__Proteobacteria	c__Alphaproteobacteria	o__Rhodospirillales	f__Acetobacteraceae	g__Acidisona

A2: GA concentration over time shown by stream source

Condition	Days	Impacted	Stream	101.0600 m/z (Area)	Concentration (mg/L)	Std Error
Biotic	7	HF+	AB	3.85E+08	79.48160912	13.11212854
Biotic	7	HF+	LL	3.96E+08	66.7273629	3.801995694
Biotic	7	HF+	NH	3.81E+08	63.97547606	1.110752354
Biotic	7	HF-	EE	4.13E+08	60.70884643	3.801995694
Biotic	7	HF-	WE	3.95E+08	61.95103937	1.110752354
Biotic	7	HF-	DR	3.97E+08	63.8216446	1.255625153
Biotic	28	HF+	AB	3.64E+08	52.58146621	1.110752354
Biotic	28	HF+	LL	3.75E+08	74.15880096	1.255625153
Biotic	28	HF+	NH	3.86E+08	65.08034858	0.335364966
Biotic	28	HF-	EE	3.92E+08	52.97379609	1.255625153
Biotic	28	HF-	WE	3.90E+08	41.42101544	0.335364966
Biotic	28	HF-	DR	4.12E+08	51.4239386	0.353042028
Biotic	56	HF+	AB	3.87E+08	34.18111899	0.335364966
Biotic	56	HF+	LL	3.73E+08	54.47144686	0.353042028
Biotic	56	HF+	NH	3.96E+08	53.11696095	2.563422329
Biotic	56	HF-	EE	4.34E+08	22.94410647	0.353042028
Biotic	56	HF-	WE	3.86E+08	28.34192364	2.563422329
Biotic	56	HF-	DR	3.83E+08	43.82851494	6.352693439
Abiotic	7	HF+	AB	3.11E+08	101.5543597	2.563422329
Abiotic	7	HF+	LL	2.68E+08	105.2199569	6.352693439
Abiotic	7	HF+	NH	2.59E+08	100.342735	0.490986499
Abiotic	7	HF-	EE	2.48E+08	110.2945362	6.352693439
Abiotic	7	HF-	WE	2.52E+08	105.2615509	0.490986499
Abiotic	7	HF-	DR	2.58E+08	105.3721828	1.458119011
Abiotic	28	HF+	AB	2.19E+08	95.35410966	0.490986499
Abiotic	28	HF+	LL	2.94E+08	98.65858198	1.458119011
Abiotic	28	HF+	NH	2.63E+08	102.0578332	1.95400352
Abiotic	28	HF-	EE	2.20E+08	103.7405871	1.458119011
Abiotic	28	HF-	WE	1.78E+08	103.4594802	1.95400352
Abiotic	28	HF-	DR	2.15E+08	110.309118	2.75098417
Abiotic	56	HF+	AB	1.51E+08	102.4418522	1.95400352
Abiotic	56	HF+	LL	2.24E+08	97.98890524	2.75098417
Abiotic	56	HF+	NH	2.21E+08	105.1594224	3.186512269
Abiotic	56	HF-	EE	1.07E+08	117.0423702	2.75098417
Abiotic	56	HF-	WE	1.29E+08	102.0744244	3.186512269
Abiotic	56	HF-	DR	1.87E+08	101.2384625	12.72747454

A3: pH changes over time in the microcosms

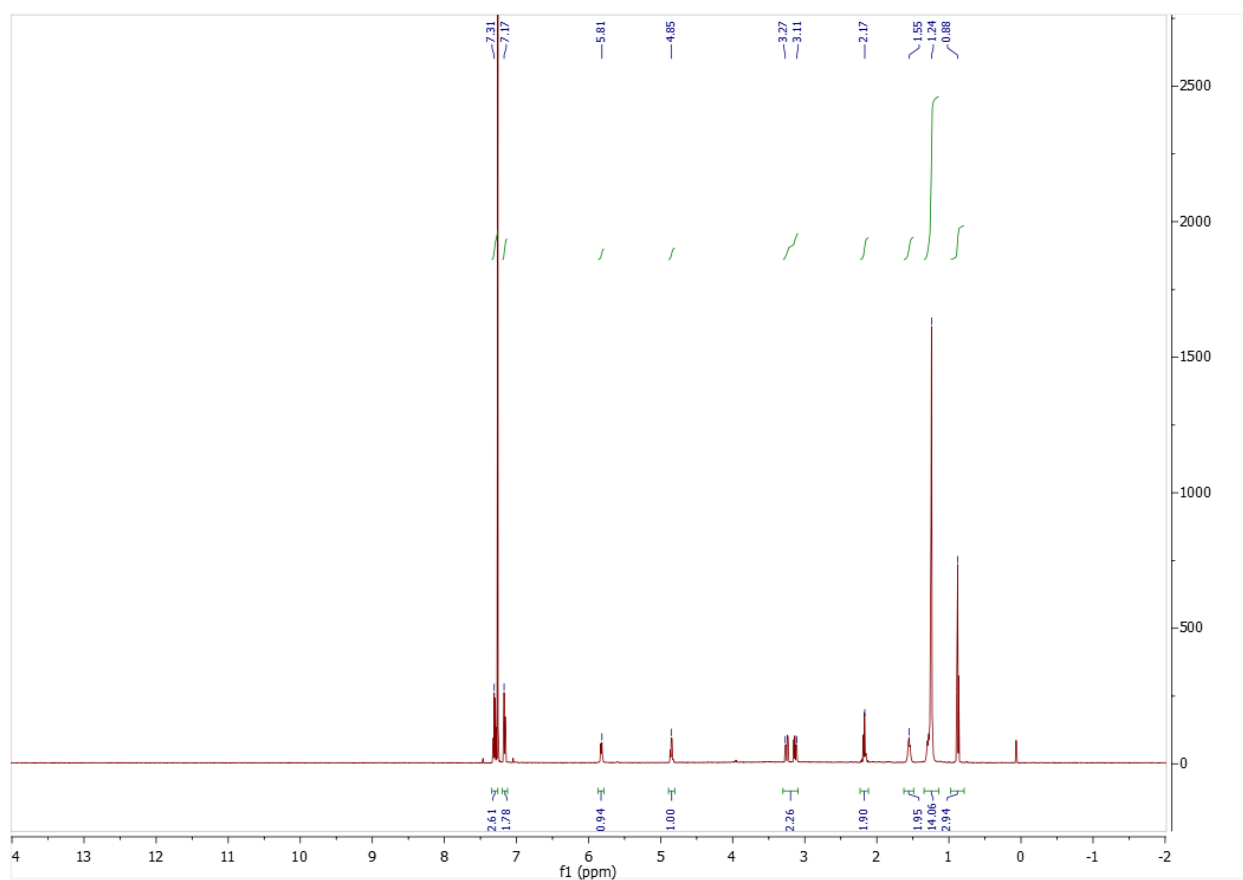
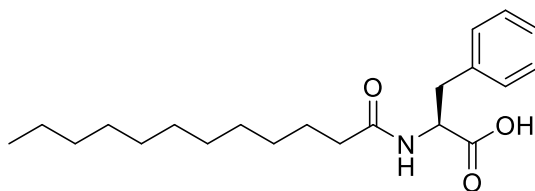
	Day 0- No GA	Day 14	Day 21	Day 28	Day 35	Day 49
Group	pH / SD	pH / SD	pH / SD	pH / SD	pH / SD	pH / SD
HF+	5.3 / 0.6	7.8 / 0.2	6.9 / 0.9	8.4 / 0.0	6.2 / 1.0	4.2 / 0.1
HF-	6.5 / 0.3	7.7 / 0.3	7.4 / 0.2	8.1 / 0.2	5.2 / 0.5	4.1 / 0.1
HF- biotic	5.3 / 0.6	7.5 / 0.2	8.4 / 0.0	7.9 / 0.4	8.2 / 0.7	5.7 / 0.5
HF- abiotic	6.5 / 0.3	7.7 / 0.2	7.6 / 0.1	7.6 / 0.3	8.4 / 0.1	6.7 / 0.3

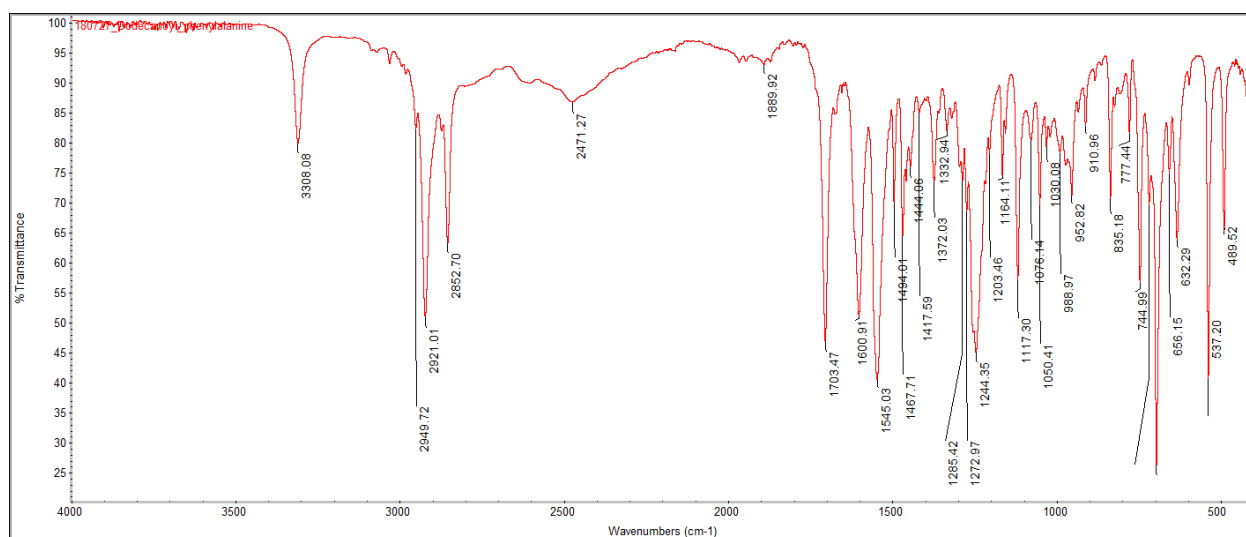
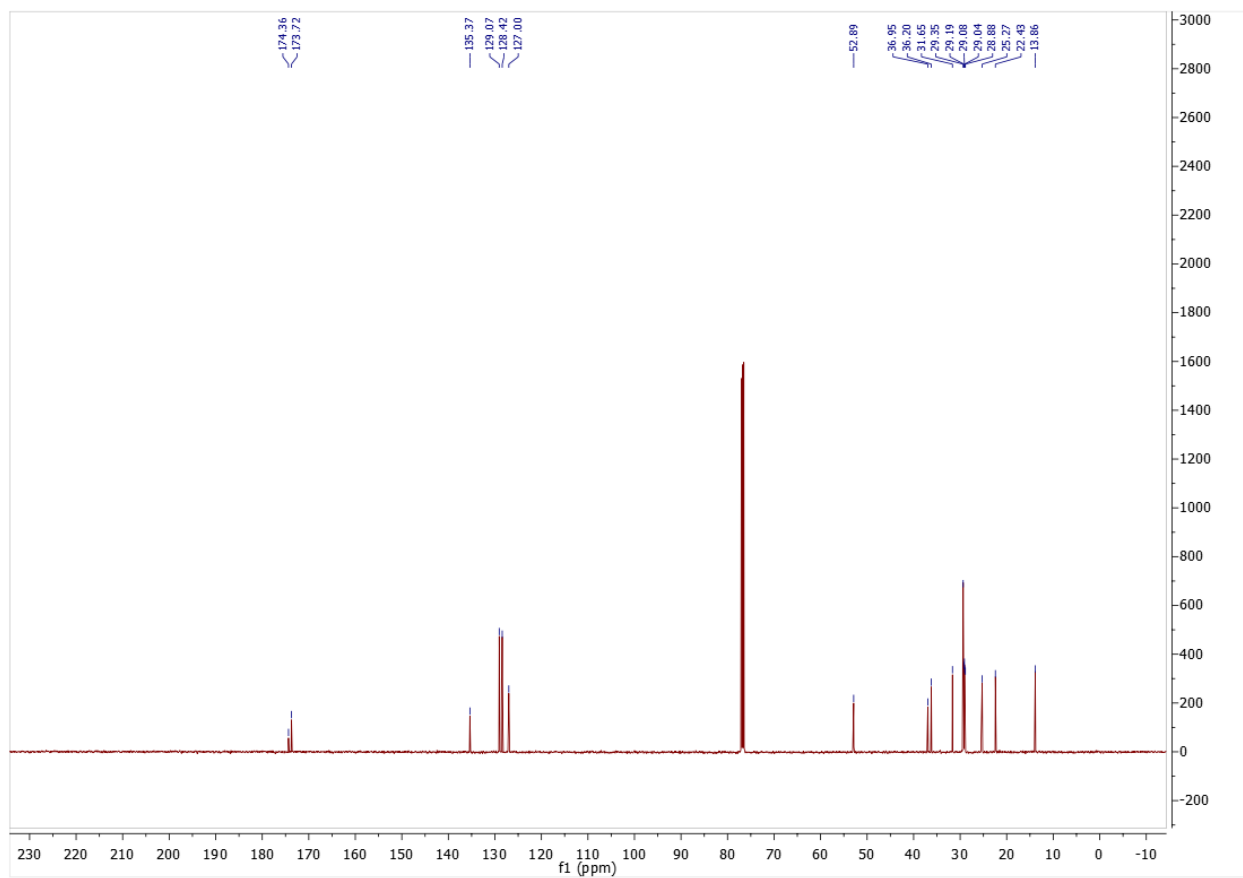
The pH each group n=9 was averaged and standard deviation (SD) calculated. Day 0-No GA was measured prior to the start of the experiment. The other samples were measured after freeze-thaw. Days not shown were not measured because samples were depleted.

Appendix B: Supporting information for chapter 3

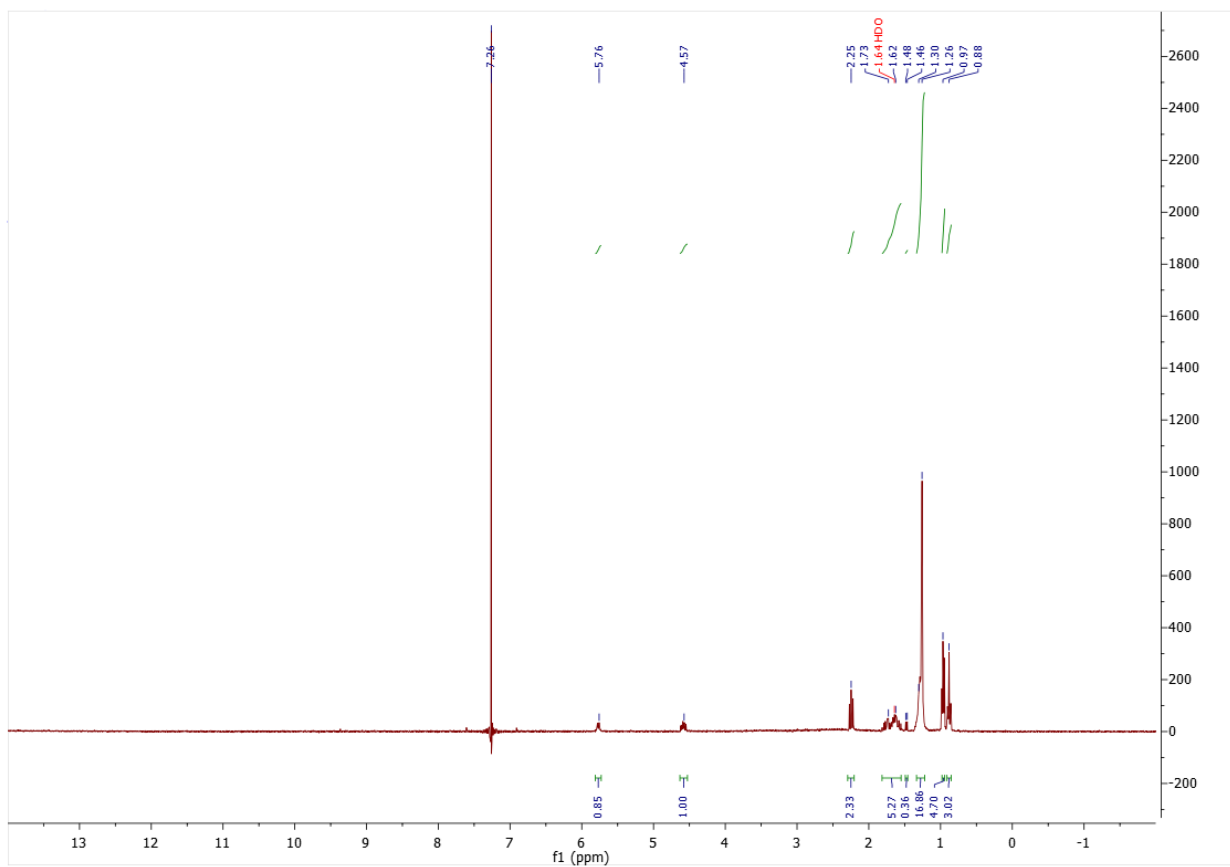
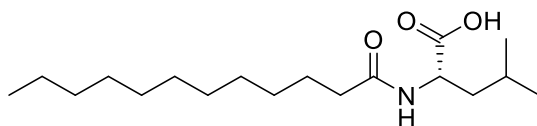
B1: Characterization data for synthetic *N*-acyl amino acids

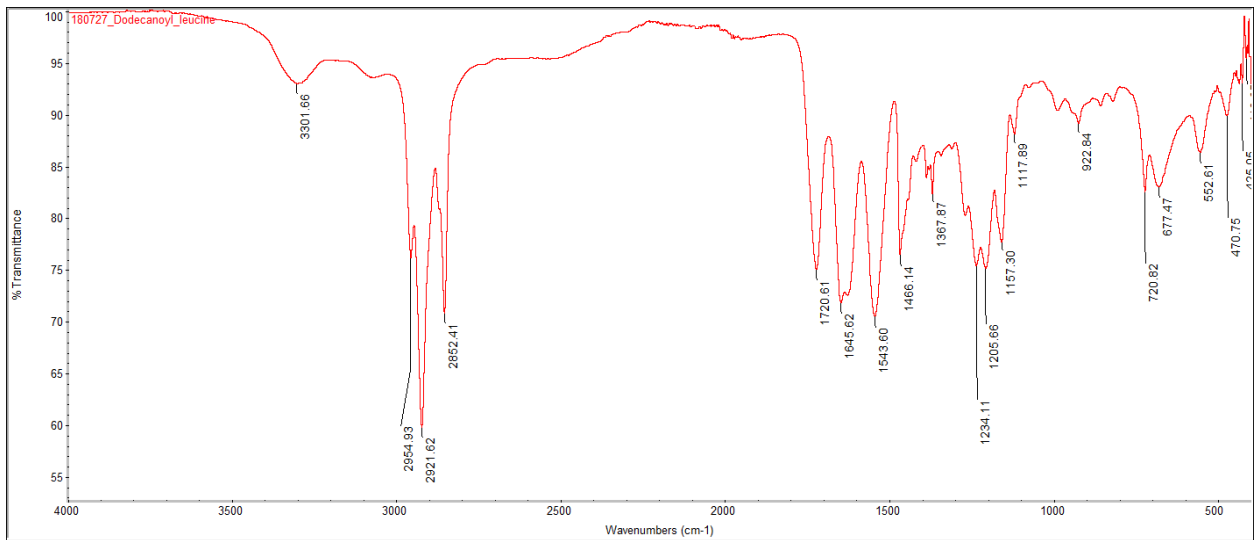
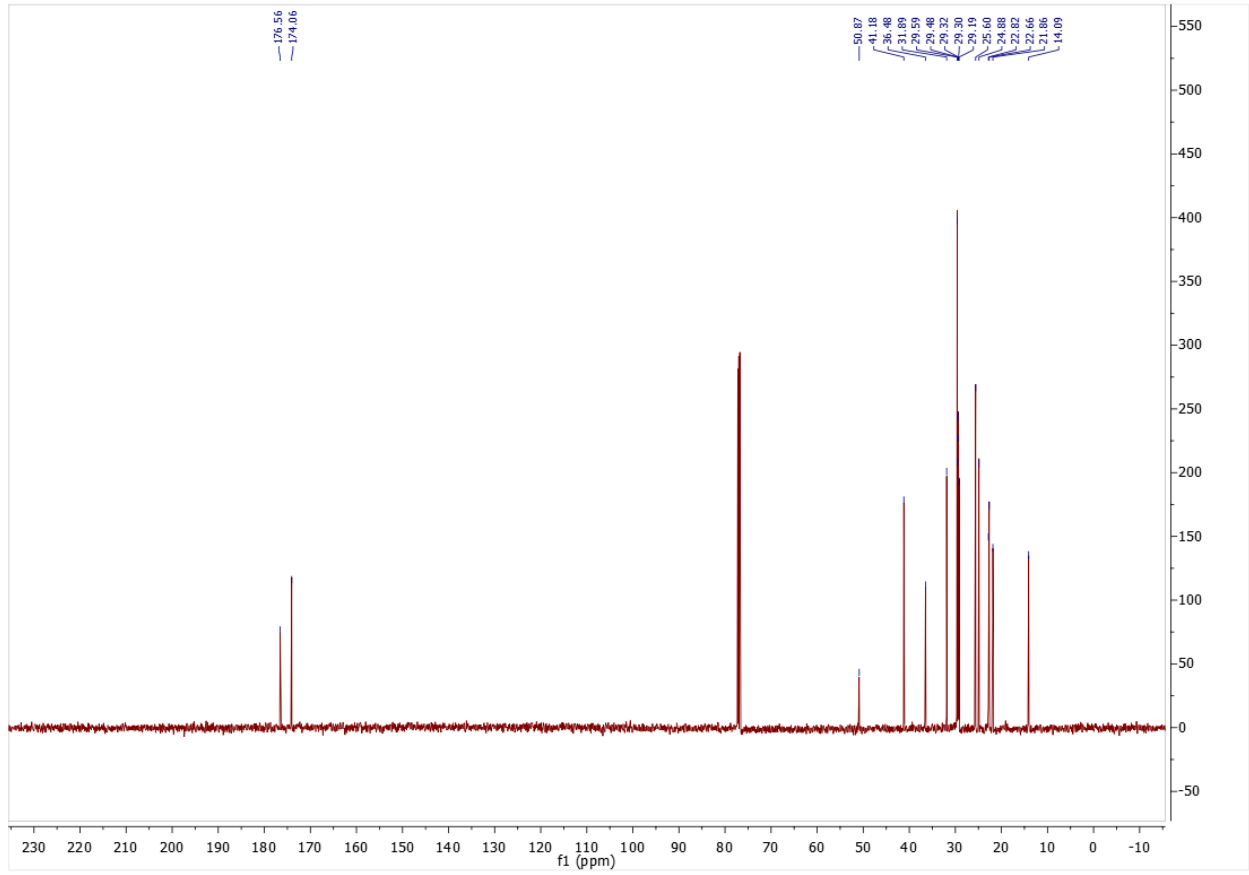
Dodecanoyl phenylalanine: ^1H NMR (500 MHz, CDCl_3) δ 0.88 (t, 3H), 1.24 (m, 14H), 1.55 (m, 2H), 2.17 (t, 2H), 3.11-3.27 (dq, 2H), 4.85 (m, 1H), 5.81 (d, 1H), 7.17 (m, 2H), 7.31 (m, 3H). ^{13}C NMR (500 MHz, CDCl_3): δ 13.86, 22.43, 25.27, 28.88, 29.04, 29.08, 29.19, 29.35, 31.65, 36.20, 36.95, 52.89, 127.00, 128.42, 129.07, 135.37, 173.72, 174.36. HRMS (ESI) m/z calculated for $\text{C}_{21}\text{H}_{33}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 346.2388, found: 346.2385. IR shown below.



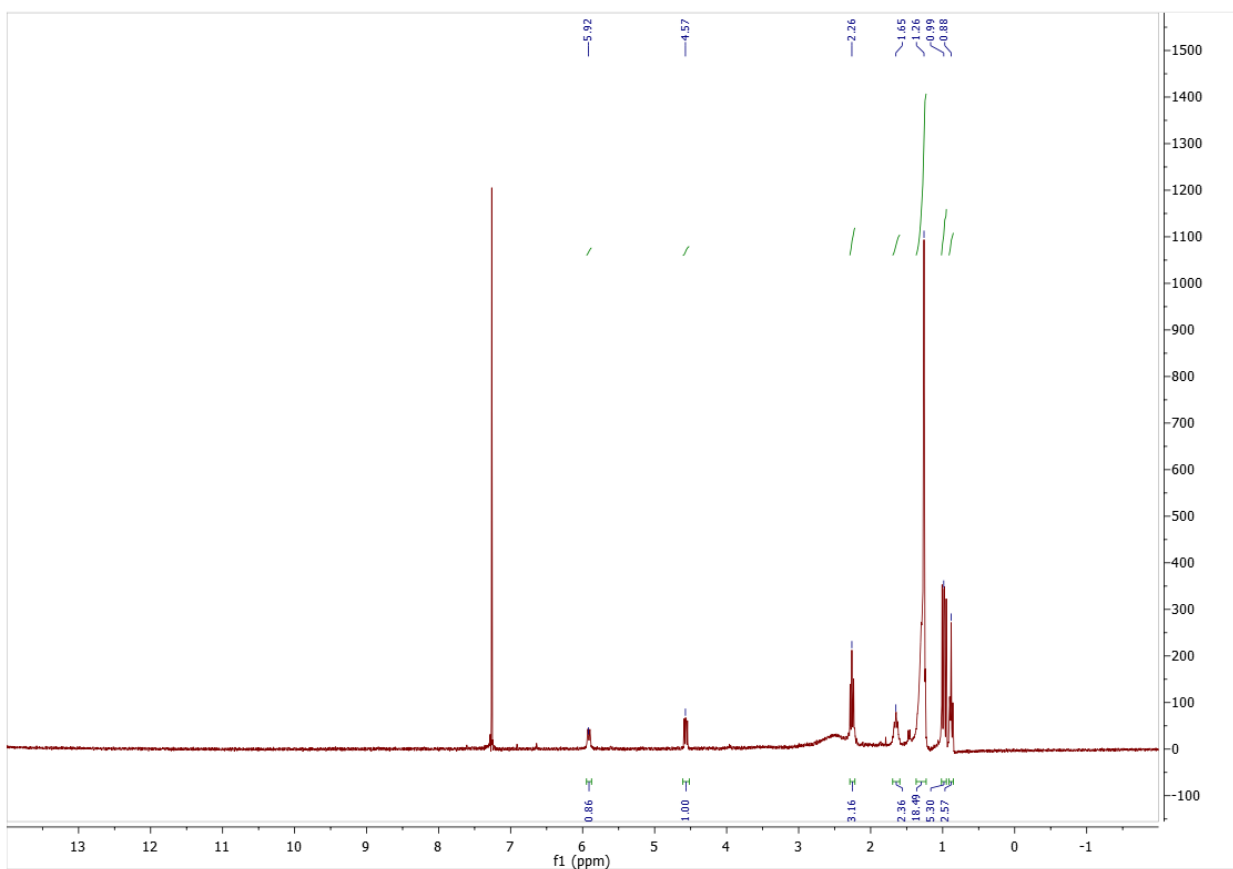
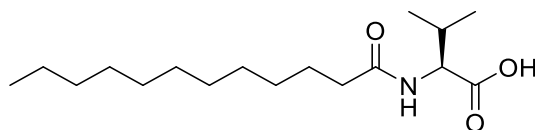


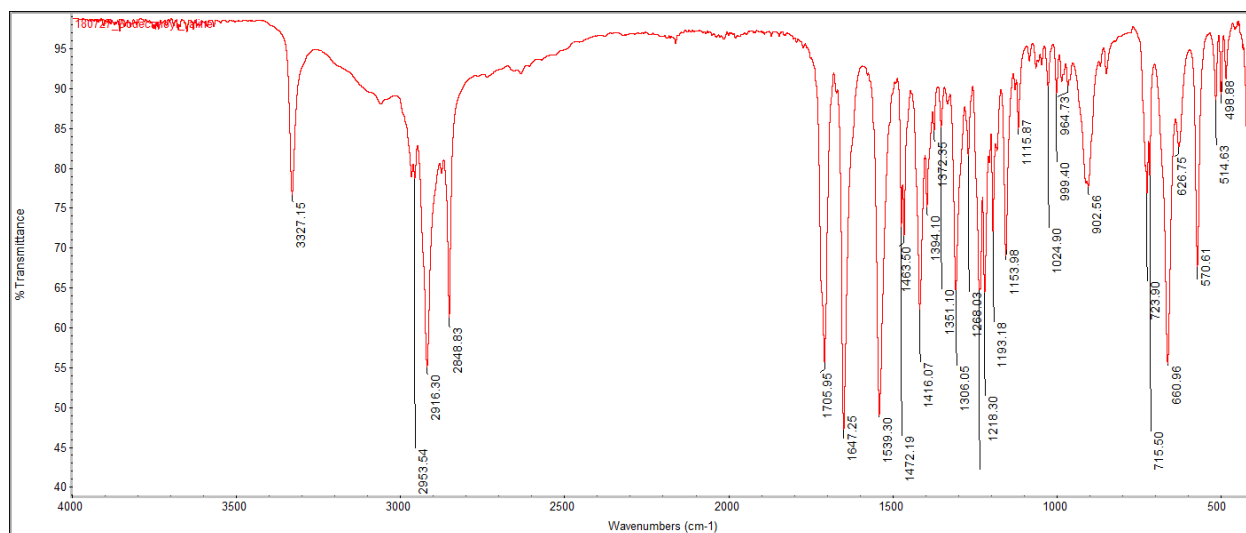
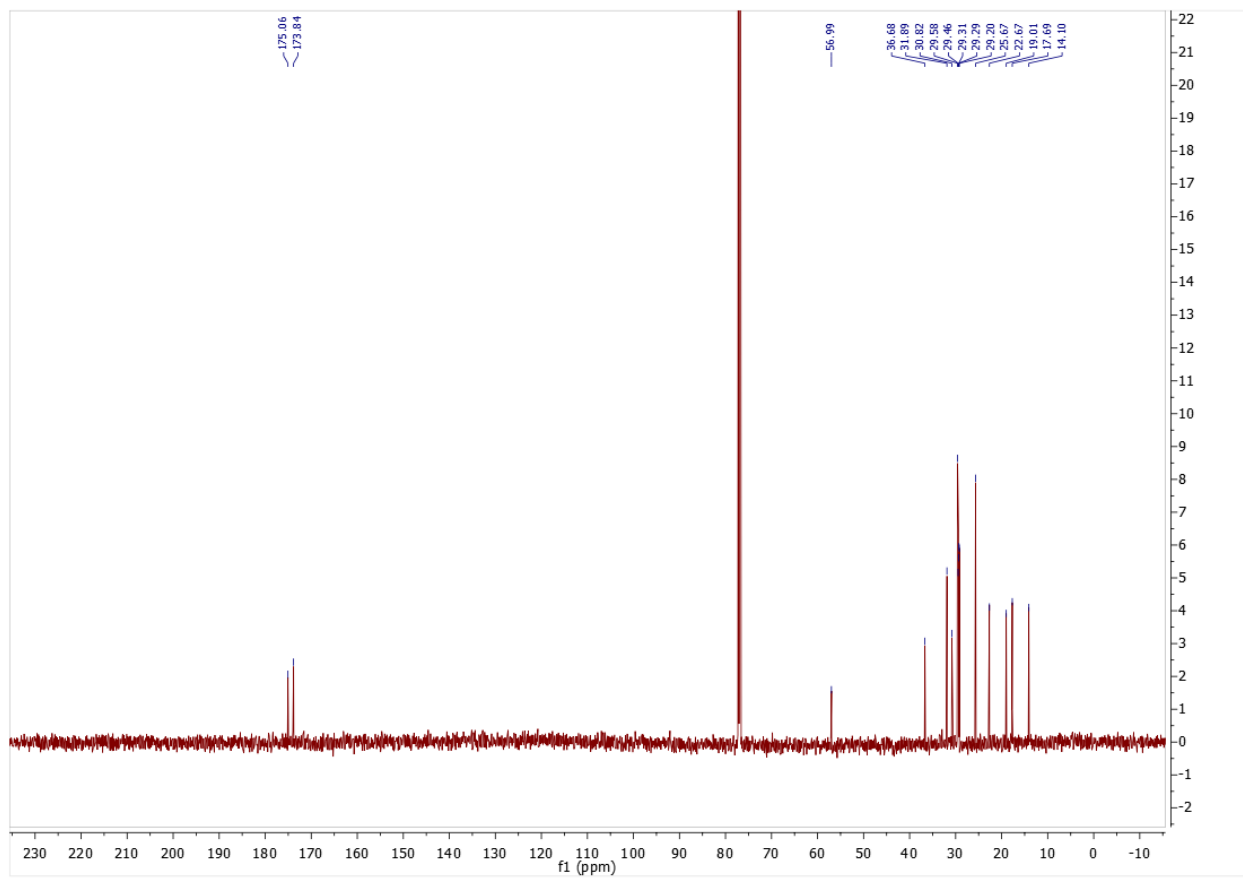
Dodecanoyl leucine: ^1H NMR (300 MHz, CDCl_3) δ 0.88 (t, 3H), 0.97 (m, 5H), 1.26 (m, 17H), 1.30-1.62 (m, 6H), 2.25 (t, 2H), 4.57 (m, 1H), 5.76 (m, 1H), 7.17. ^{13}C NMR (600 MHz, CDCl_3): δ 14.09, 21.86, 22.66, 22.82, 24.88, 25.60, 29.19, 29.30, 29.32, 29.48, 29.59, 31.89, 36.48, 41.18, 50.87, 174.06, 176.56. HRMS (ESI) m/z calculated for $\text{C}_{18}\text{H}_{35}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 312.2554, found: 312.2545. IR shown below.



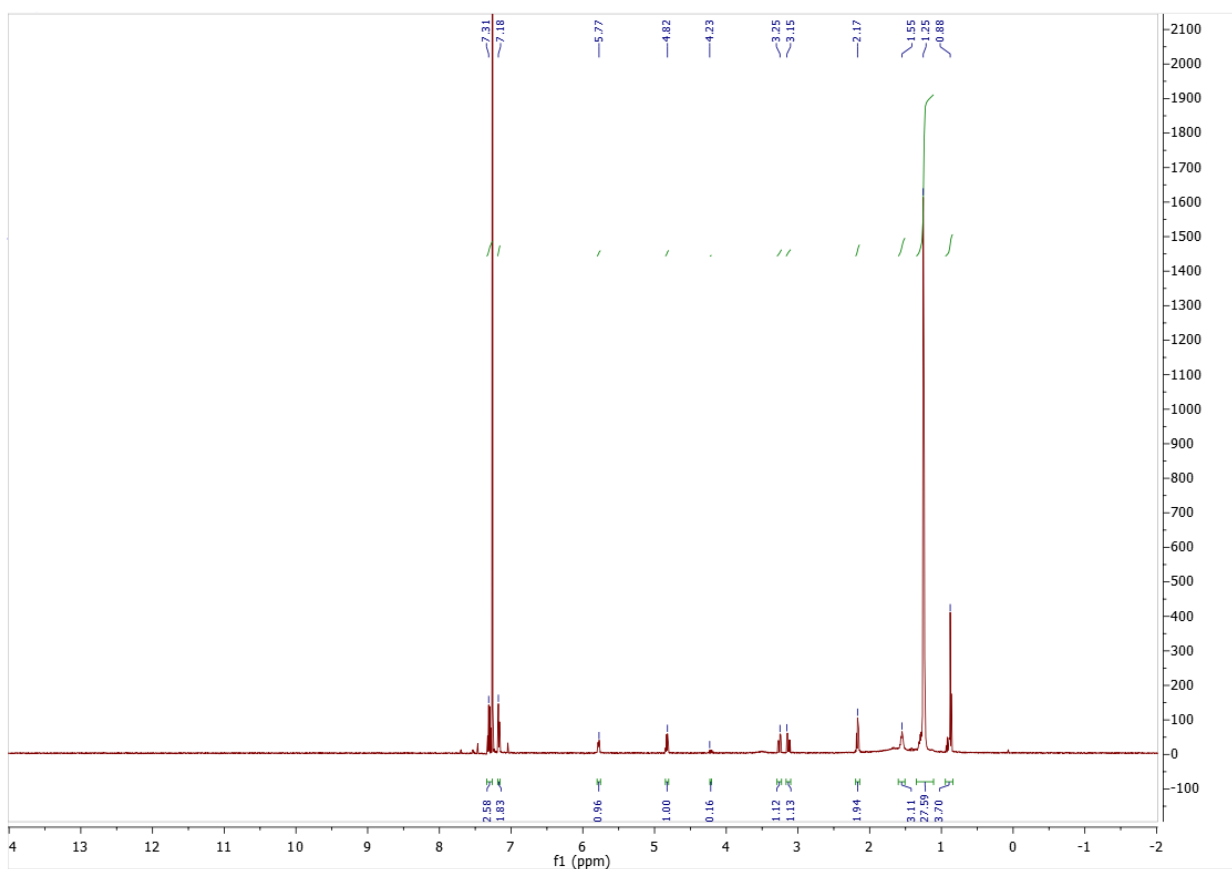
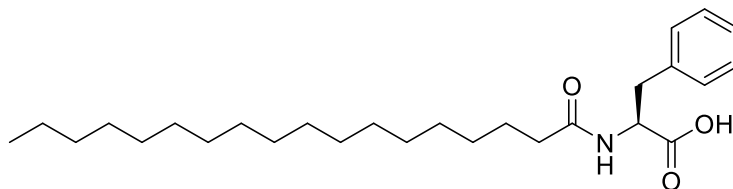


Dodecanoyl valine: ^1H NMR (300 MHz, CDCl_3) δ 0.88 (t, 3H), 0.99 (m, 5H), 1.26 (m, 18H), 1.65 (t, 2H), 2.26 (m, 3H), 4.57 (m, 1H), 5.92 (d, 1H). ^{13}C NMR (600 MHz, CDCl_3): δ 14.10, 17.69, 19.01, 22.67, 25.67, 29.20, 29.29, 29.31, 29.46, 29.58, 30.82, 31.89, 36.68, 56.99, 173.84, 175.06. HRMS (ESI) m/z calculated for $\text{C}_{17}\text{H}_{33}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 298.2388, found: 298.2385. IR shown below.

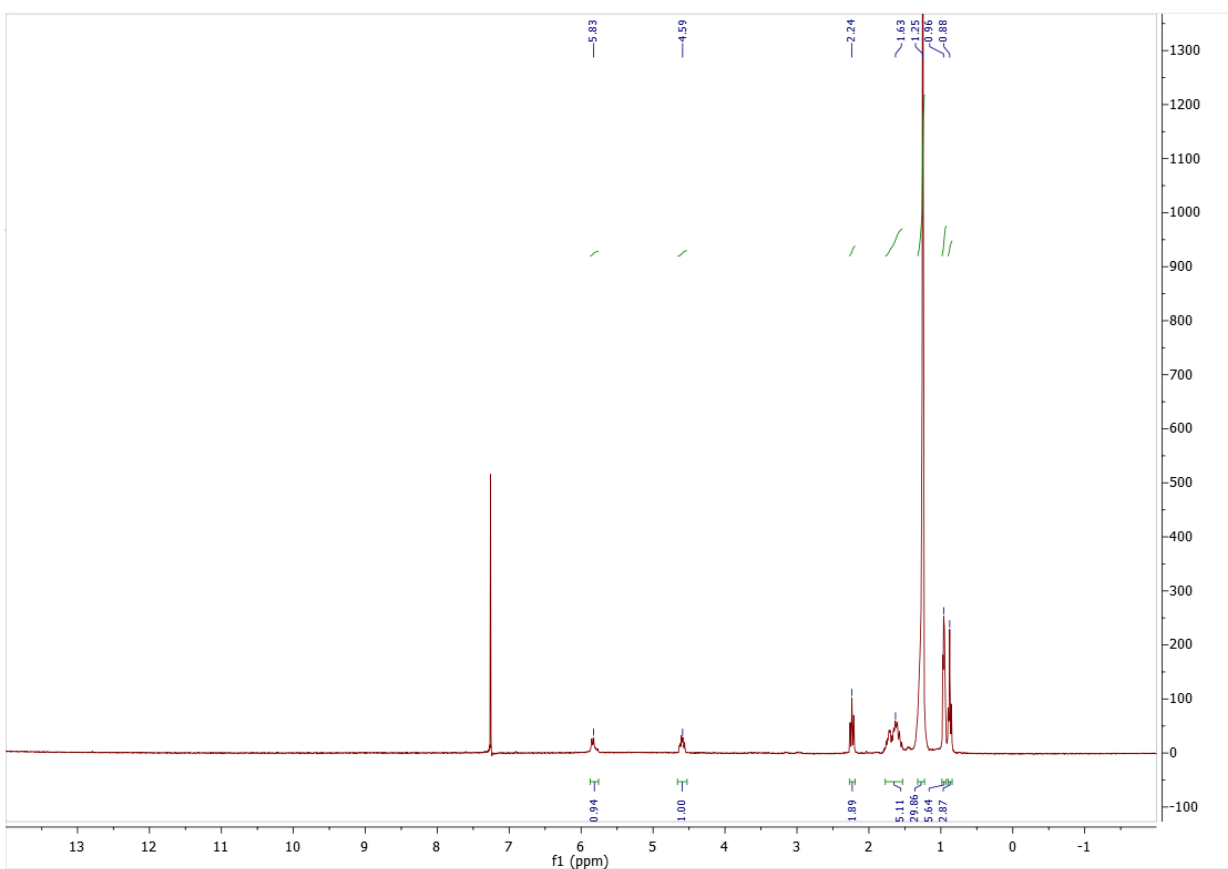
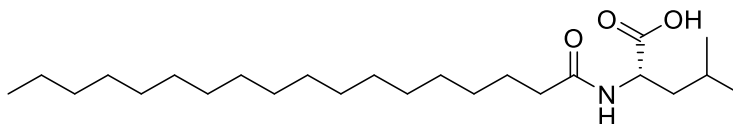


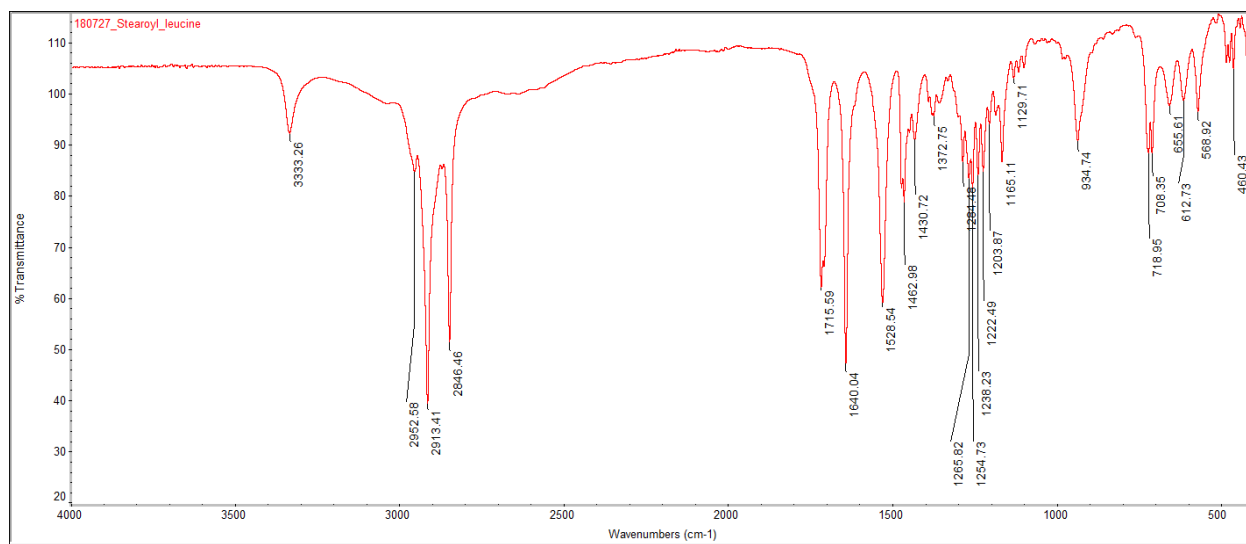
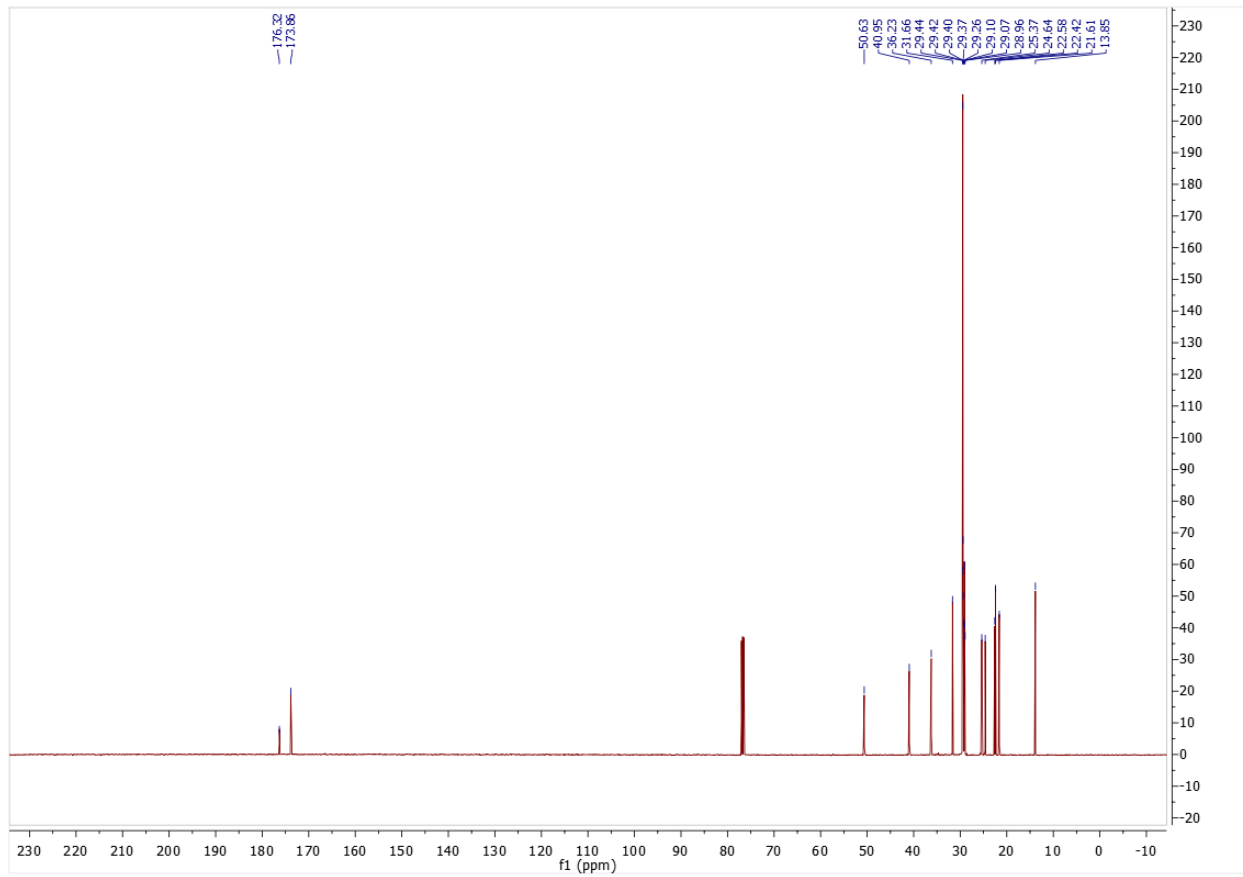


Stearoyl phenylalanine: ^1H NMR (500 MHz, CDCl_3) δ 0.88 (t, 3H), 1.25 (m, 28H), 1.55 (m, 3H), 2.17 (t, 2H), 3.15 (dd, 1H), 3.25 (dd, 1H), 4.82 (m, 1H), 5.77 (d, 1H) 7.18 (m, 2H), 7.31 (m, 3H). ^{13}C NMR (500 MHz, CDCl_3): δ 13.86, 22.43, 25.29, 28.90, 29.05, 29.11, 29.22, 29.38, 29.41, 29.42, 29.45, 31.67, 36.20, 36.99, 52.86, 126.97, 128.38, 129.09, 135.41, 173.74, 174.37. HRMS (ESI) m/z calculated for $\text{C}_{27}\text{H}_{45}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 430.3327, found: 430.3321. IR shown below.

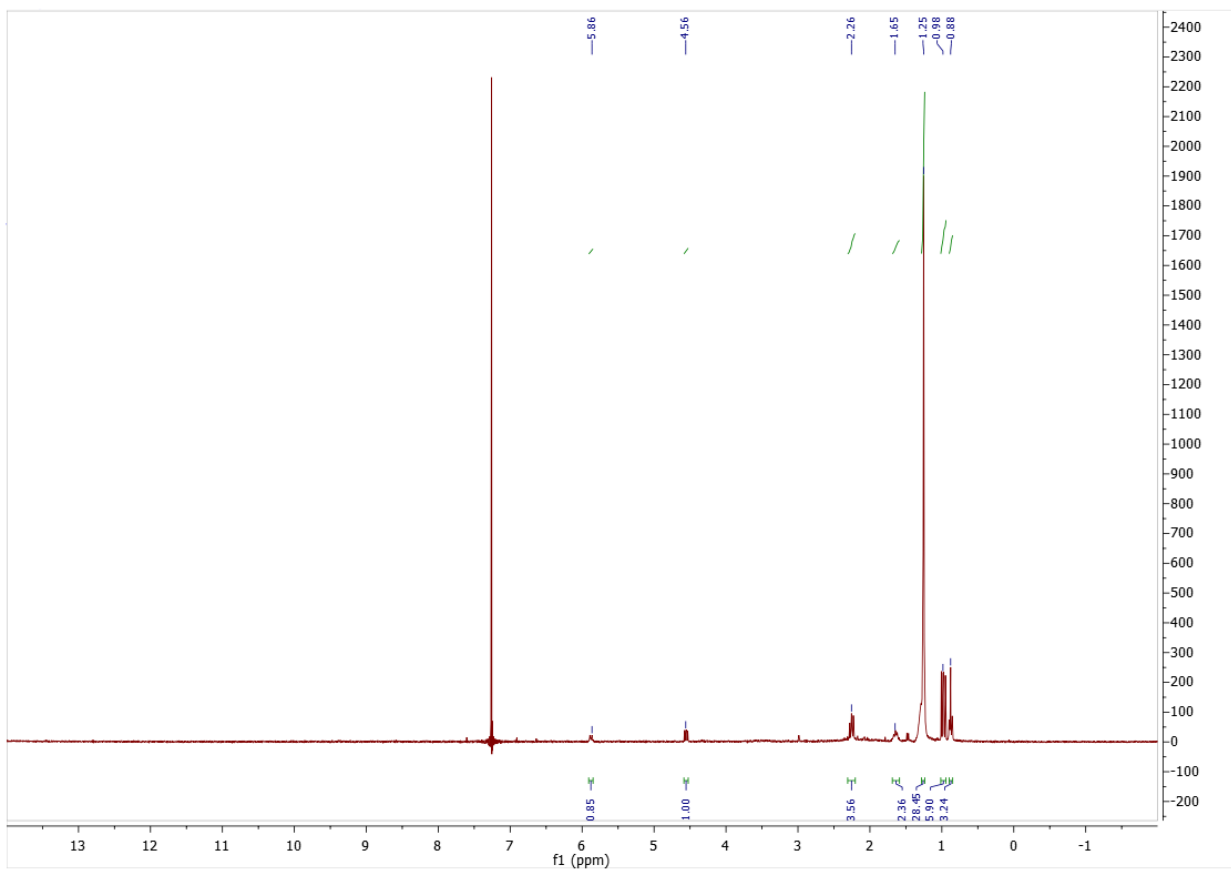
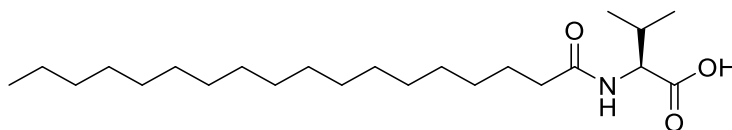


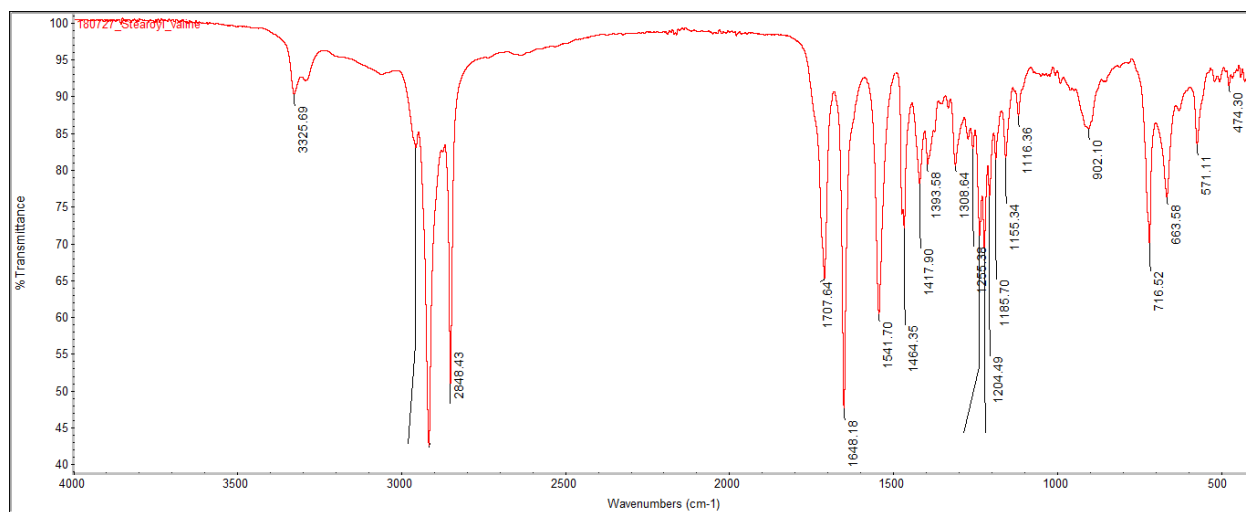
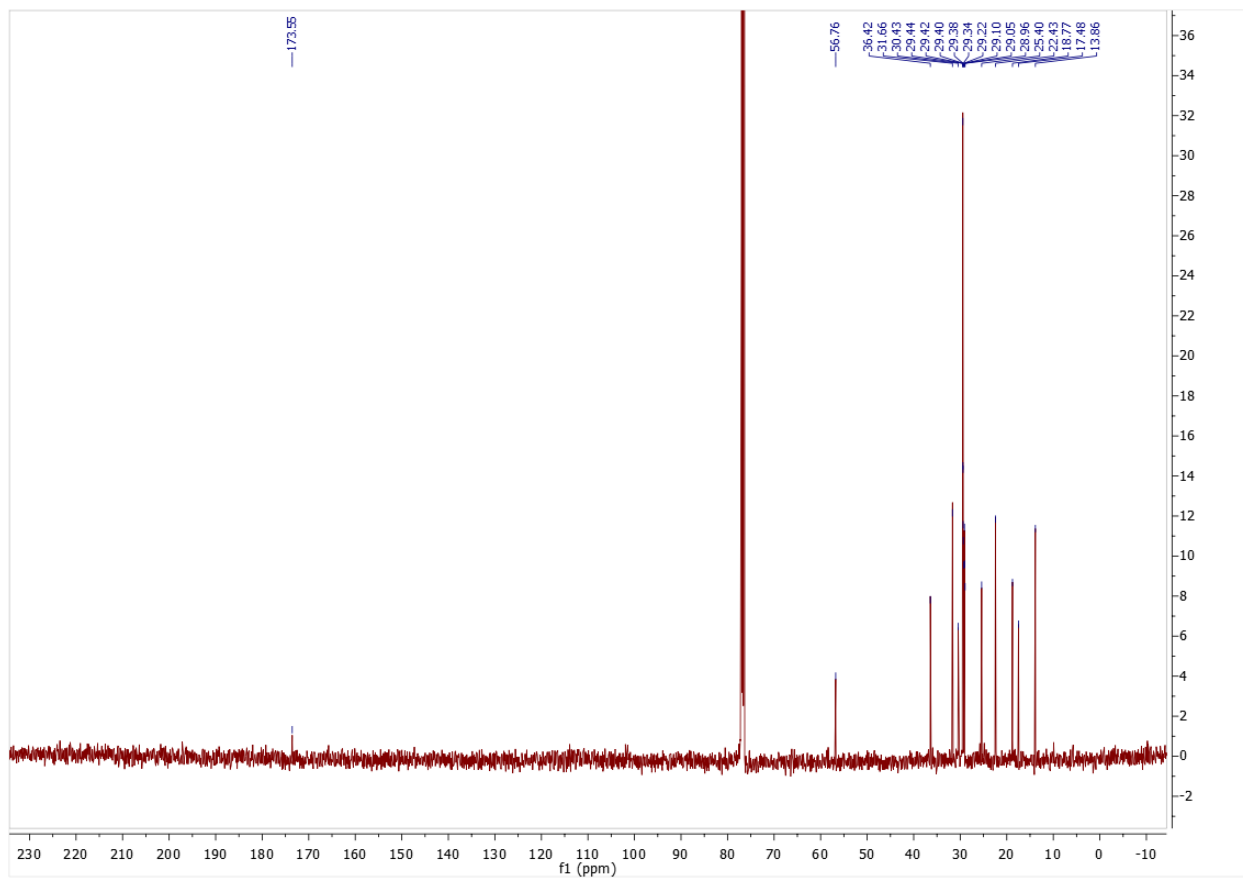
Stearoyl leucine: ^1H NMR (300 MHz, CDCl_3) δ 0.88 (t, 3H), 0.96 (m, 5H), 1.25 (m, 30H), 1.63 (m, 5H), 2.24 (t, 2H), 4.59 (m, 1H), 5.83 (m, 1H). ^{13}C NMR (500 MHz, CDCl_3): δ 13.85, 21.61, 22.42, 22.58, 24.64, 25.37, 28.96, 29.07, 29.10, 29.26, 29.37, 29.40, 29.42, 29.44, 31.66, 36.23, 40.95, 50.63, 173.86, 176.32. HRMS (ESI) m/z calculated for $\text{C}_{24}\text{H}_{47}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 396.3483, found: 396.3487. IR shown below.



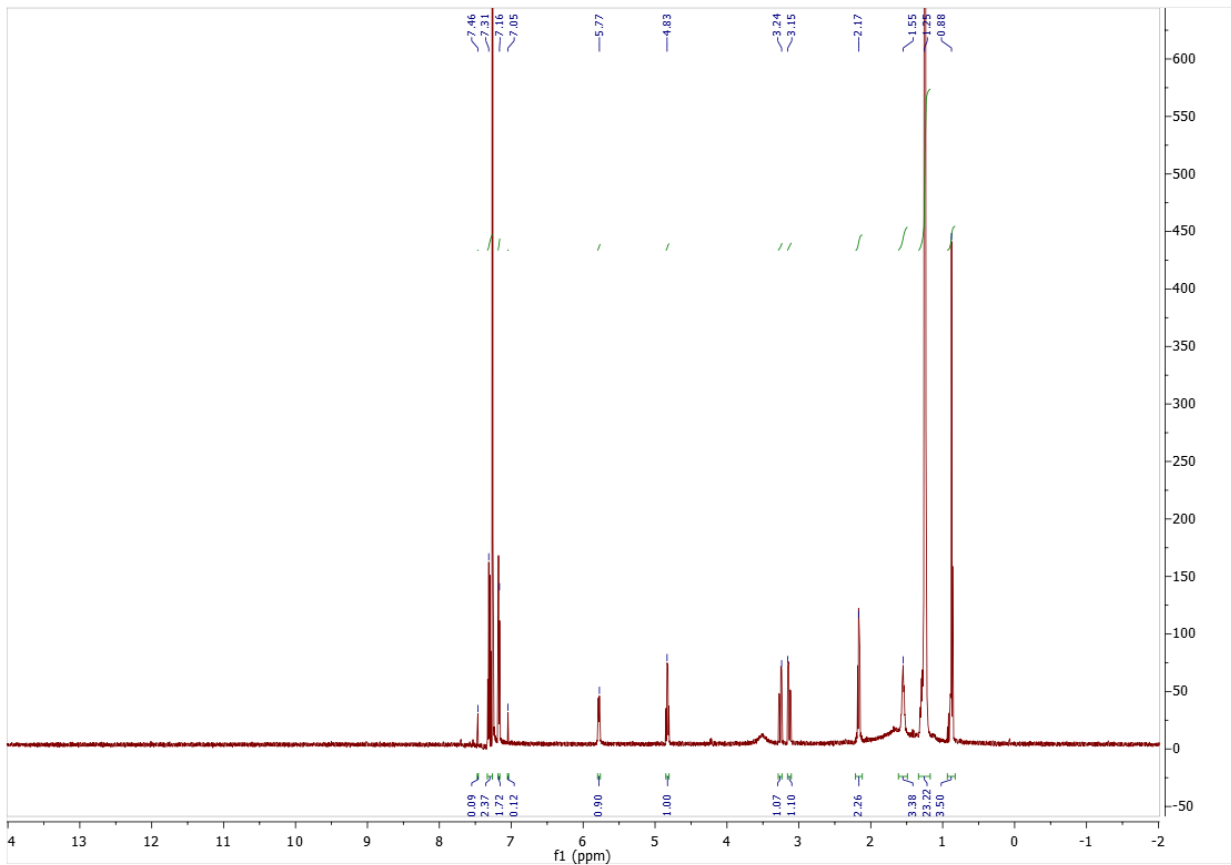
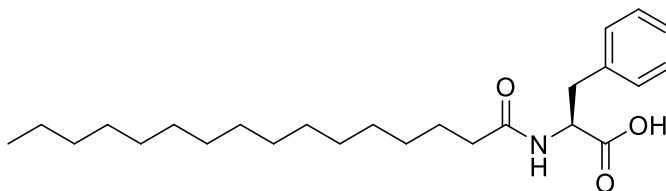


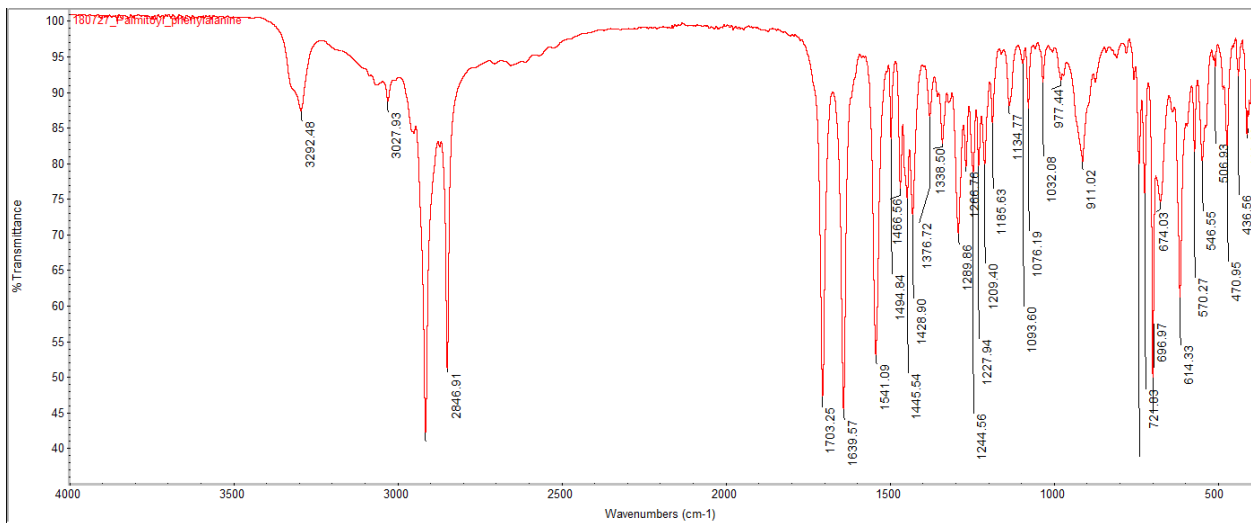
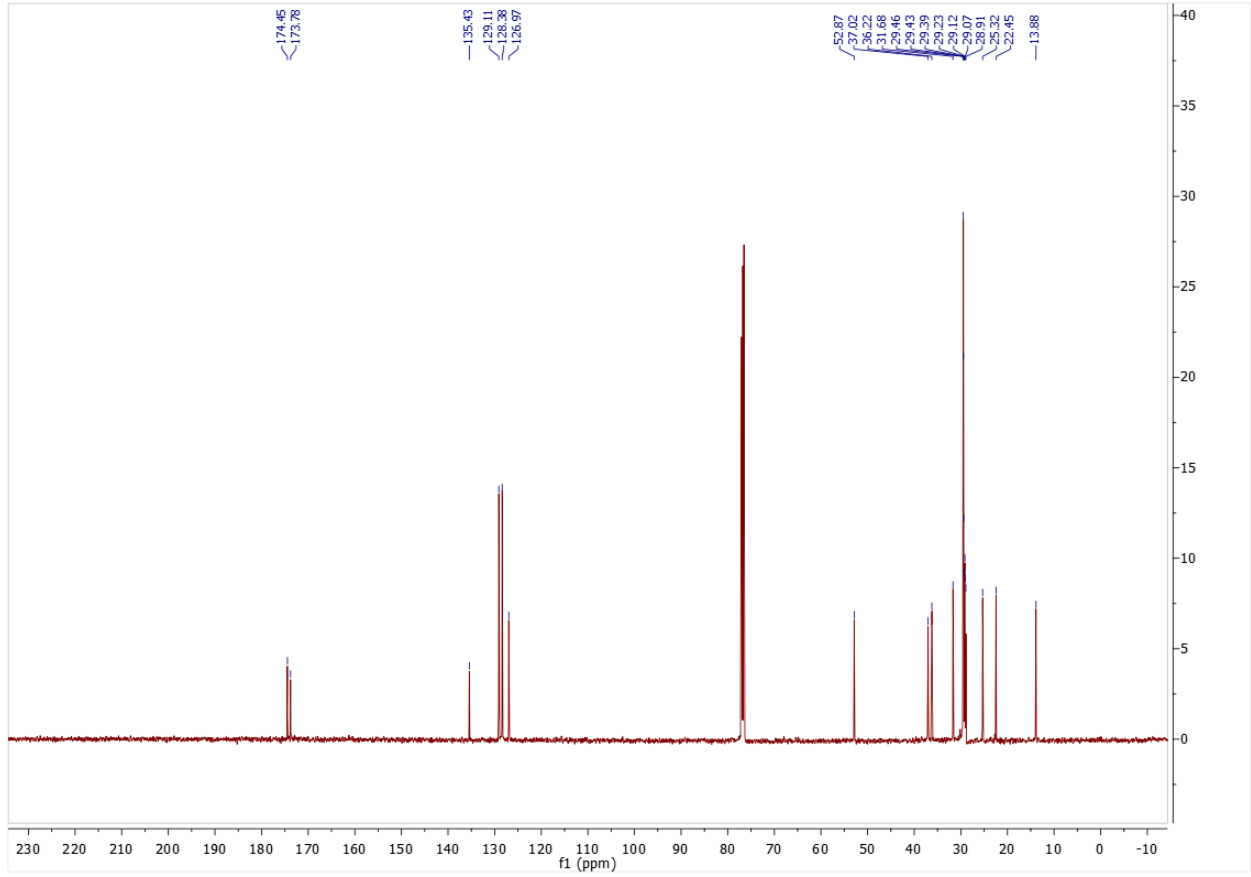
Stearoyl valine: ^1H NMR (300 MHz, CDCl_3) δ 0.88 (t, 3H), 0.98 (m, 6H), 1.25 (m, 28H), 1.65 (t, 2H), 2.26 (m, 4H), 4.56 (m, 1H), 5.86 (d, 1H). ^{13}C NMR (500 MHz, CDCl_3): δ 13.86, 17.48, 18.77, 22.43, 25.40, 28.96, 29.05, 29.10, 29.22, 29.34, 29.38, 29.40, 29.42, 29.44, 30.43, 31.66, 36.42, 56.78, 173.55. HRMS (ESI) m/z calculated for $\text{C}_{23}\text{H}_{45}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 382.3387, found: 382.3327. IR shown below.



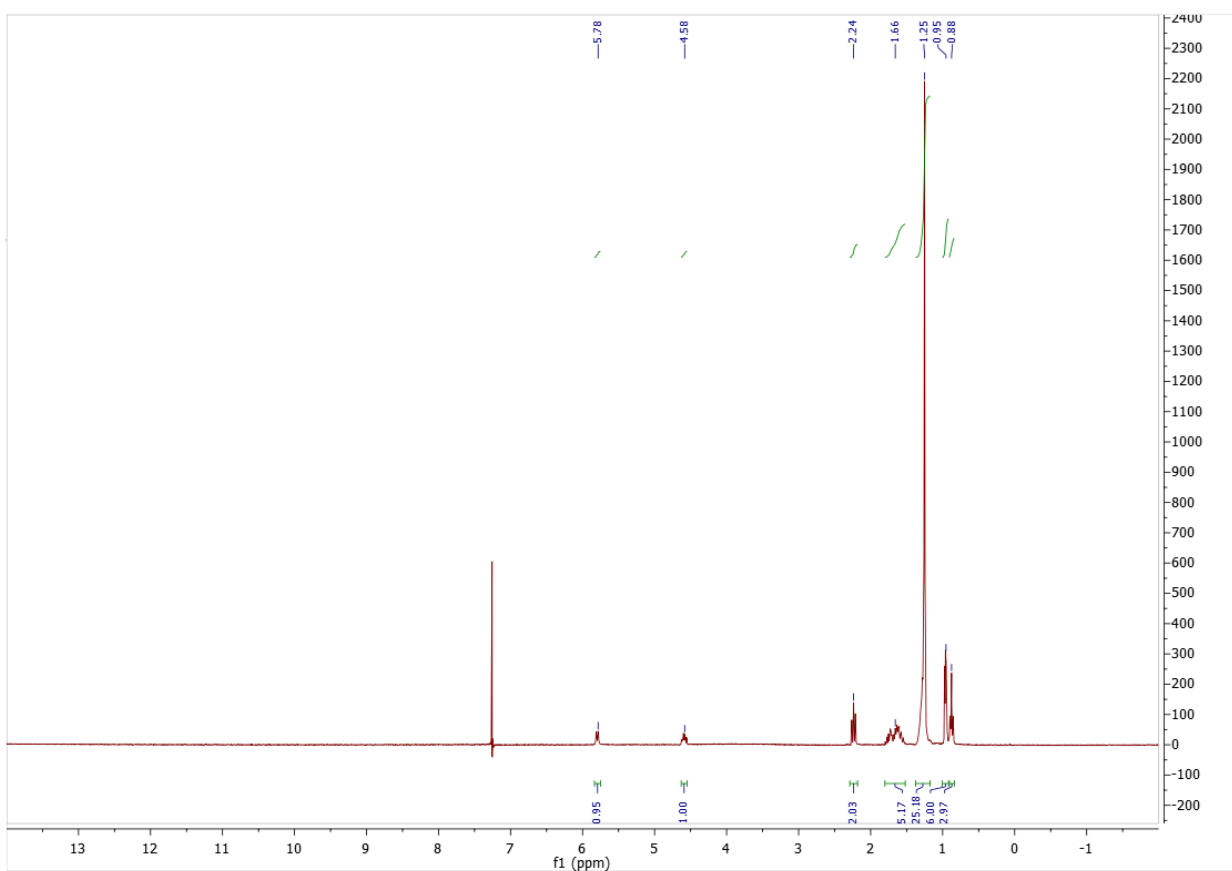
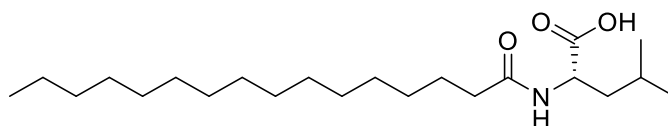


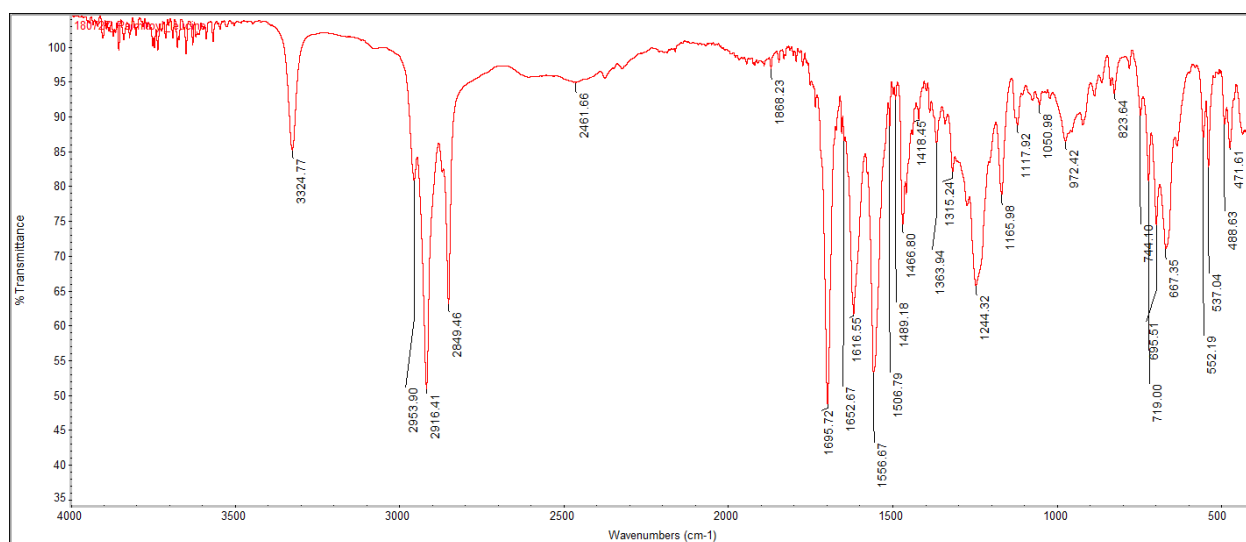
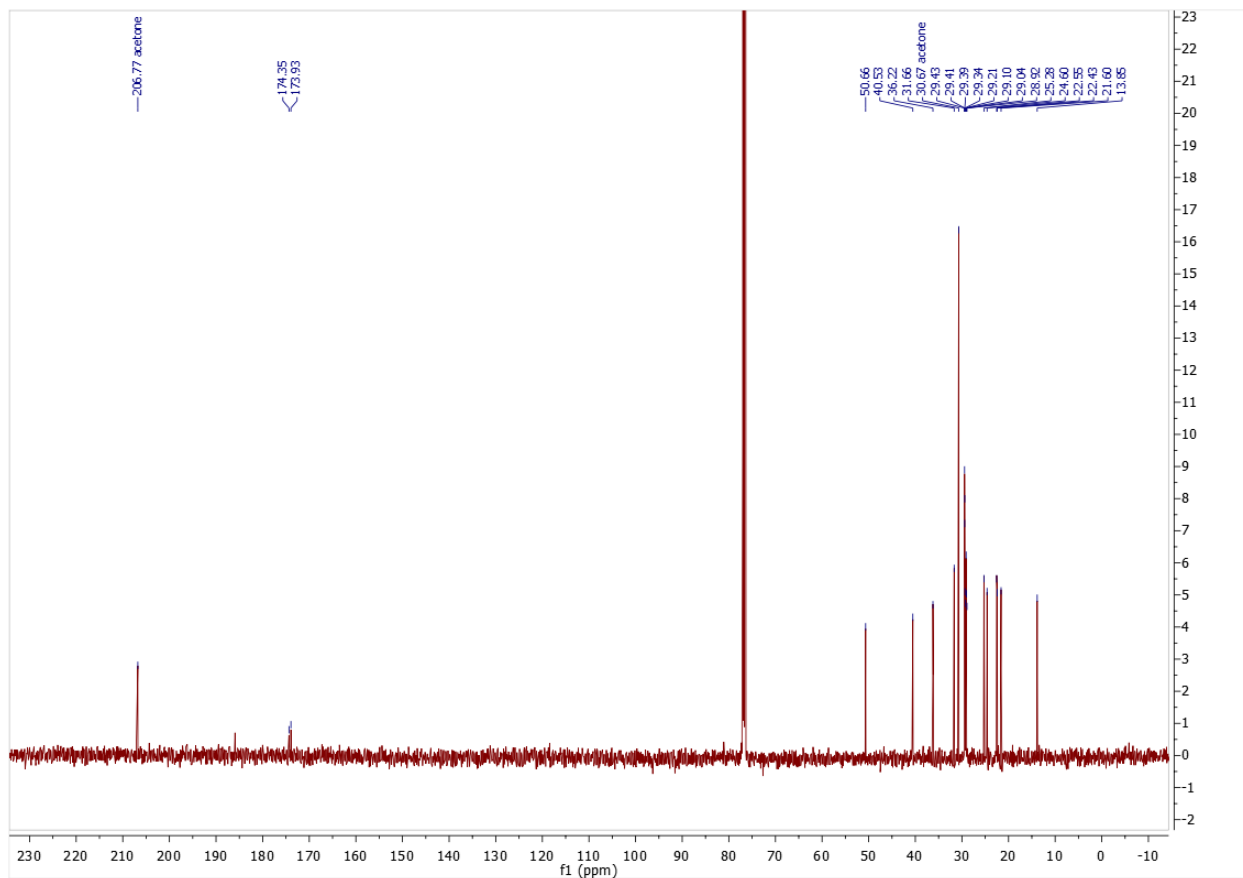
Palmitoyl phenylalanine: ^1H NMR (500 MHz, CDCl_3) δ 0.88 (t, 3H), 1.25 (m, 24H), 1.55 (m, 3H), 2.17 (t, 2H), 3.15 (dd, 1H), 3.24 (dd, 1H), 4.83 (m, 1H), 5.77 (d, 1H) 7.16 (m, 2H), 7.31 (m, 3H). ^{13}C NMR (500 MHz, CDCl_3): δ 13.88, 22.45, 25.32, 28.91, 29.07, 29.12, 29.23, 29.39, 29.43, 29.46, 31.68, 36.22, 37.02, 52.87, 126.97, 128.38, 129.11, 135.43, 173.78, 174.45. HRMS (ESI) m/z calculated for $\text{C}_{25}\text{H}_{41}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 402.3014, found: 402.3014. IR shown below.



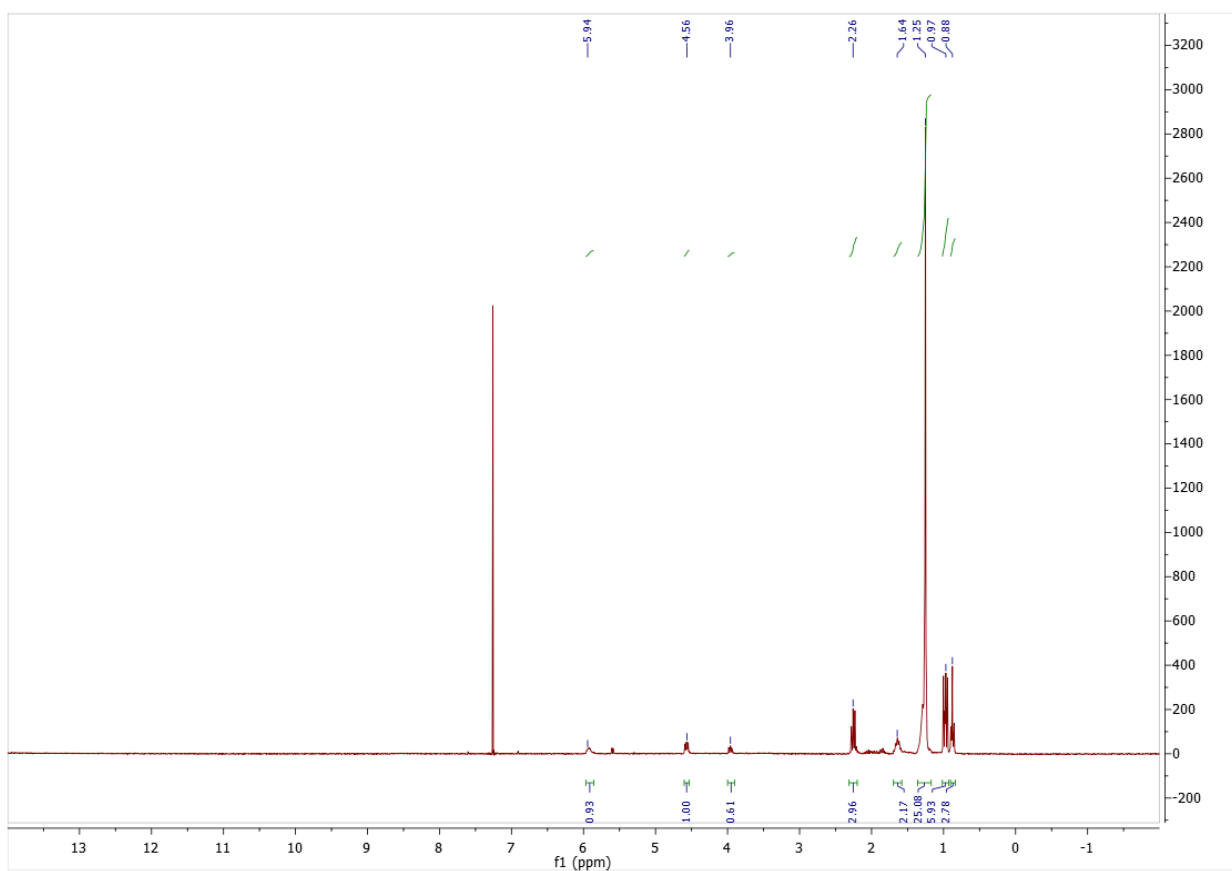
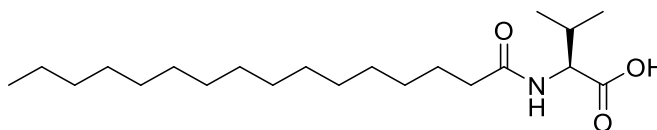


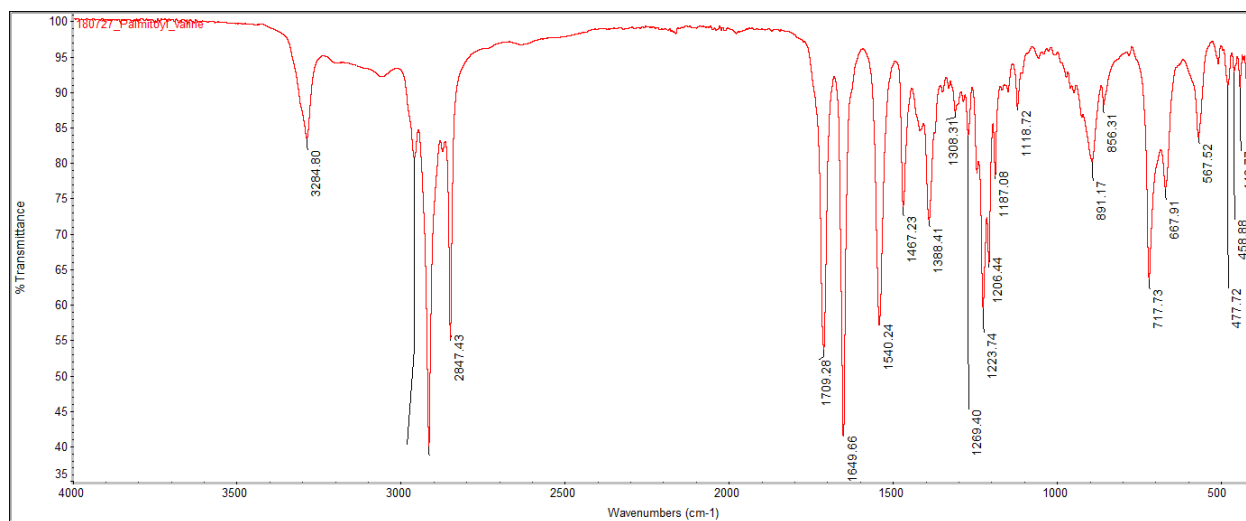
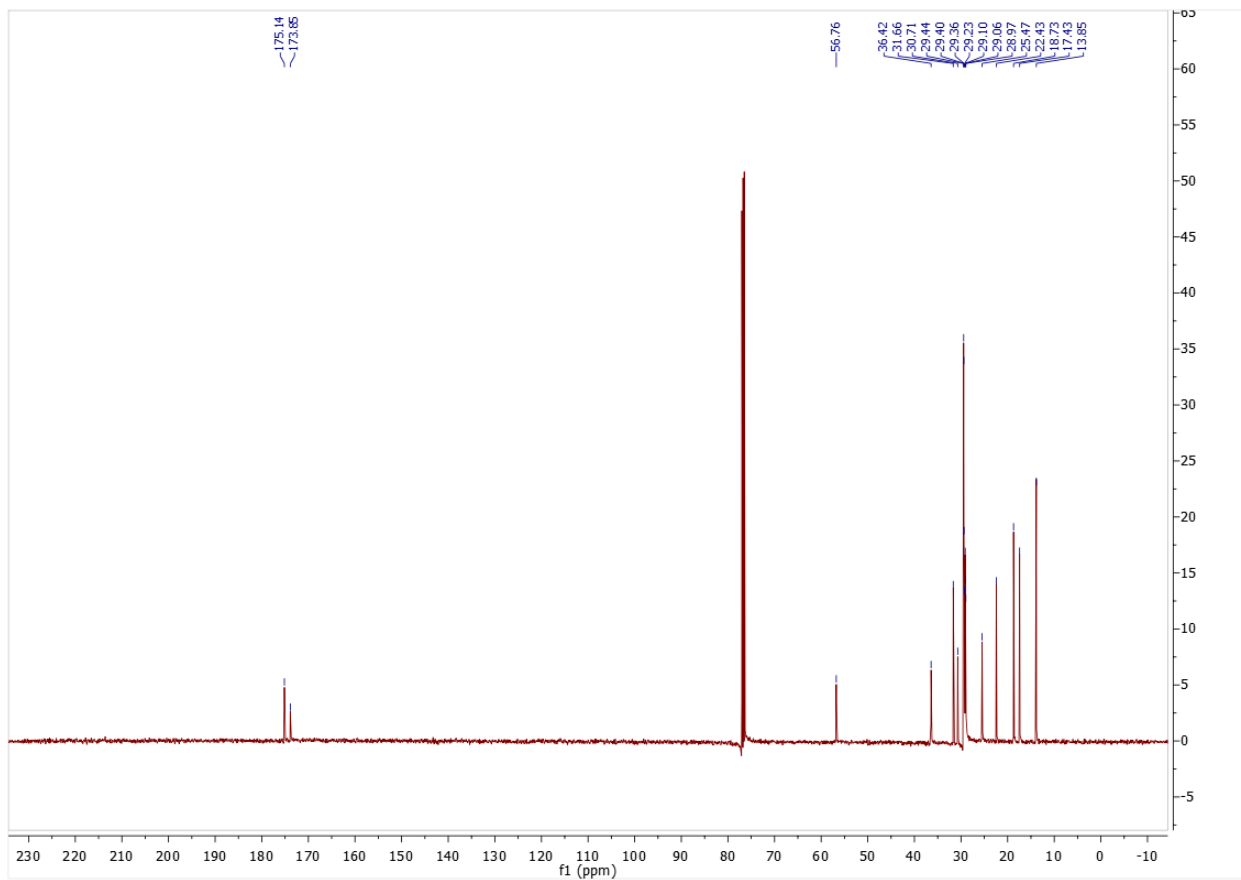
Palmitoyl leucine: ^1H NMR (300 MHz, CDCl_3) δ 0.88 (t, 3H), 0.95 (m, 6H), 1.25 (m, 25H), 1.66 (m, 5H), 2.24 (t, 2H), 4.58 (m, 1H), 5.78 (m, 1H). ^{13}C NMR (500 MHz, CDCl_3): δ 13.85, 21.60, 22.43, 22.55, 24.60, 25.28, 28.92, 29.04, 29.10, 29.21, 29.34, 29.39, 29.41, 29.43, 31.66, 36.22, 40.53, 50.66, 173.93, 174.35. HRMS (ESI) m/z calculated for $\text{C}_{22}\text{H}_{43}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 368.3170, found: 368.3174. IR shown below.



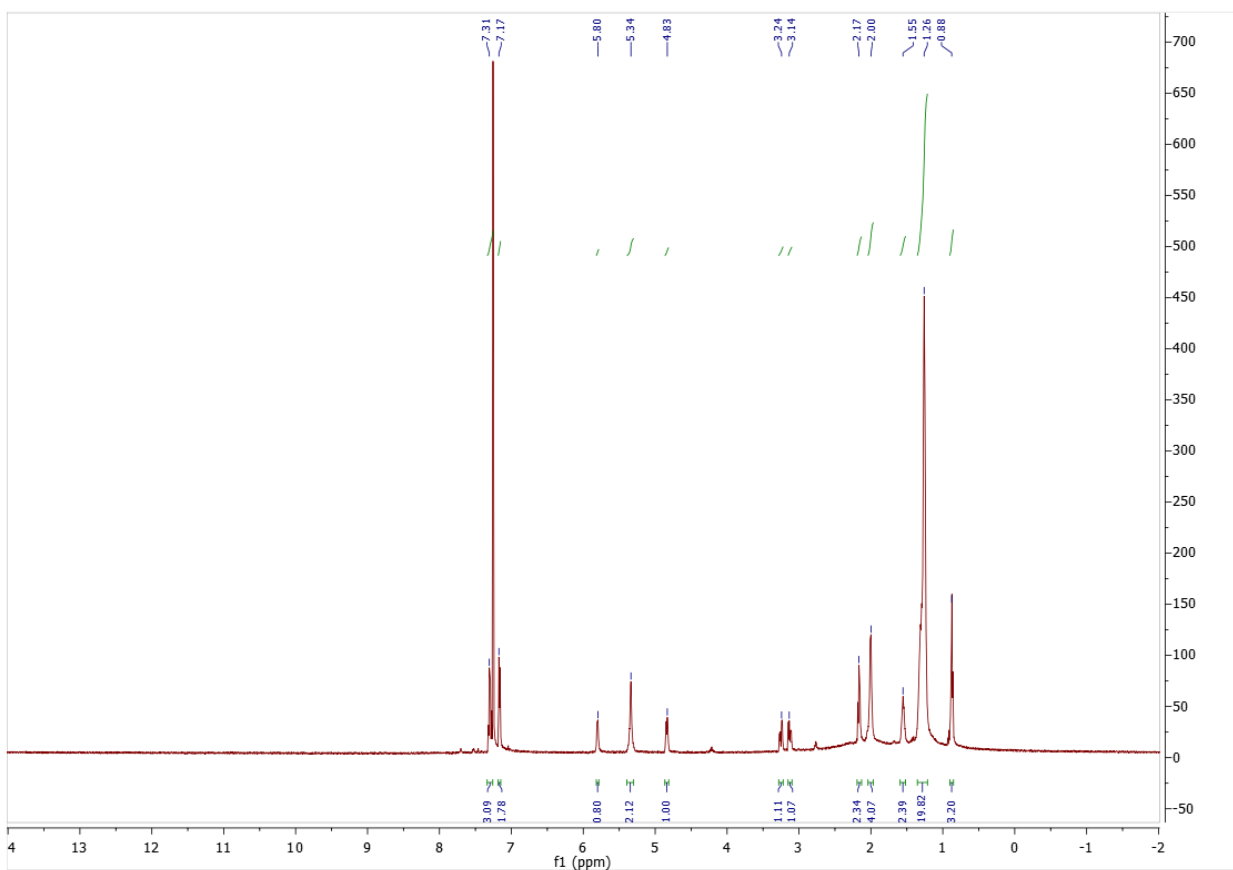
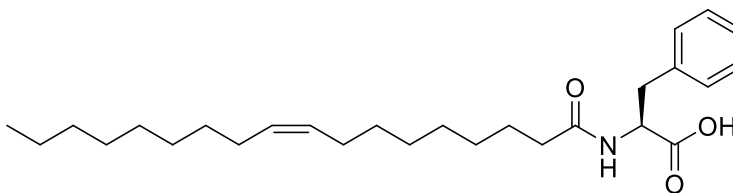


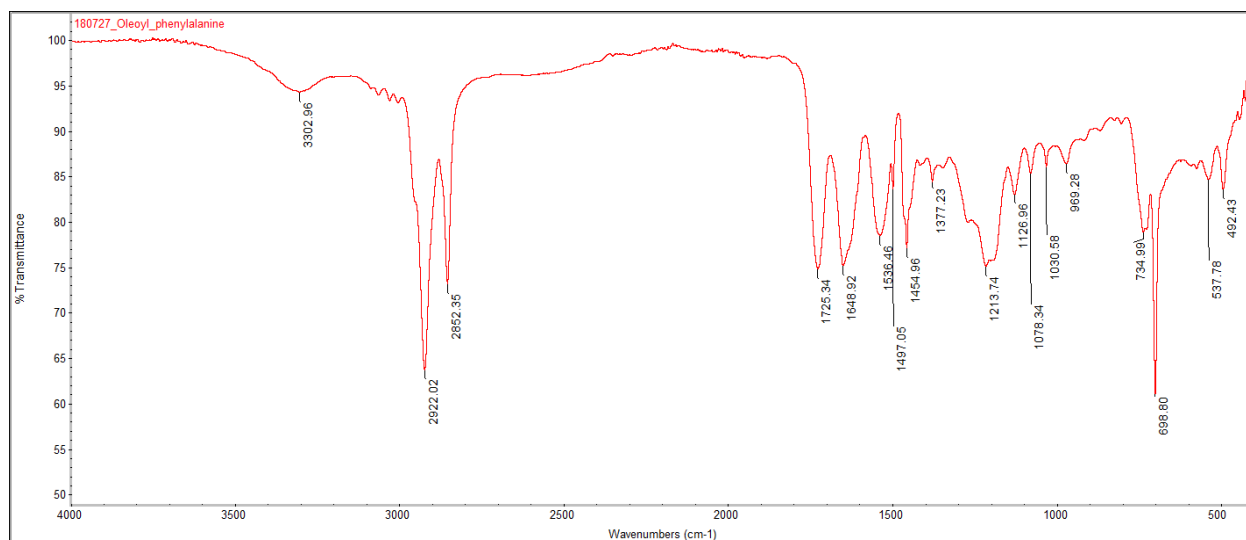
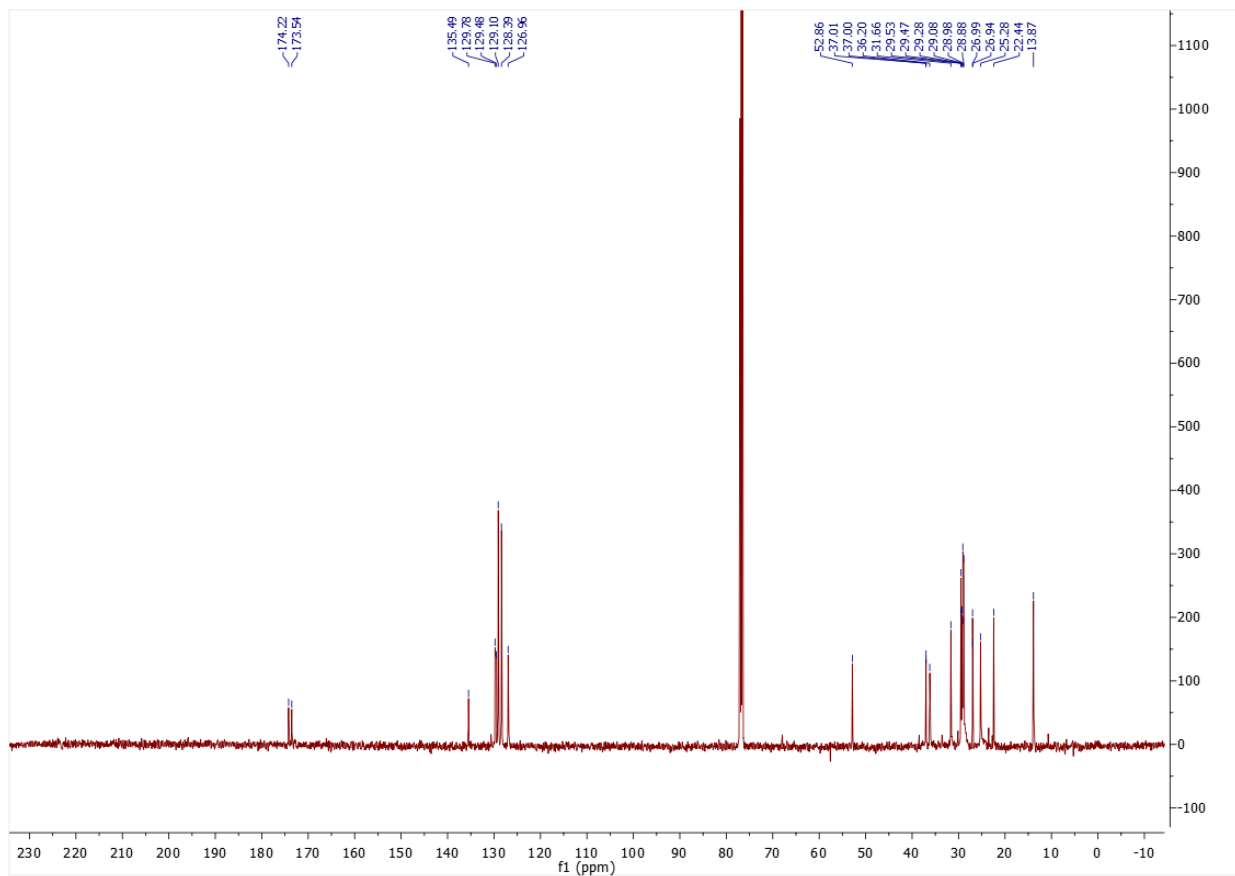
Palmitoyl valine: ^1H NMR (300 MHz, CDCl_3) δ 0.88 (t, 3H), 0.97 (m, 6H), 1.25 (m, 25H), 1.64 (t, 2H), 2.26 (m, 3H), 4.56 (m, 1H), 5.94 (d, 1H). ^{13}C NMR (500 MHz, CDCl_3): δ 13.85, 17.43, 18.73, 22.43, 25.47, 28.97, 29.06, 29.10, 29.23, 29.36, 29.40, 29.44, 30.71, 31.66, 36.42, 56.76, 173.85, 175.14. HRMS (ESI) m/z calculated for $\text{C}_{21}\text{H}_{41}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 354.3014, found: 354.3018. IR shown below.



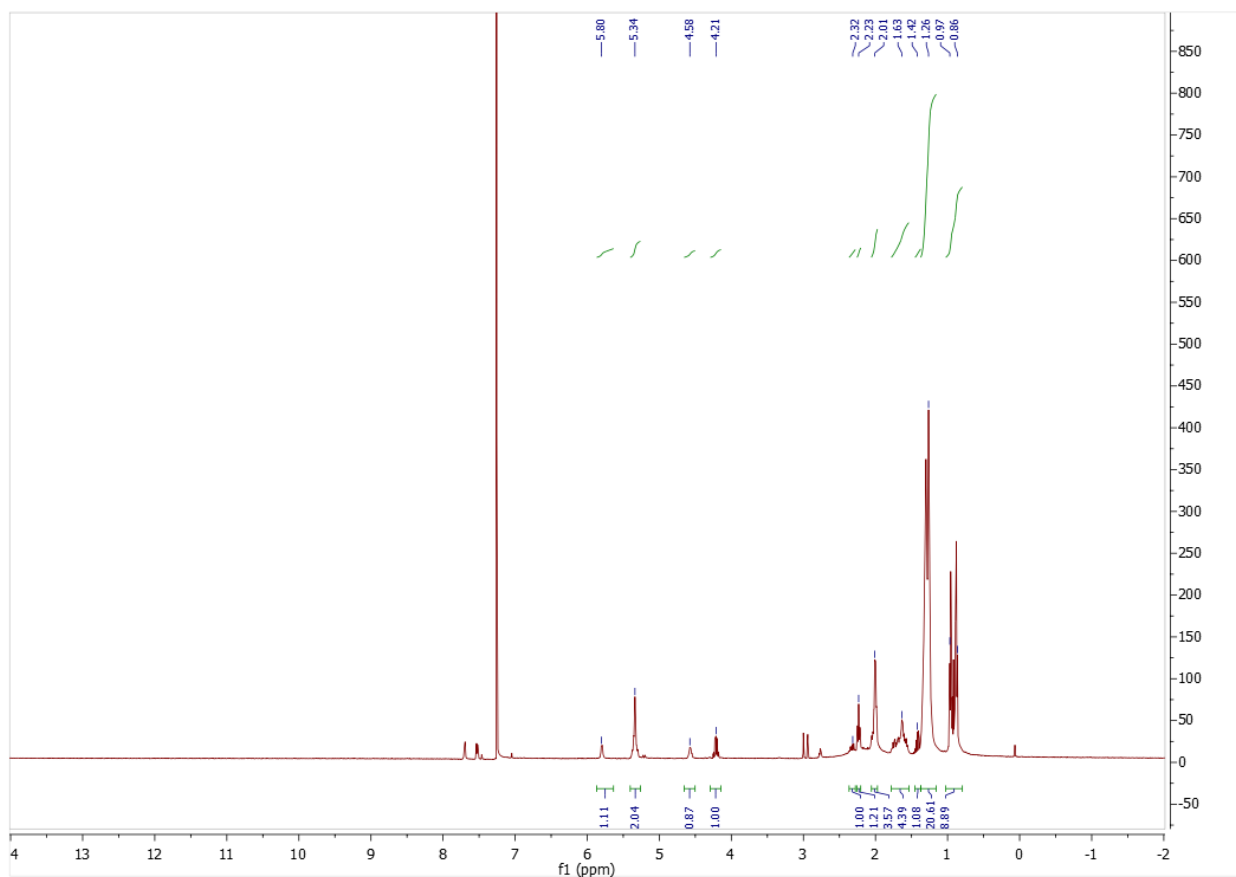
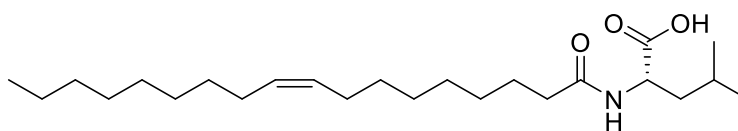


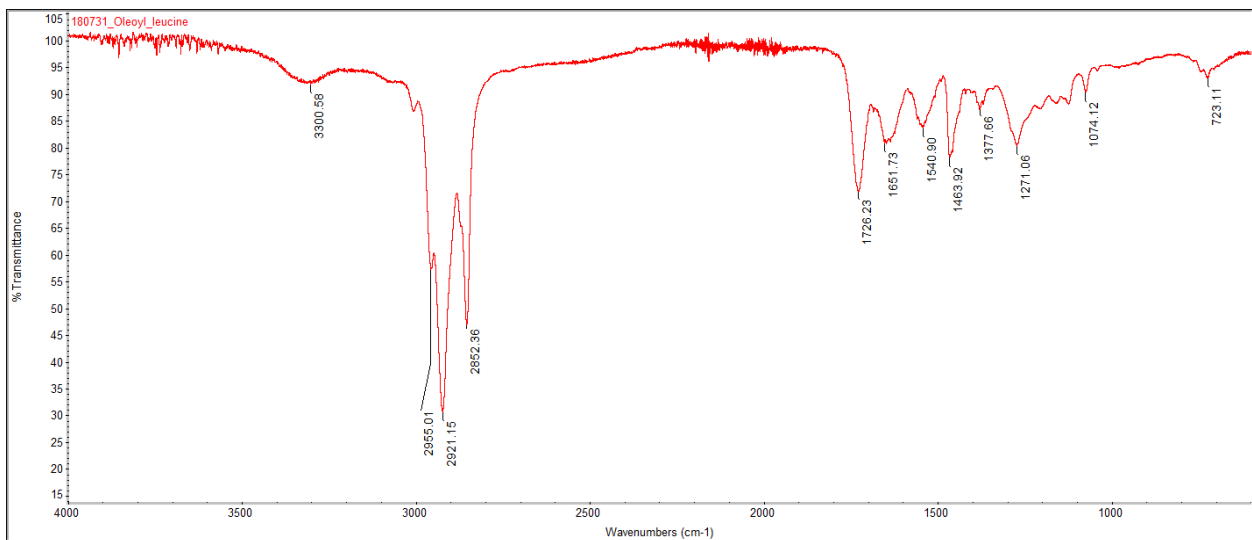
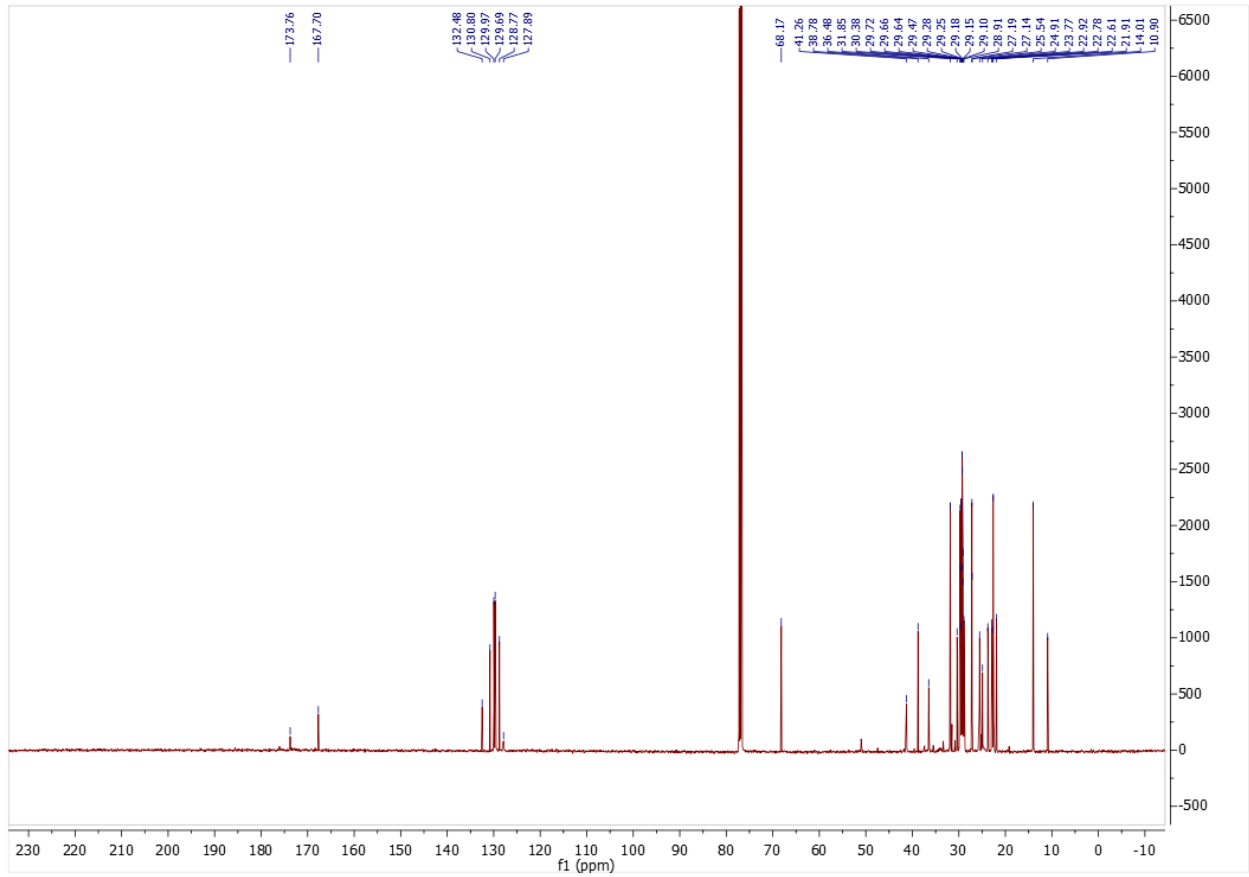
Oleoyl phenylalanine: ^1H NMR (300 MHz, CDCl_3) δ 0.88 (t, 3H), 1.26 (m, 20H), 1.55 (m, 2H), 2.00 (m, 4H), 2.17 (t, 2H), 3.14 (dd, 1H), 3.24 (dd, 1H), 4.83 (m, 1H), 5.34 (m, 2H), 5.80 (d, 1H), 7.17 (m, 2H), 7.31 (m, 3H). ^{13}C NMR (500 MHz, CDCl_3): δ 13.87, 22.44, 25.28, 26.94, 26.99, 28.88, 28.98, 29.08, 29.28, 29.47, 29.53, 31.66, 36.20, 37.00, 52.86, 126.98, 128.39, 129.10, 129.48, 129.78, 135.49, 173.54, 174.22. HRMS (ESI) m/z calculated for $\text{C}_{27}\text{H}_{43}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 428.3170, found: 428.3164. IR shown below.





Oleoyl leucine: ^1H NMR (300 MHz, CDCl_3) δ 0.86- 0.96 (m, 9H), 1.26 (m, 20H), 1.42 (m, 1H), 1.63 (m, 4H), 2.01 (m, 3H), 2.23 (m, 1H), 2.32 (m, 1H), 4.21 (m, 1H), 4.58 (q, 1H), 5.34 (m, 2H), 5.79 (m, 1H), 5.80 (d, 1H). ^{13}C NMR (500 MHz, CDCl_3): δ 10.90, 14.01, 21.91, 22.61, 22.78, 22.92, 23.77, 24.91, 25.54, 27.14, 27.19, 28.91, 29.10, 29.15, 29.18, 29.25, 29.28, 29.47, 29.64, 29.66, 29.72, 30.38, 31.85, 36.48, 38.78, 41.26, 68.17, 127.89, 128.77, 129.97, 130.80, 132.48, 167.70, 173.76. HRMS (ESI) m/z calculated for $\text{C}_{24}\text{H}_{45}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 394.3327, found: 394.3322. IR shown below.

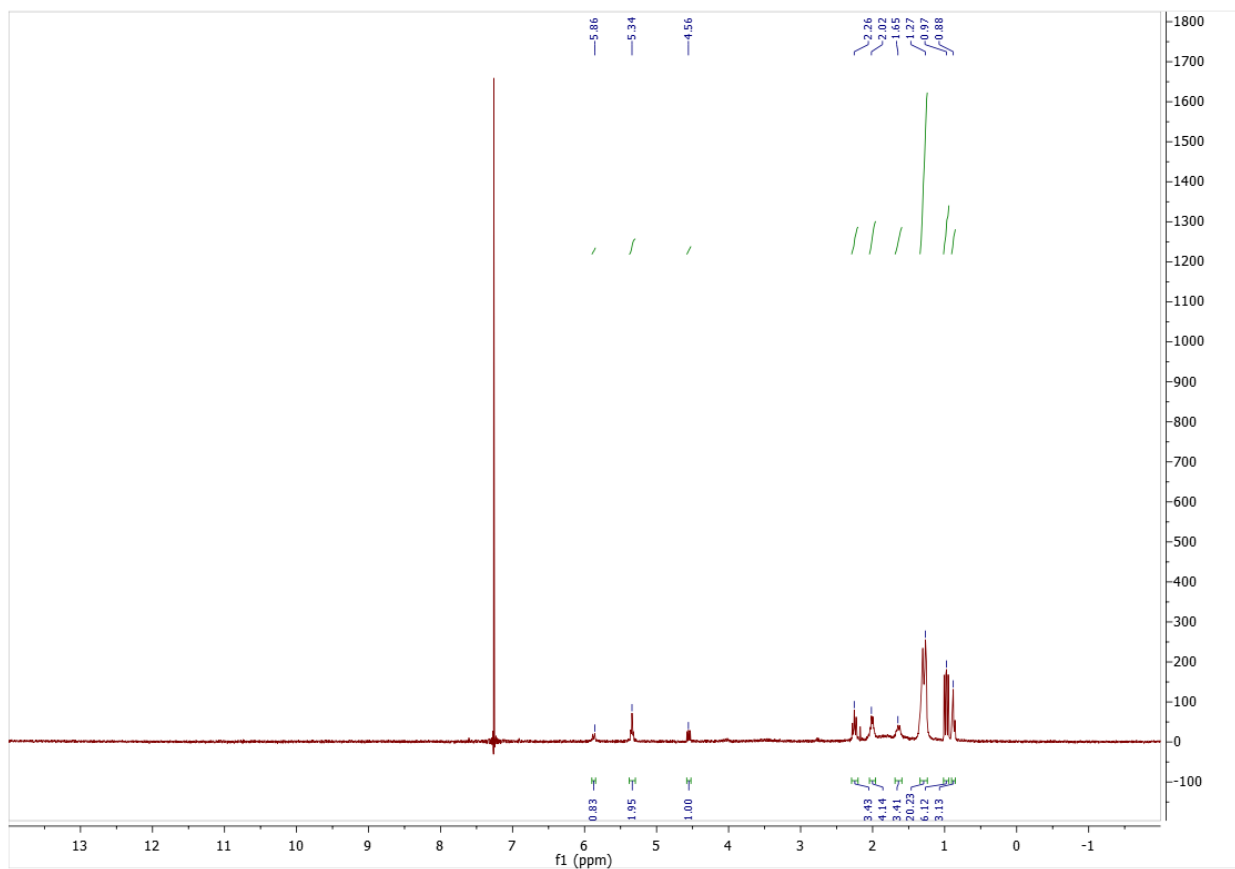
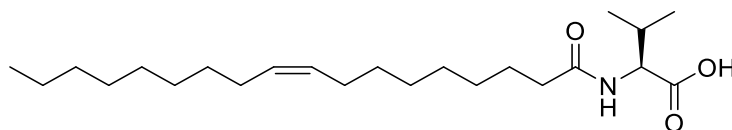


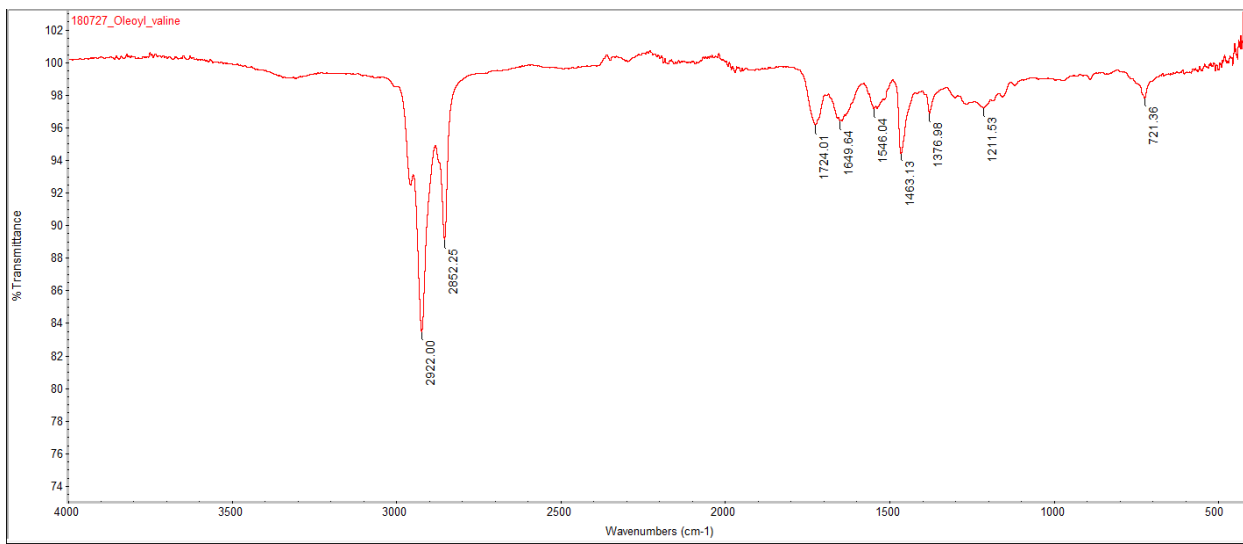
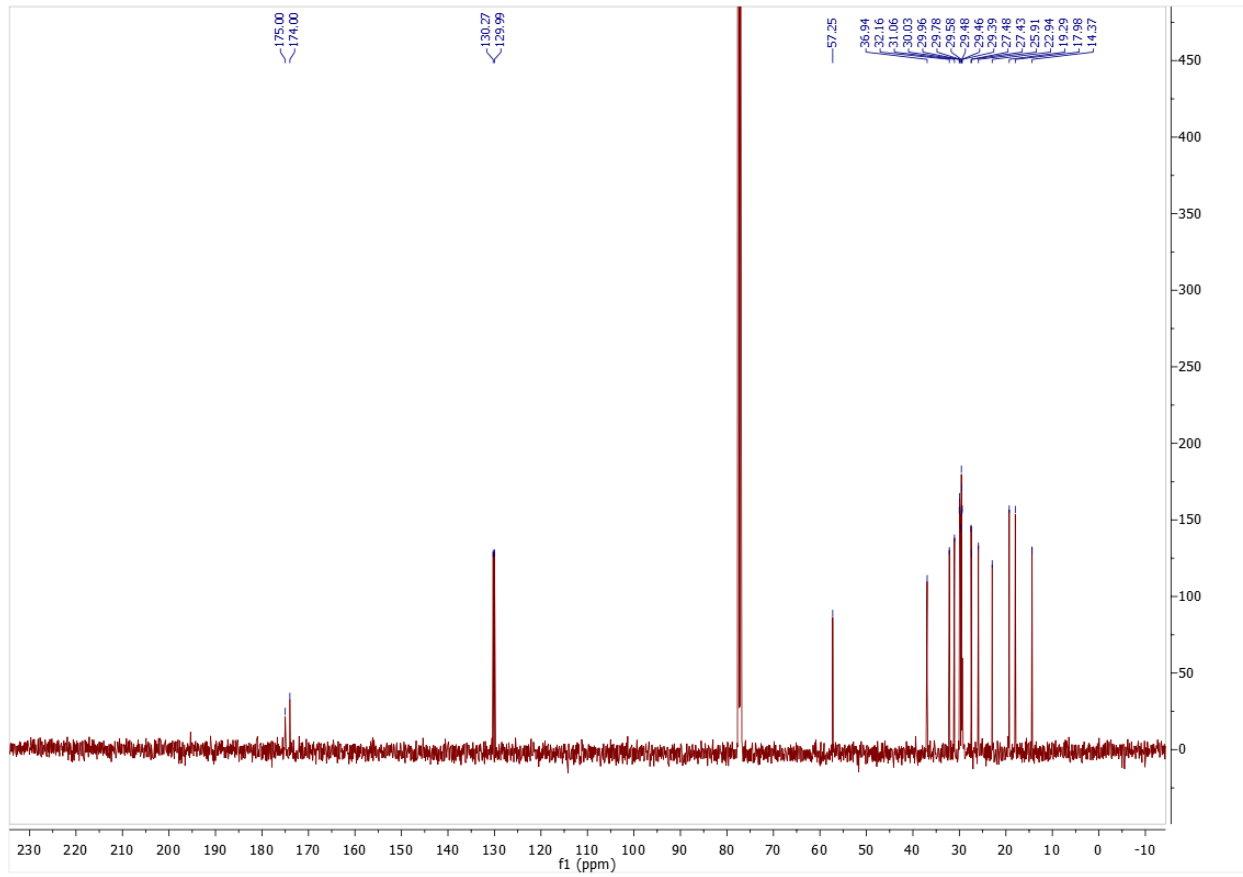


Oleoyl valine: ^1H NMR (300 MHz, CDCl_3) δ 0.88 (t, 3H), 0.97 (m, 6H), 1.27 (m, 20H), 1.65 (m, 3H), 2.02 (m, 4H), 2.26 (m, 3H), 4.56 (m, 1H), 5.34 (t, 2H), 5.86 (d, 1H).

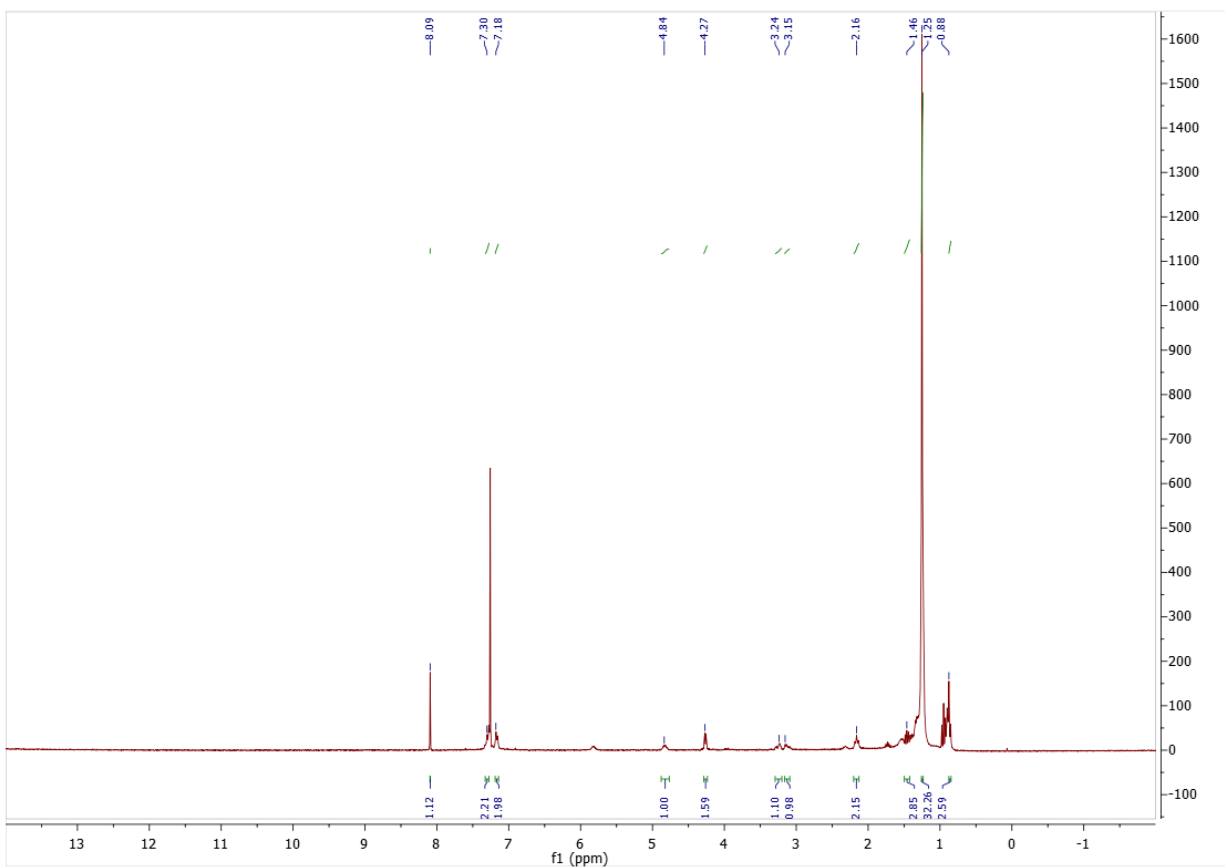
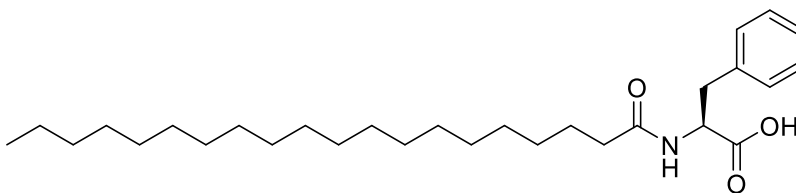
^{13}C NMR (500 MHz, CDCl_3): δ 14.37, 17.98, 19.29, 22.94, 25.91, 27.43, 27.48, 29.39, 29.46, 29.48, 29.58, 29.78, 29.96, 30.03, 31.06, 36.94, 57.25, 129.99, 130.27, 174.00, 175.00. HRMS (ESI) m/z calculated for $\text{C}_{23}\text{H}_{43}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 380.3170, found: 380.3173.

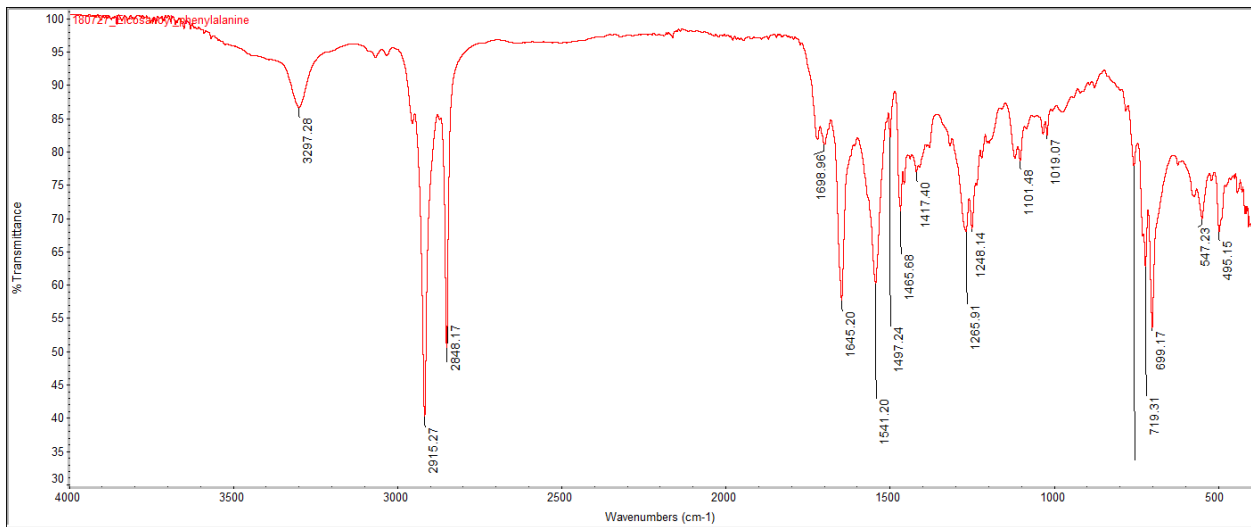
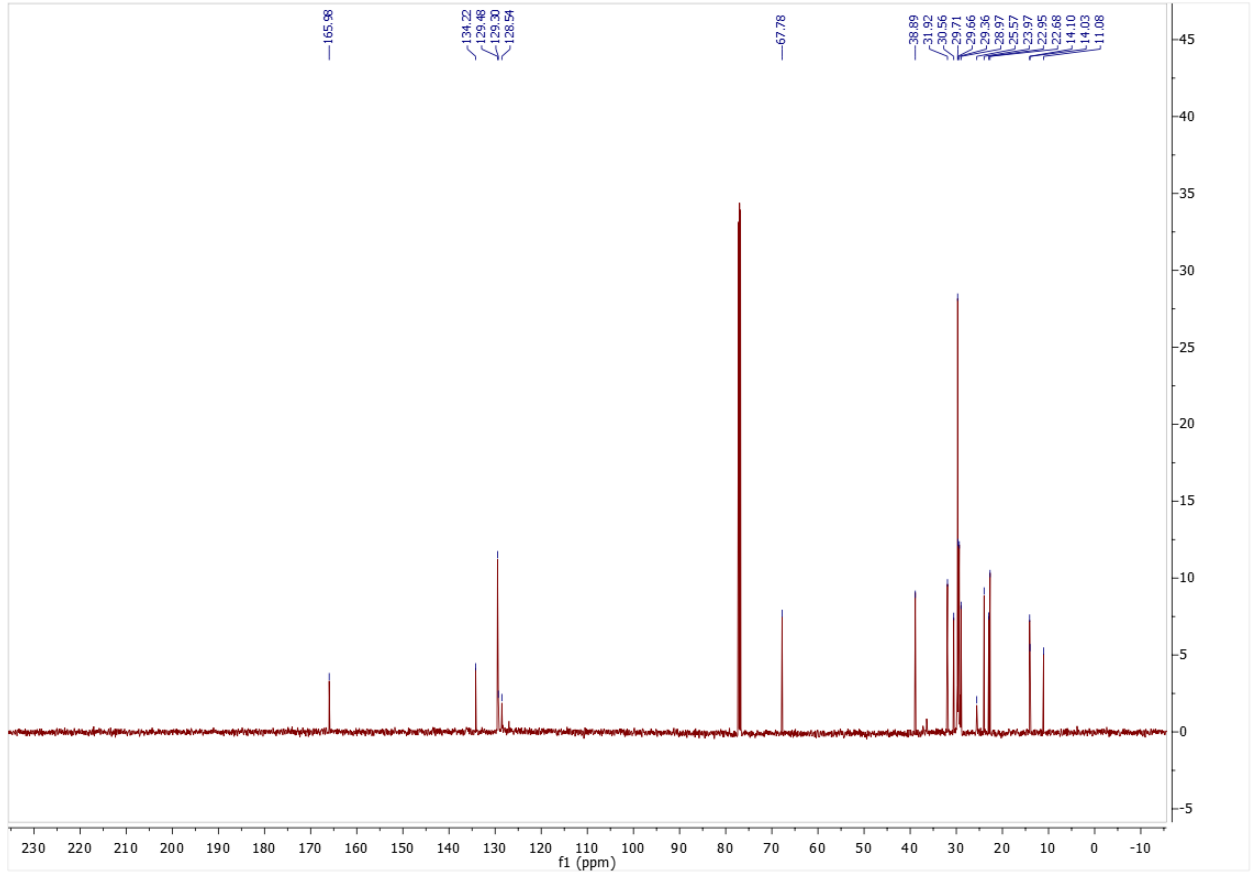
IR shown below.



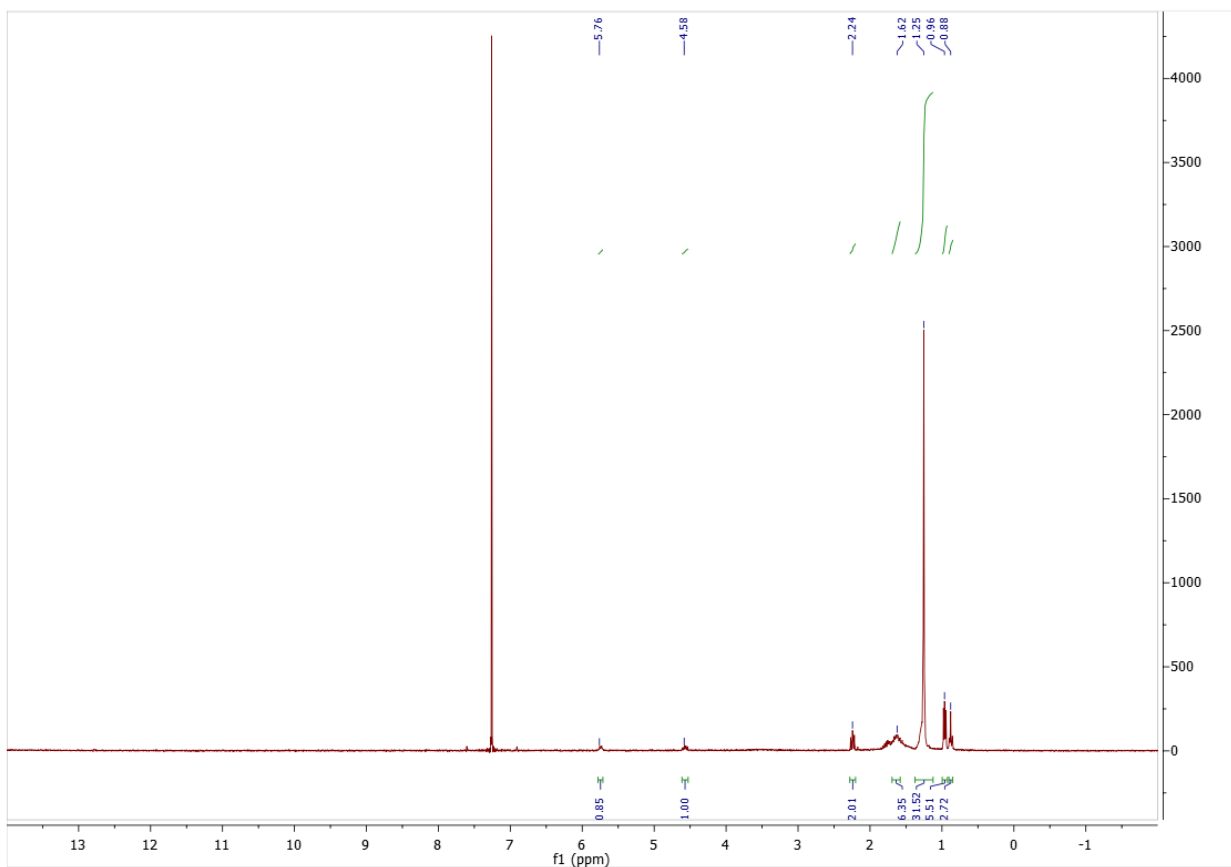
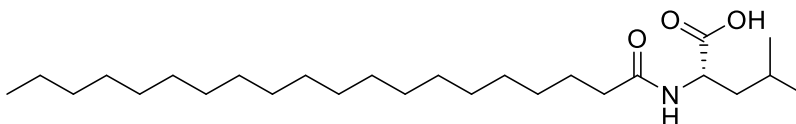


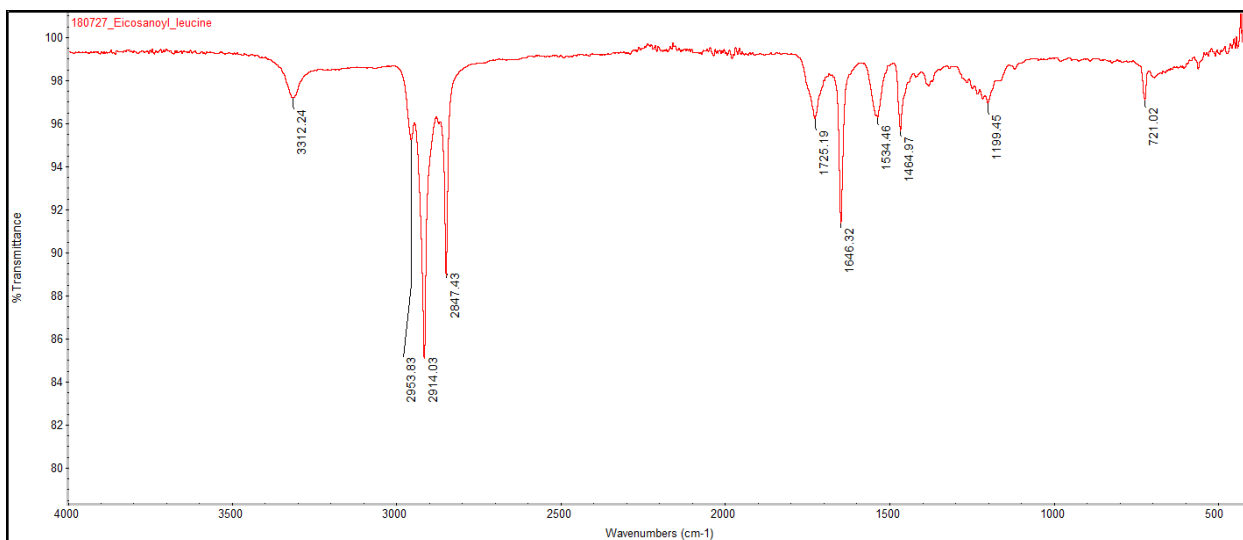
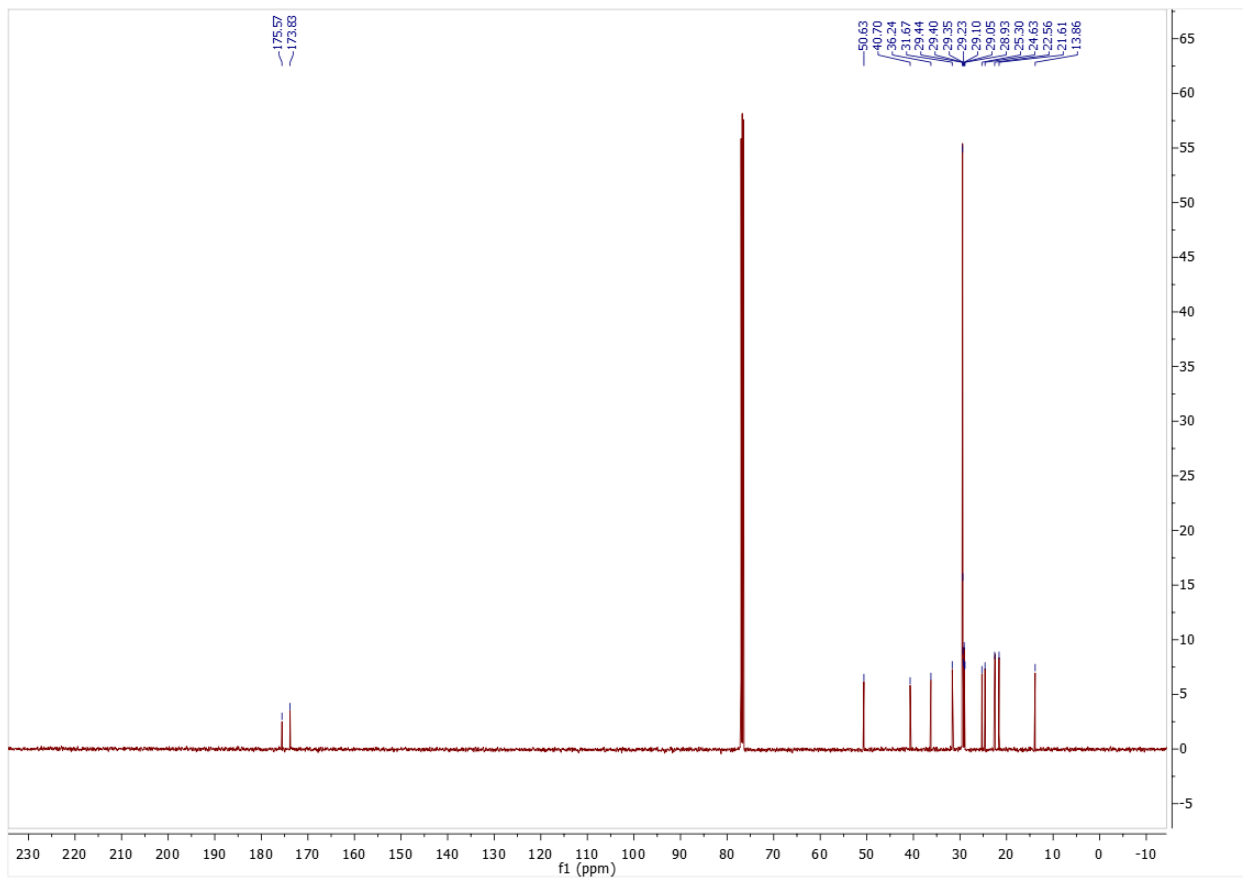
Eicosanoyl phenylalanine: ^1H NMR (300 MHz, CDCl_3) δ 0.88 (t, 3H), 1.25 (m, 32H), 1.46 (m, 3H), 2.16 (t, 2H), 3.15 (dd, 1H), 3.24 (dd, 1H), 4.27 (t, 2H), 4.84 (m, 1H), 7.18 (m, 2H), 7.30 (m, 2H). ^{13}C NMR (600 MHz, CDCl_3): δ 11.08, 14.03, 14.10, 22.68, 22.95, 23.97, 25.57, 28.97, 29.36, 29.66, 29.71, 30.56, 31.92, 38.89, 67.78, 128.54, 129.30, 129.48, 134.22. HRMS (ESI) m/z calculated for $\text{C}_{29}\text{H}_{49}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 458.3640, found: 458.3636. IR shown below.



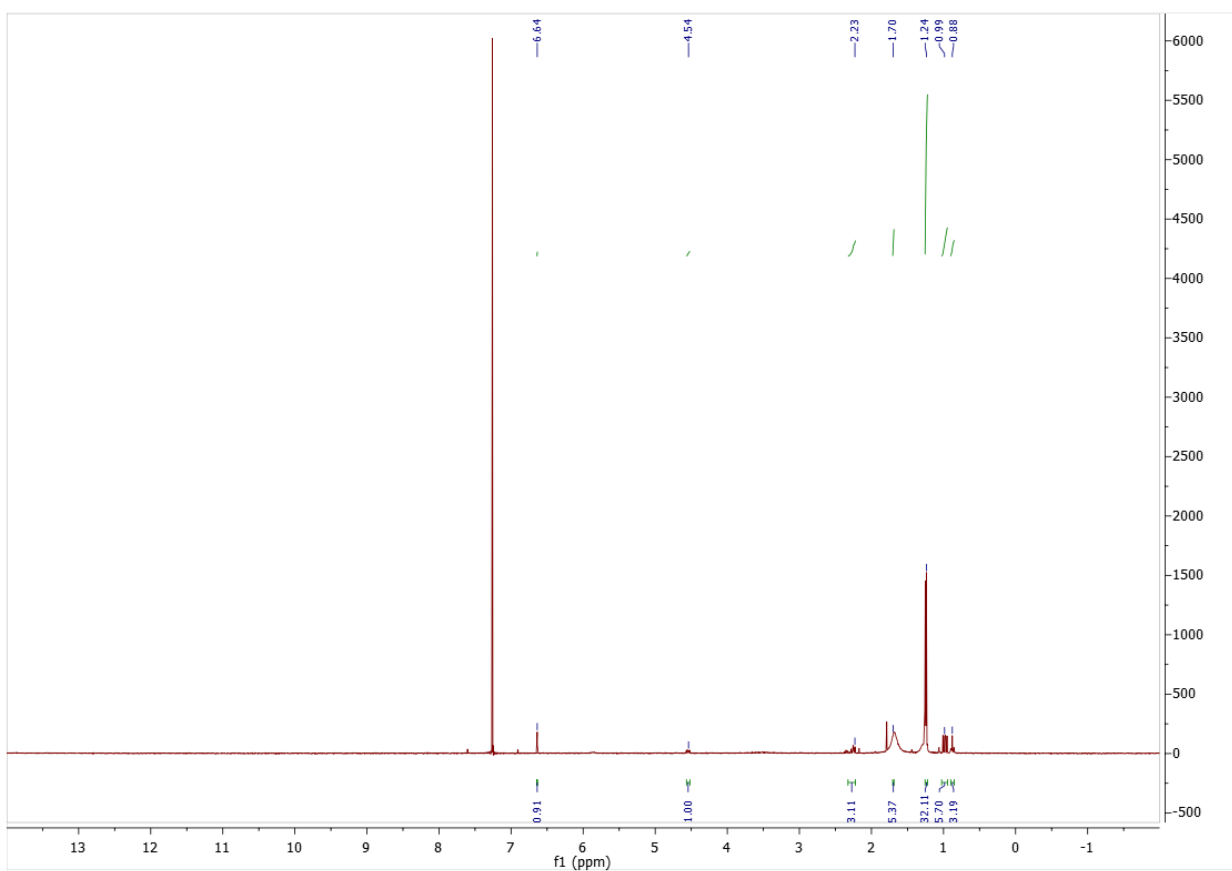
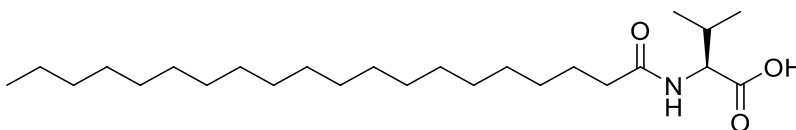


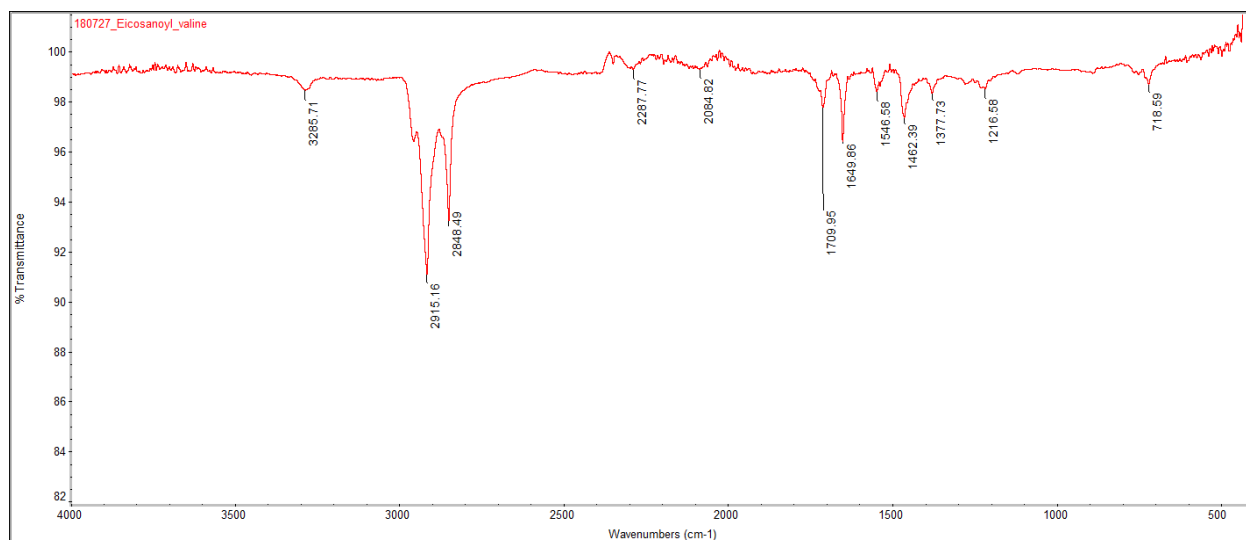
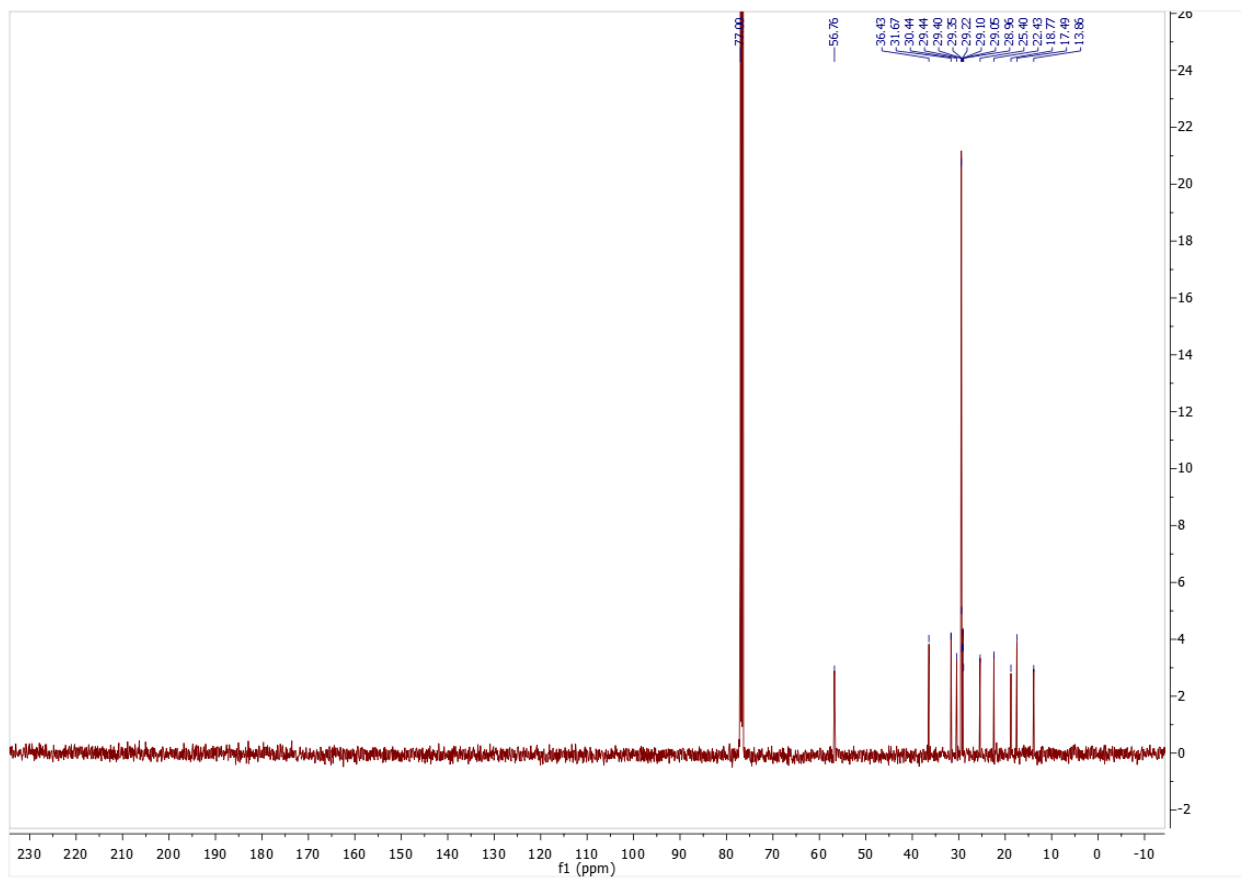
Eicosanoyl leucine: ^1H NMR (300 MHz, CDCl_3) δ 0.88 (t, 3H), 0.96 (m, 6H), 1.25 (m, 32H), 1.62 (m, 6H), 2.24 (m, 2H), 4.58 (m, 1H), 5.76 (m, 1H). ^{13}C NMR (500 MHz, CDCl_3): δ 13.86, 21.61, 22.56, 24.63, 25.30, 28.93, 29.05, 29.10, 29.23, 29.35, 29.40, 29.44, 31.67, 36.24, 40.70, 50.63, 173.83, 175.57. HRMS (ESI) m/z calculated for $\text{C}_{26}\text{H}_{51}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 424.3796, found: 424.3798. IR shown below.



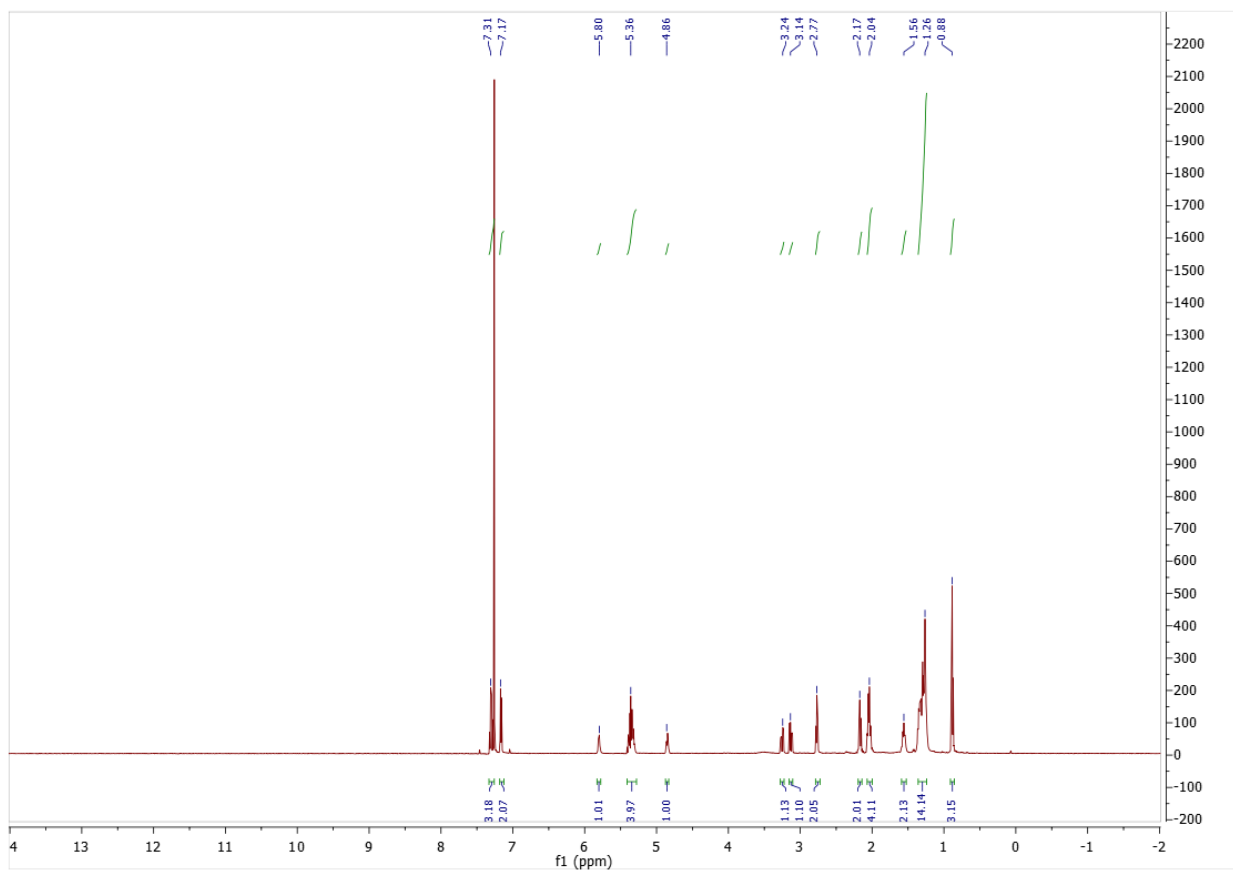
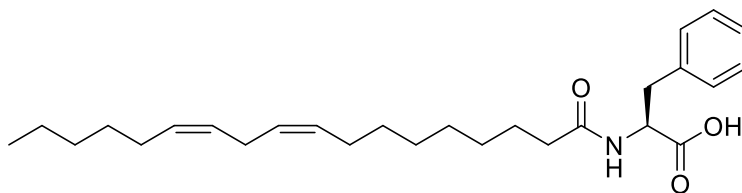


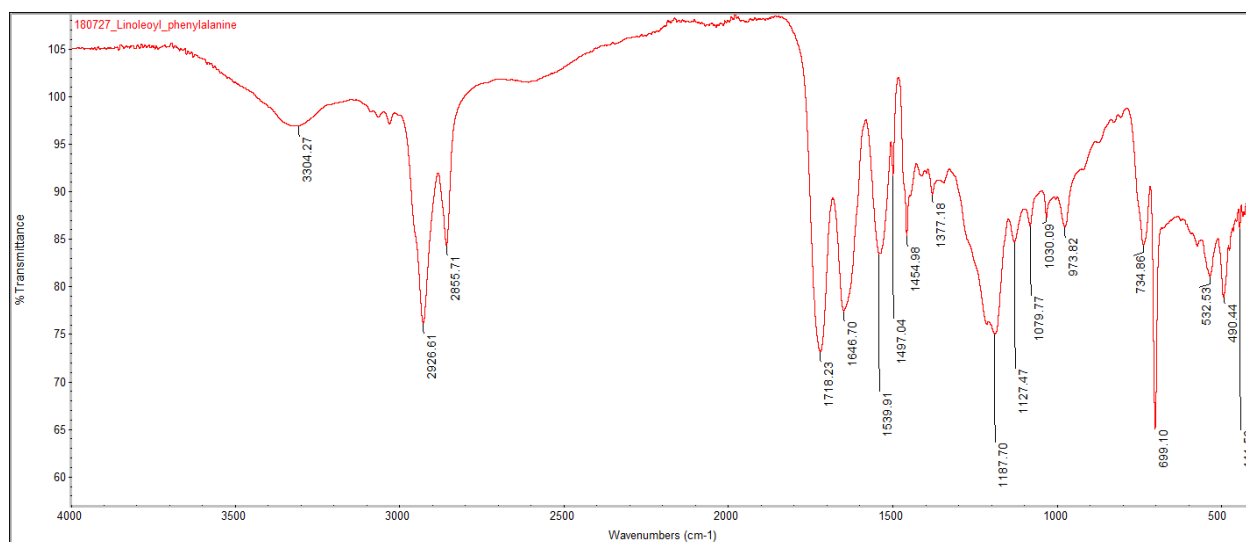
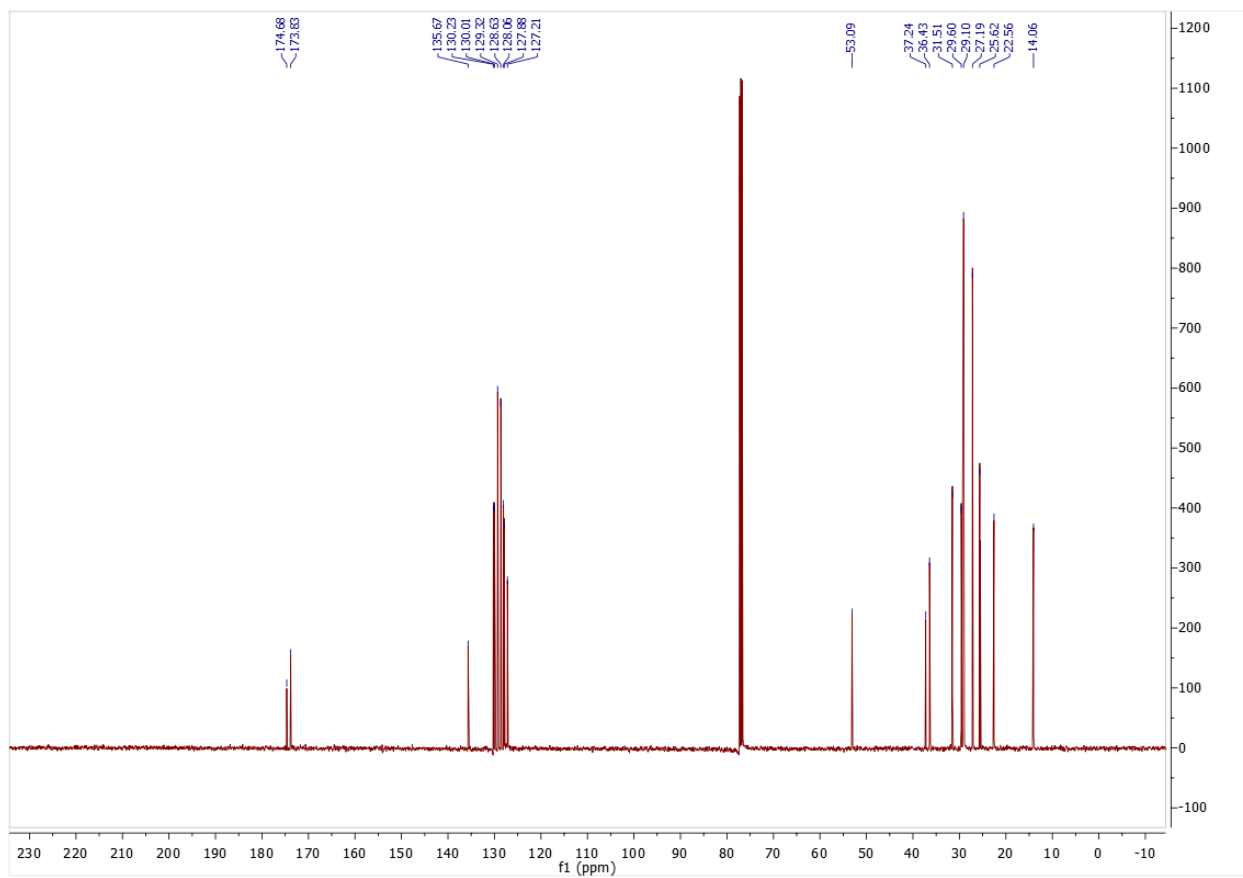
Eicosanoyl valine: ^1H NMR (300 MHz, CDCl_3) δ 0.88 (t, 3H), 0.99 (m, 6H), 1.24 (m, 32H), 1.70 (m, 5H), 2.23 (m, 3H), 4.54 (m, 1H), 6.64 (d, 1H). ^{13}C NMR (500 MHz, CDCl_3): δ 13.86, 17.49, 18.77, 22.48, 25.40, 28.96, 29.05, 29.10, 29.22, 29.35, 29.40, 29.44, 30.44, 31.67, 36.43, 56.76. HRMS (ESI) m/z calculated for $\text{C}_{25}\text{H}_{49}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 410.3640, found: 410.3641. IR shown below.



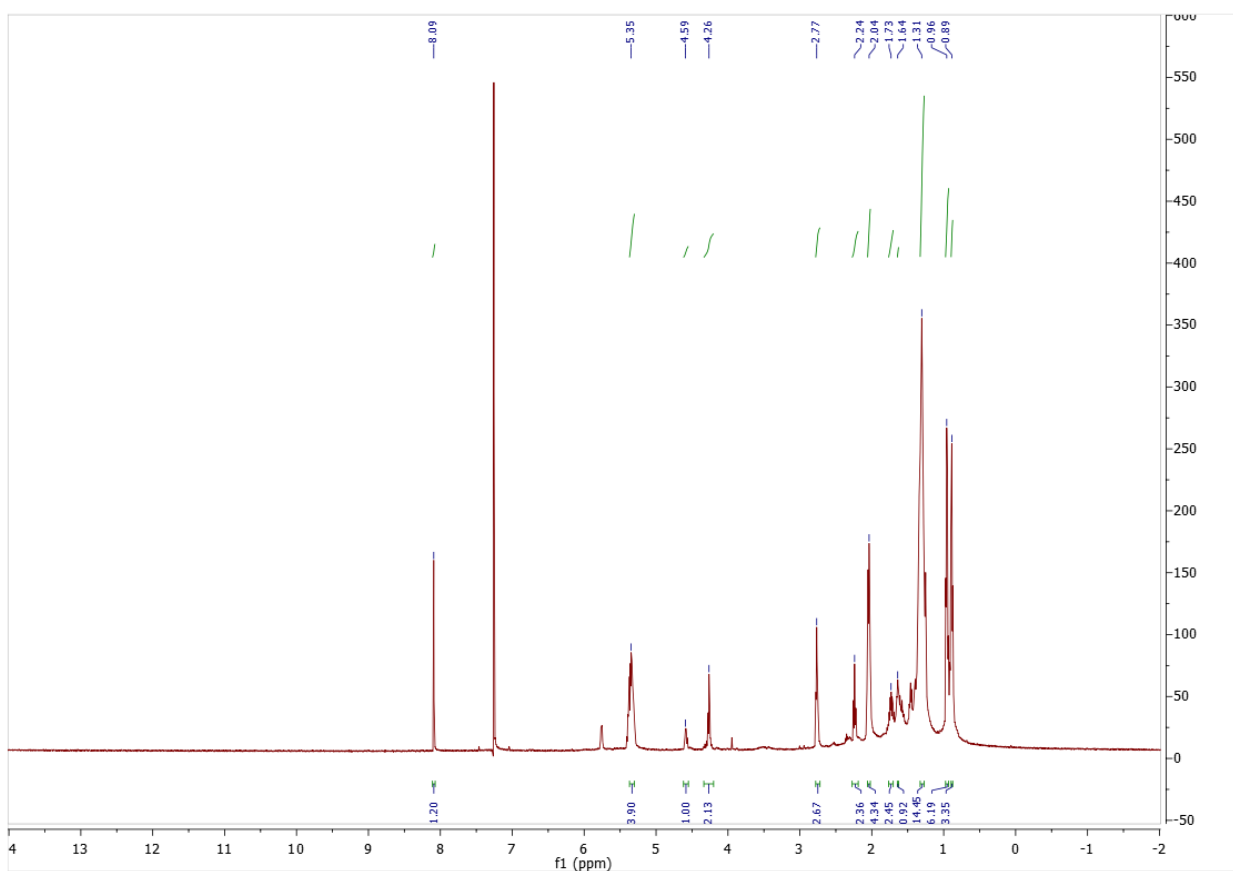
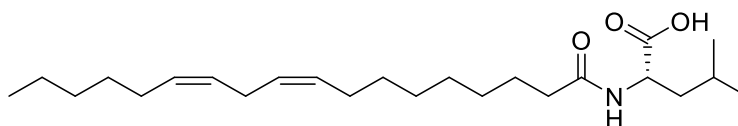


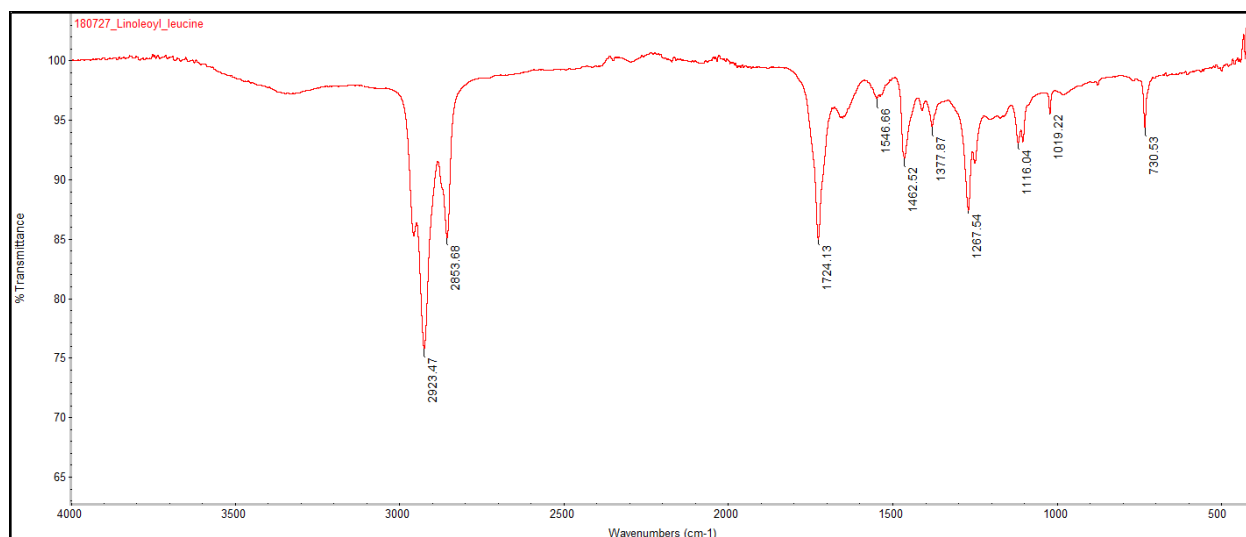
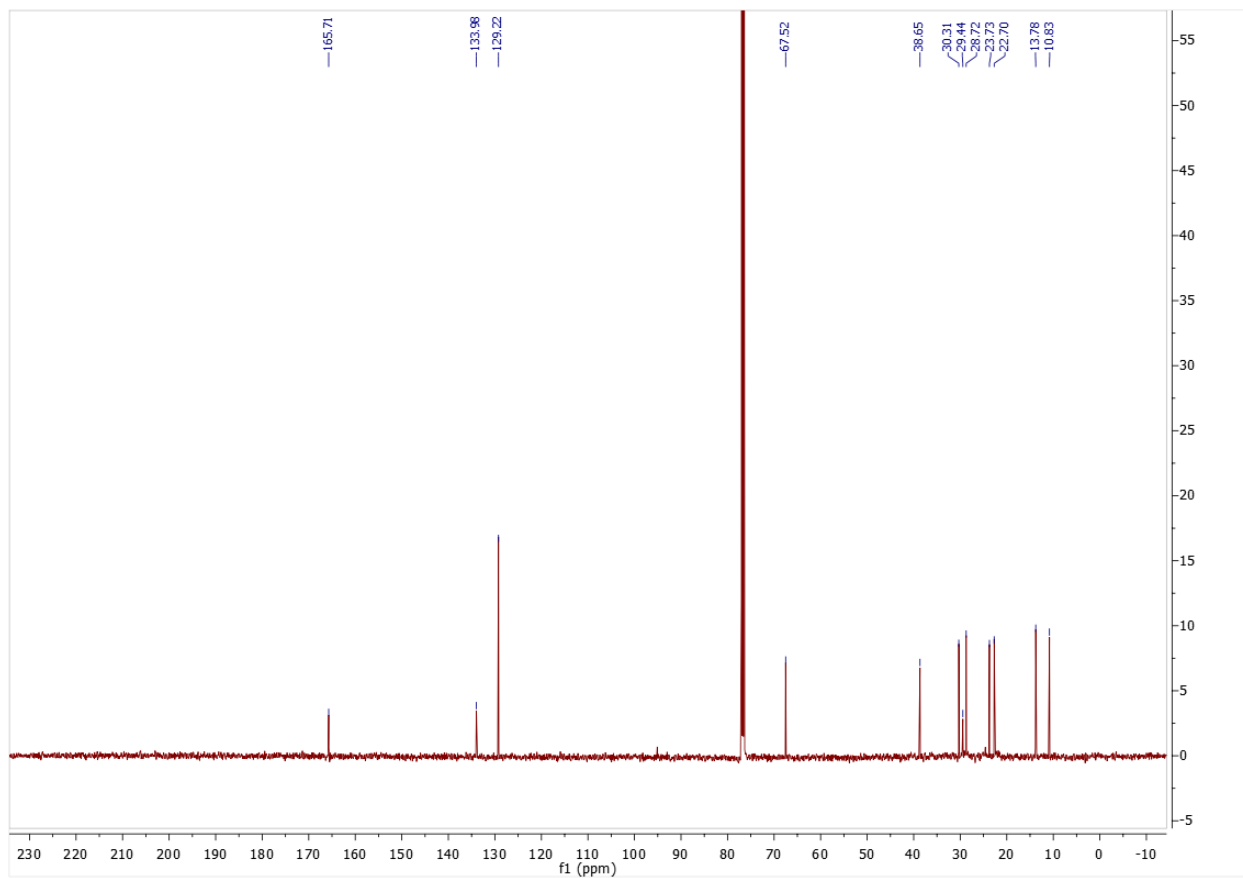
Linoleoyl phenylalanine: $^1\text{H NMR}$ (300 MHz, CDCl_3) δ 0.88 (t, 3H), 1.26 (m, 14H), 1.56 (m, 2H), 2.04 (m, 4H), 2.17 (t, 2H), 2.77 (t, 2H), 3.14 (dd, 1H), 3.24 (dd, 1H), 4.86 (m, 1H), 5.36 (m, 4H), 5.80 (d, 1H), 7.17 (m, 2H), 7.31 (m, 3H). $^{13}\text{CNMR}$ (500 MHz, CDCl_3): δ 14.06, 22.56, 25.62, 27.19, 29.10, 29.60, 31.51, 36.43, 37.24, 53.09, 127.21, 127.88, 128.06, 128.63, 129.32, 130.01, 130.23, 135.67, 173.83, 174.68. HRMS (ESI) m/z calculated for $\text{C}_{27}\text{H}_{41}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 426.3014, found: 426.3013. IR shown below.



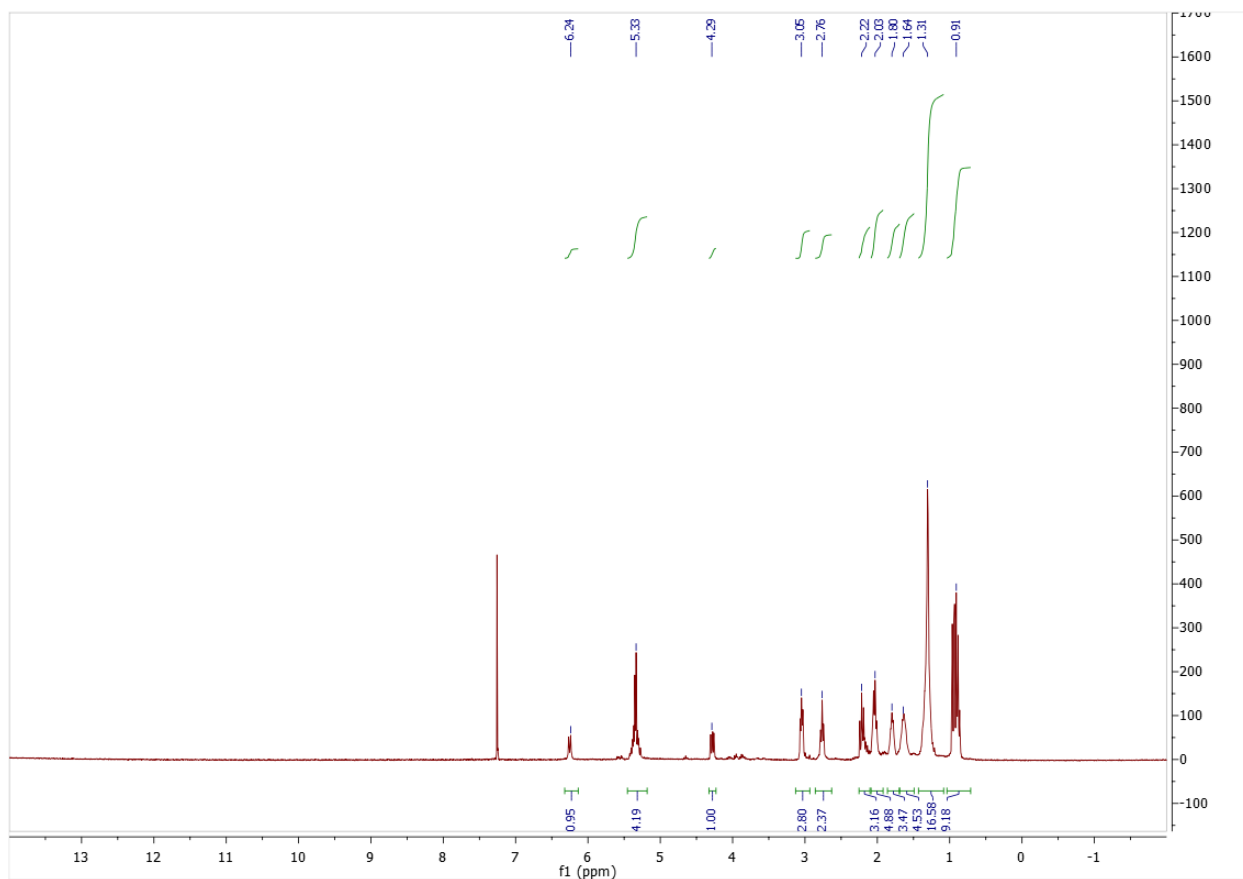
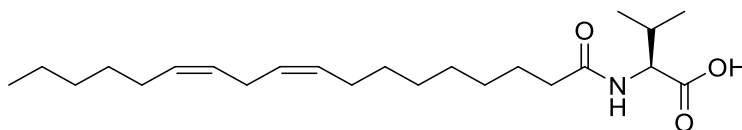


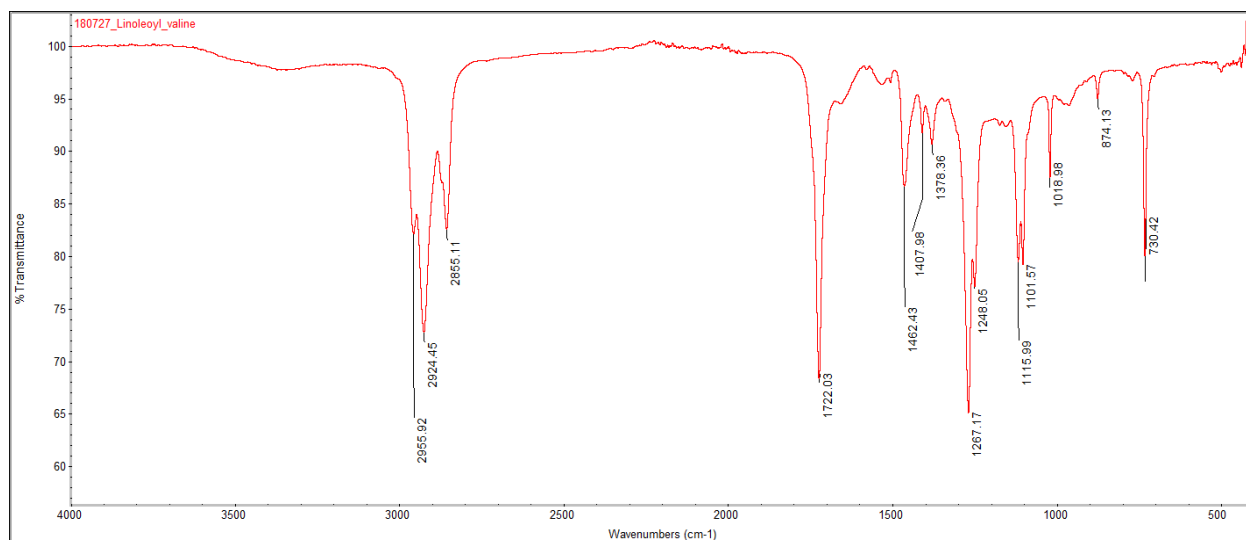
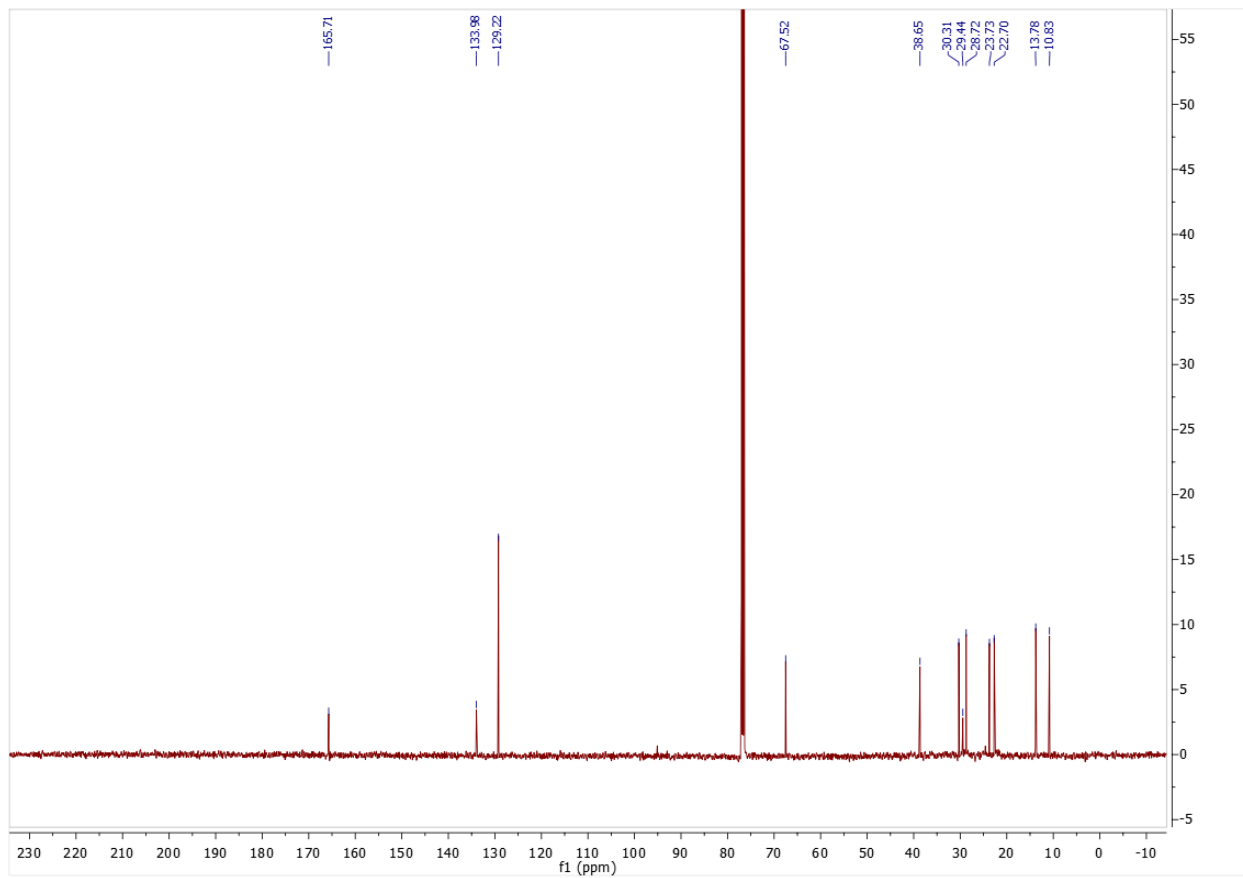
Linoleoyl leucine: ^1H NMR (300 MHz, CDCl_3) δ 0.89 (t, 3H), 0.96 (m, 6H), 1.31 (m, 14H), 1.64 (m, 1H), 1.73 (m, 2H), 2.04 (m, 4H), 2.24 (m, 2H), 2.77 (t, 3H), 4.26 (t, 2H), 4.59 (m, 1H), 5.35 (m, 4H), 8.09 (s, 1H). ^{13}C NMR (500 MHz, CDCl_3): δ 10.83, 13.78, 22.70, 23.73, 28.72, 29.44, 30.31, 38.65, 67.52, 129.22, 133.98, 165.71. HRMS (ESI) m/z calculated for $\text{C}_{24}\text{H}_{42}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 392.3170, found: 392.3168. IR shown below.



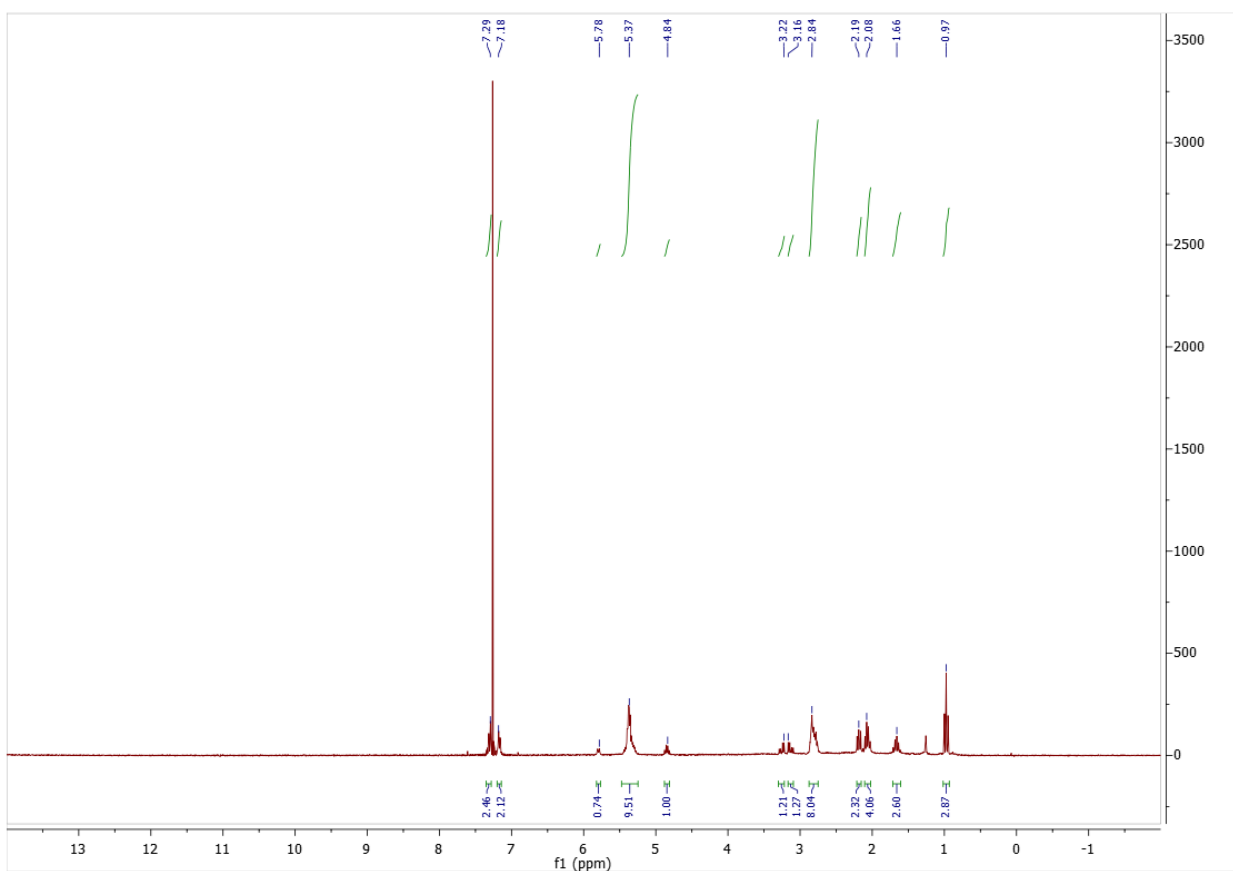
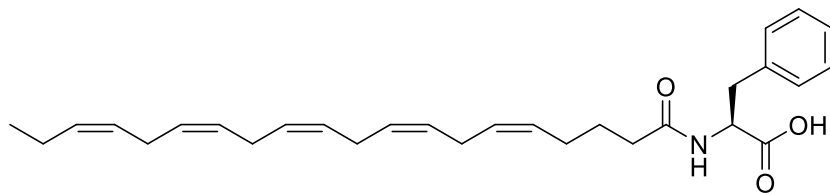


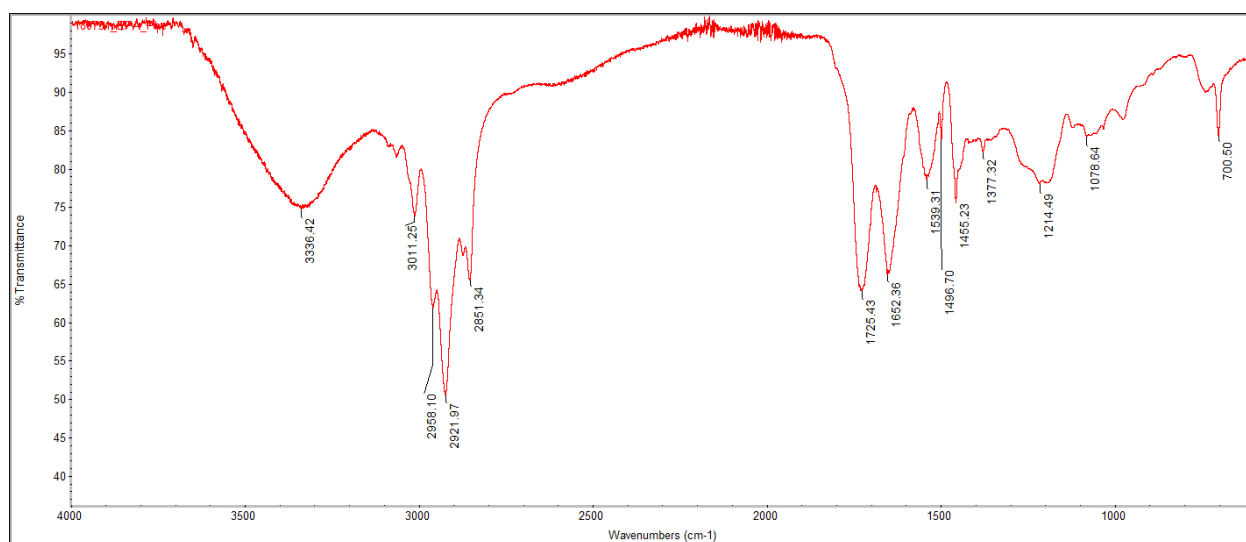
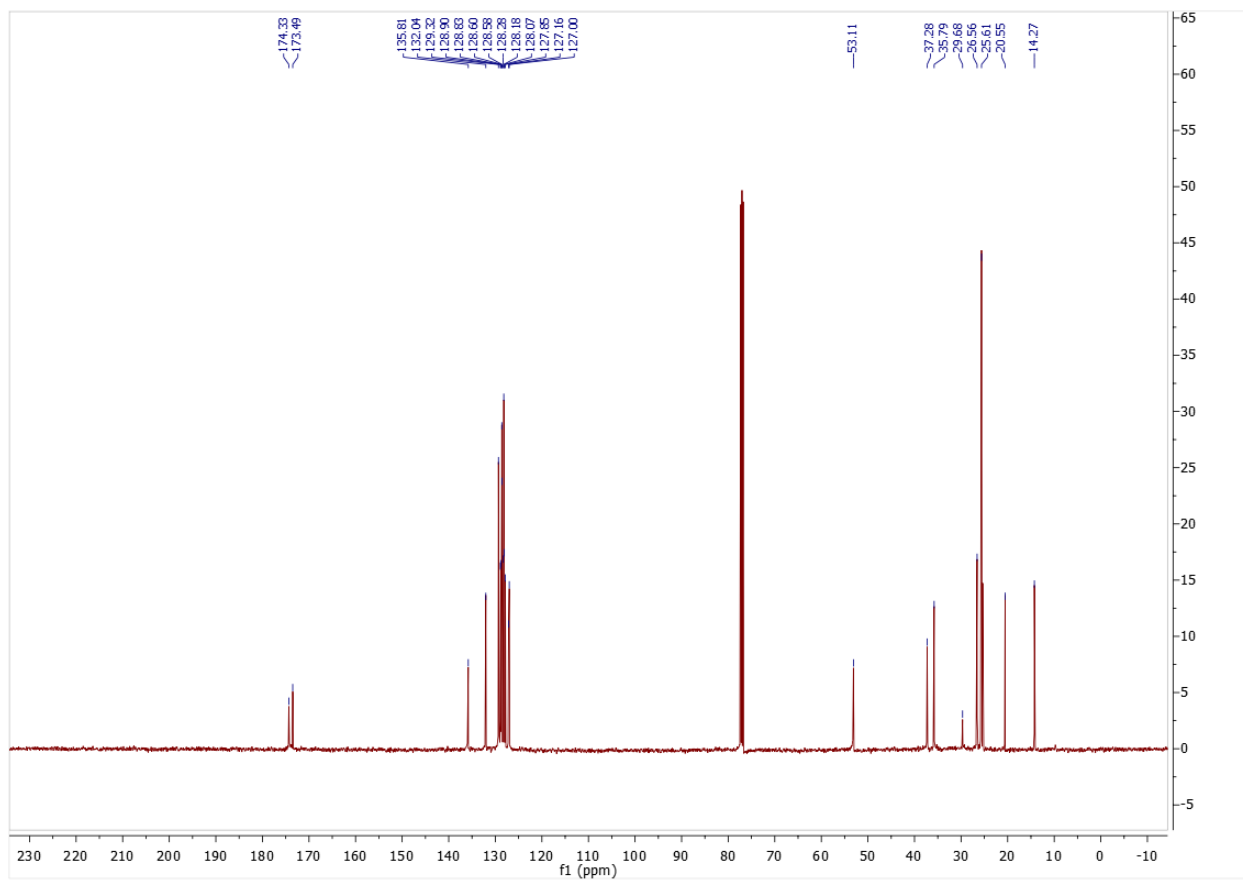
Linoleoyl valine: $^1\text{H NMR}$ (300 MHz, CDCl_3) δ 0.91 (m, 9H), 1.31 (m, 16H), 1.64 (m, 4H), 1.80 (m, 4H), 2.03 (m, 5H), 2.22 (m, 3H), 2.76 (t, 2H), 3.05 (t, 2H), 4.29 (q, 1H), 5.33 (m, 4H), 6.24 (d, 1H). $^{13}\text{CNMR}$ (500 MHz, CDCl_3): δ 13.81, 17.78, 19.10, 22.21, 22.31, 22.36, 25.36, 25.56, 26.94, 28.92, 29.03, 29.05, 29.08, 29.38, 30.78, 31.25, 36.54, 44.18, 58.51, 127.63, 127.78, 129.77, 129.96, 173.24, 176.25. IR shown below.



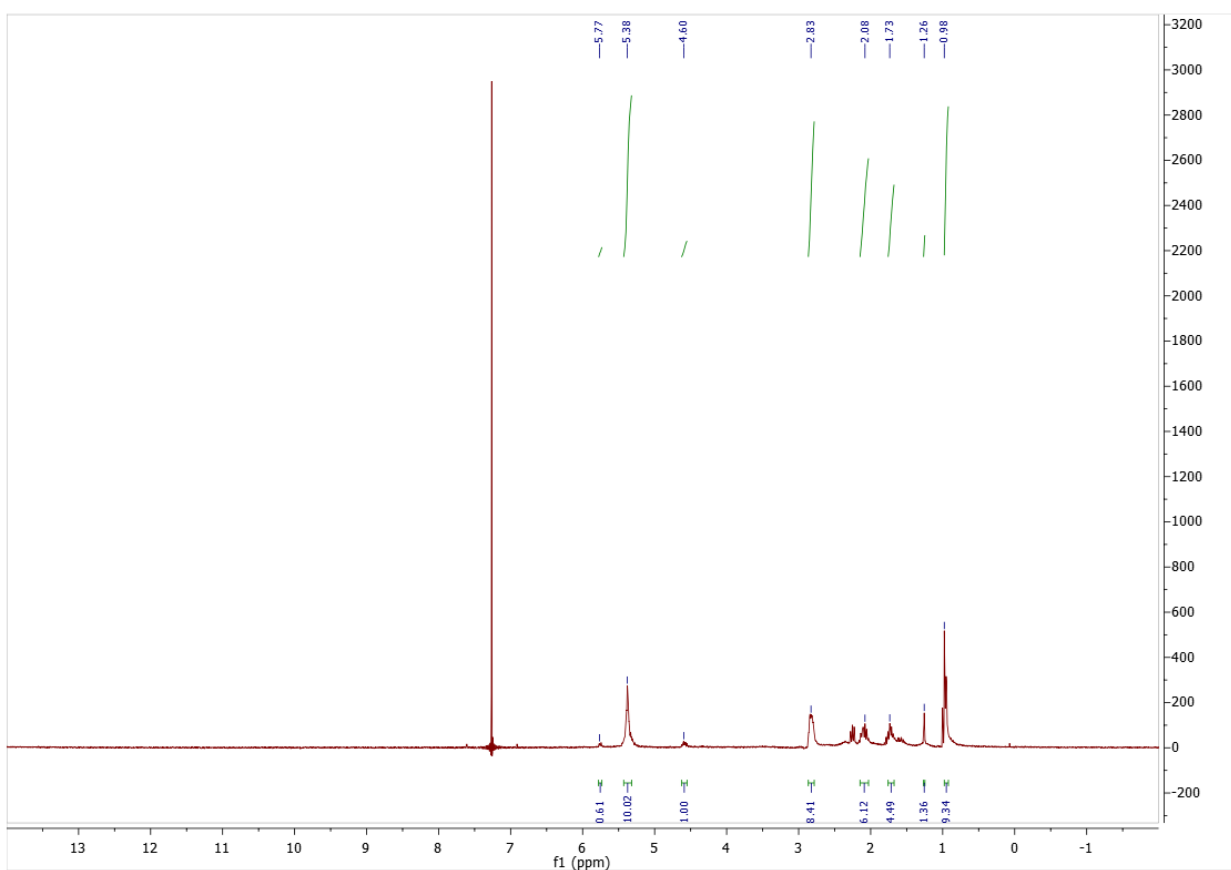
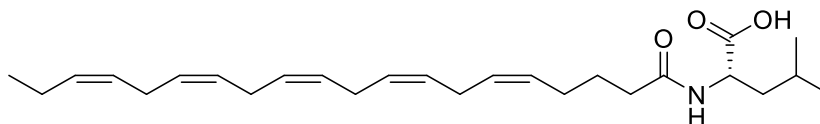


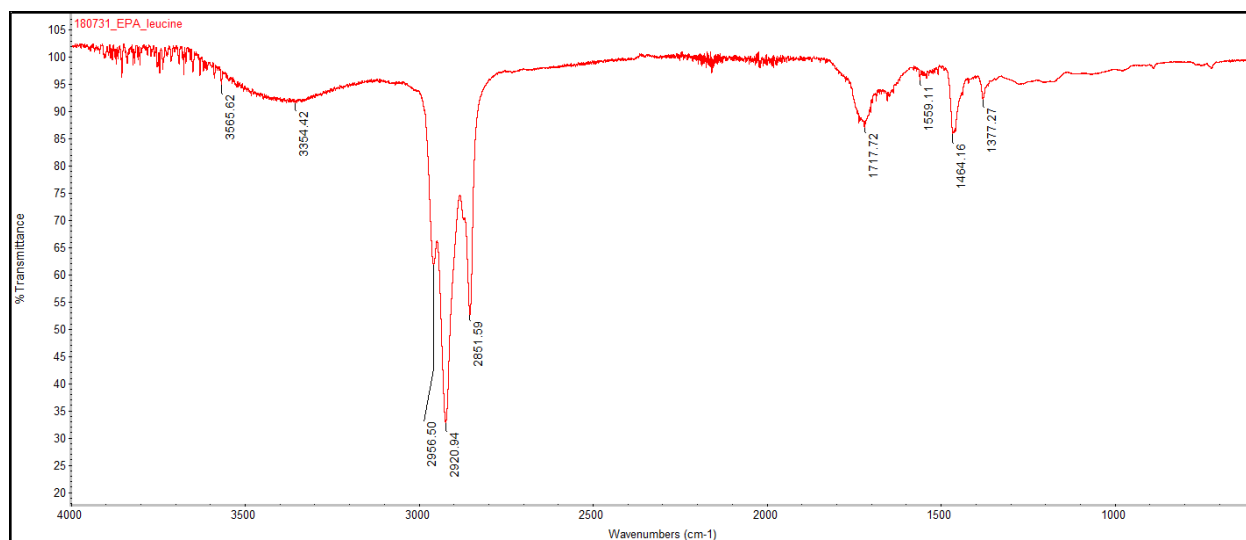
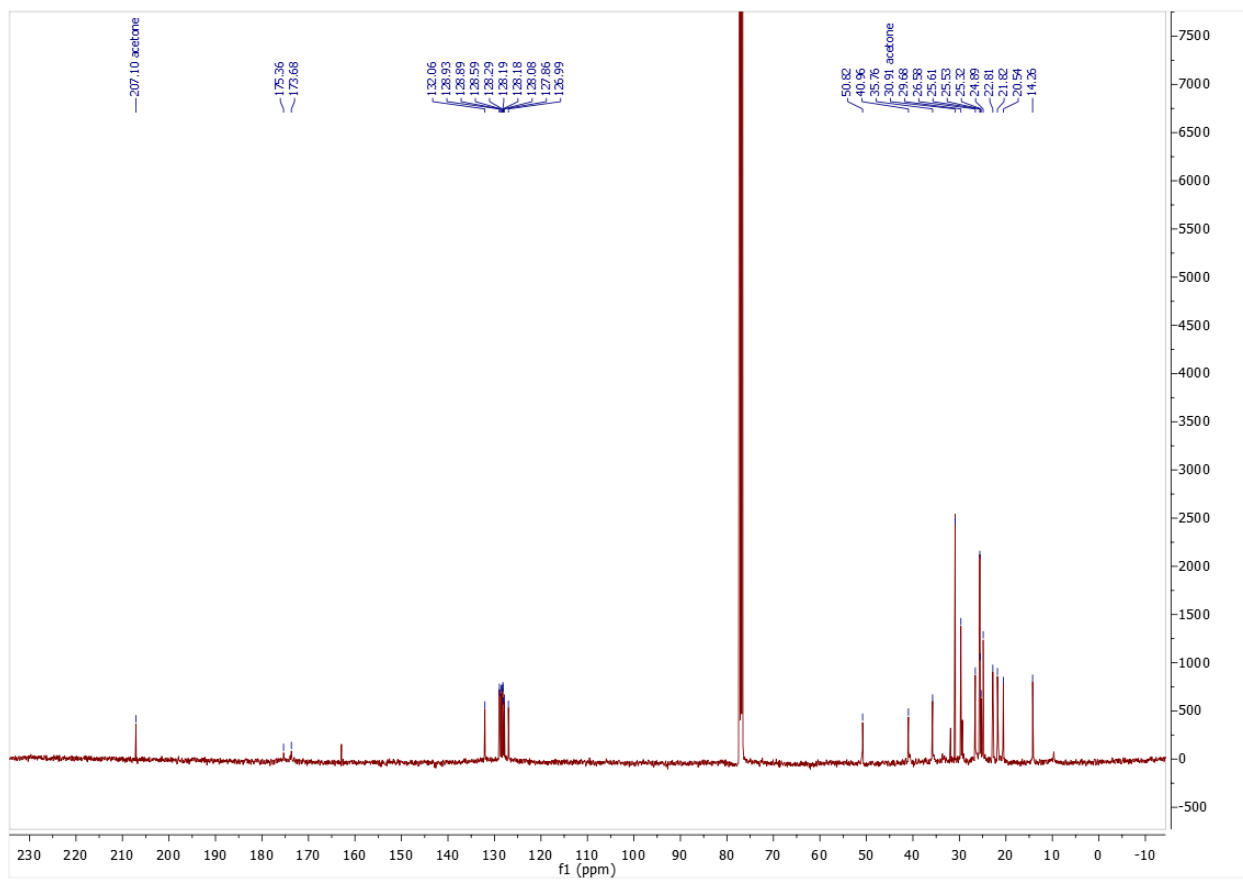
Eicosapentaenoyl phenylalanine: ^1H NMR (300 MHz, CDCl_3) δ 0.97 (t, 3H), 1.66 (m, 3H), 2.08 (m, 4H), 2.19 (m, 2H), 2.84 (m, 8H), 3.16 (m, 1H), 3.22 (m, 1H), 4.84 (m, 1H), 5.37 (m, 10H), 5.78 (d, 1H), 7.18 (m, 2H), 7.29 (m, 3H). ^{13}C NMR (500 MHz, CDCl_3): δ 14.27, 20.55, 25.61, 26.56, 29.68, 35.79, 37.28, 53.11, 127.00, 127.16, 127.85, 128.07, 128.18, 128.28, 128.58, 128.60, 128.83, 128.90, 129.32, 132.04, 135.81, 173.49, 174.33. HRMS (ESI) m/z calculated for $\text{C}_{29}\text{H}_{39}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 448.2857, found: 448.2854



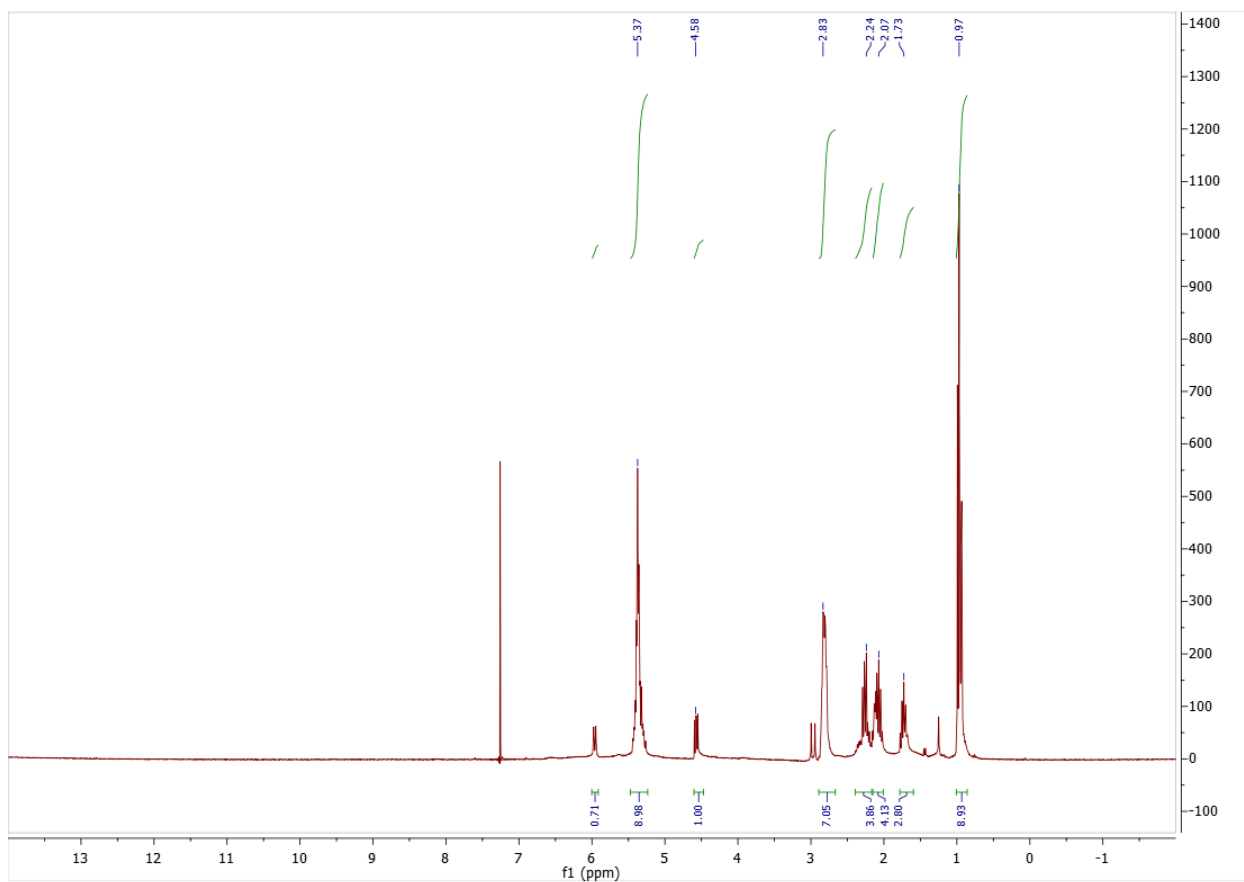
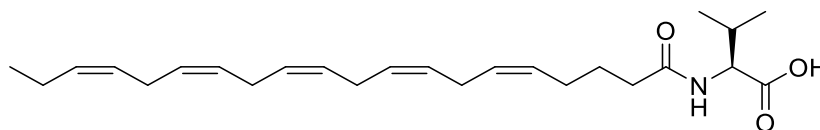


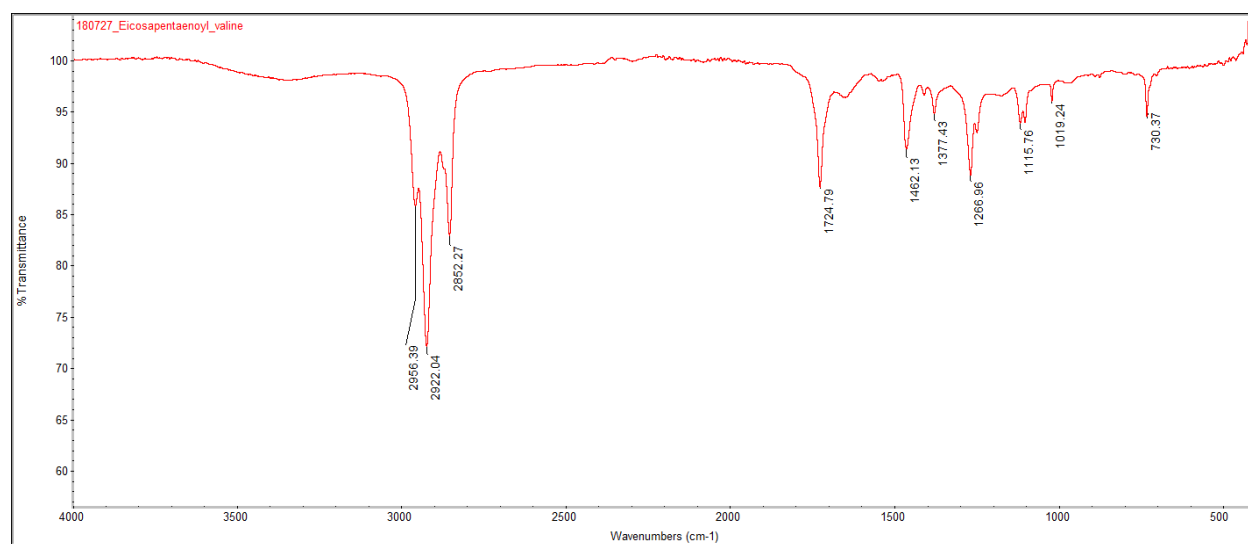
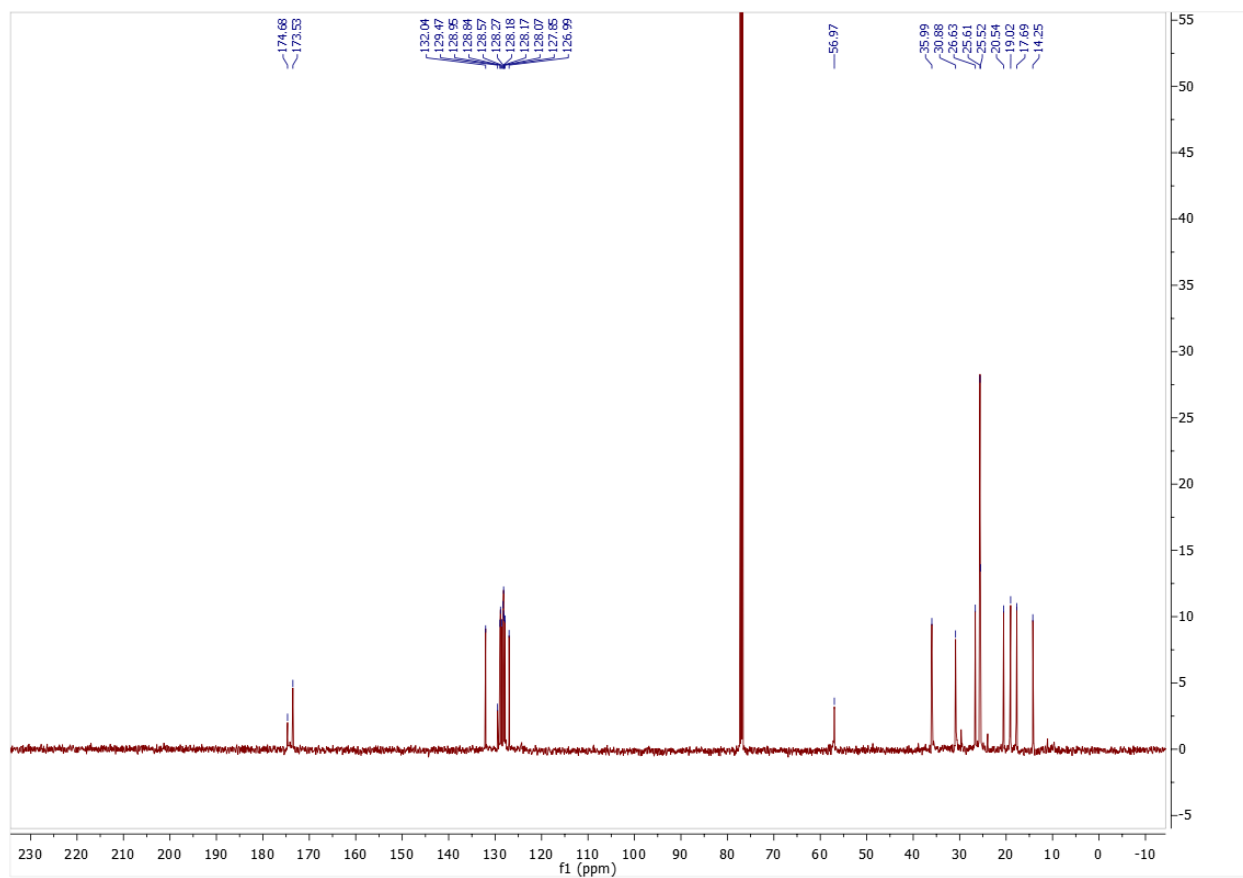
Eicosapentaenoyl leucine: ^1H NMR (300 MHz, CDCl_3) δ 0.98 (m, 9H), 1.26 (m, 1H), 1.73 (m, 5H), 2.08 (m, 6H), 2.83 (m, 8H), 4.60 (m, 1H), 5.38 (m, 10H), 5.77 (d, 1H). ^{13}C NMR (500 MHz, CDCl_3): δ 14.26, 20.54, 21.82, 22.81, 24.89, 25.32, 25.53, 25.61, 26.58, 29.68, 35.76, 40.96, 50.82, 126.99, 127.86, 128.08, 128.18, 128.19, 128.29, 128.59, 128.89, 128.93, 132.06, 173.68, 175.36. HRMS (ESI) m/z calculated for $\text{C}_{26}\text{H}_{41}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 414.3014, found: 414.3012. IR shown below.



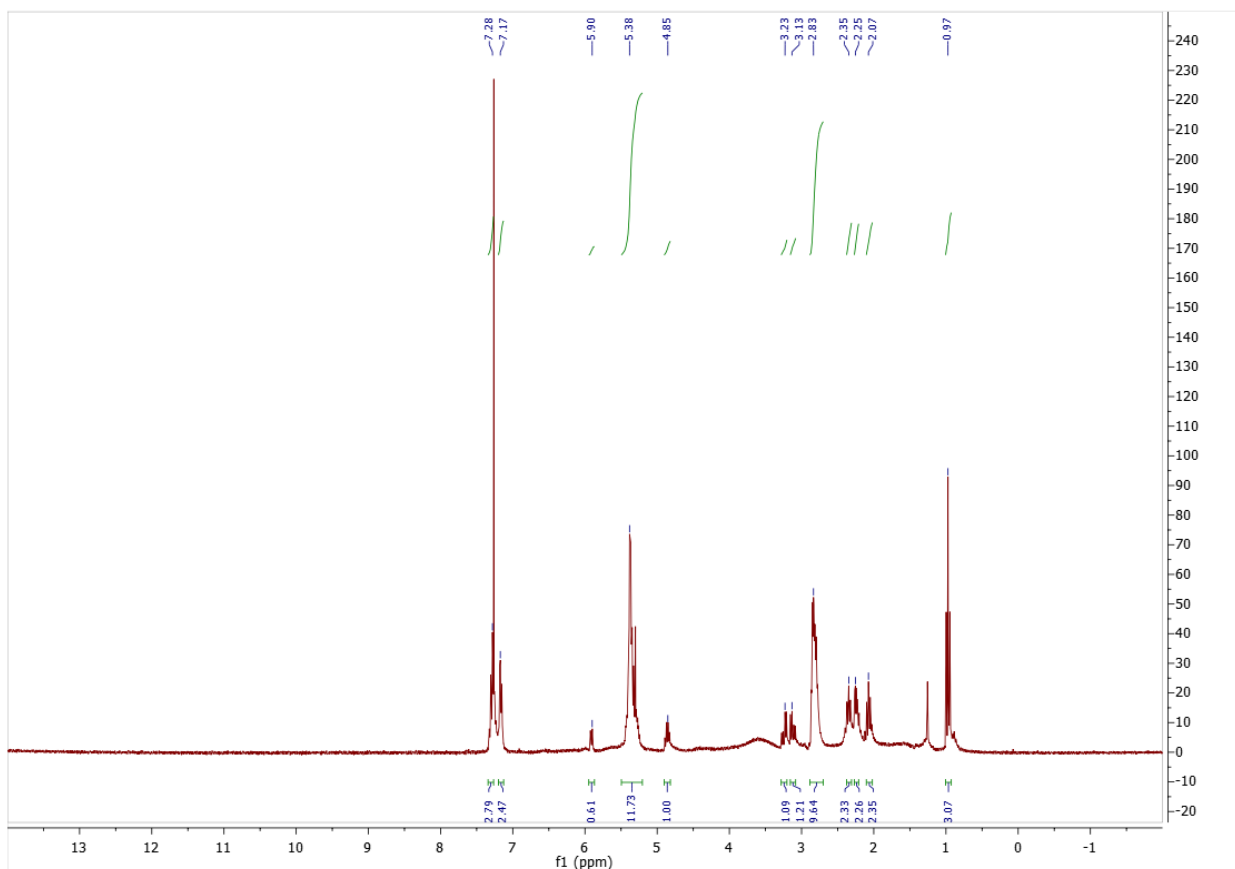
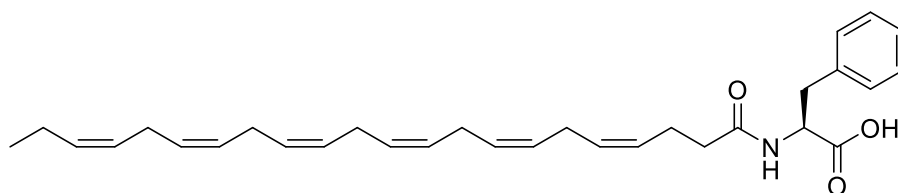


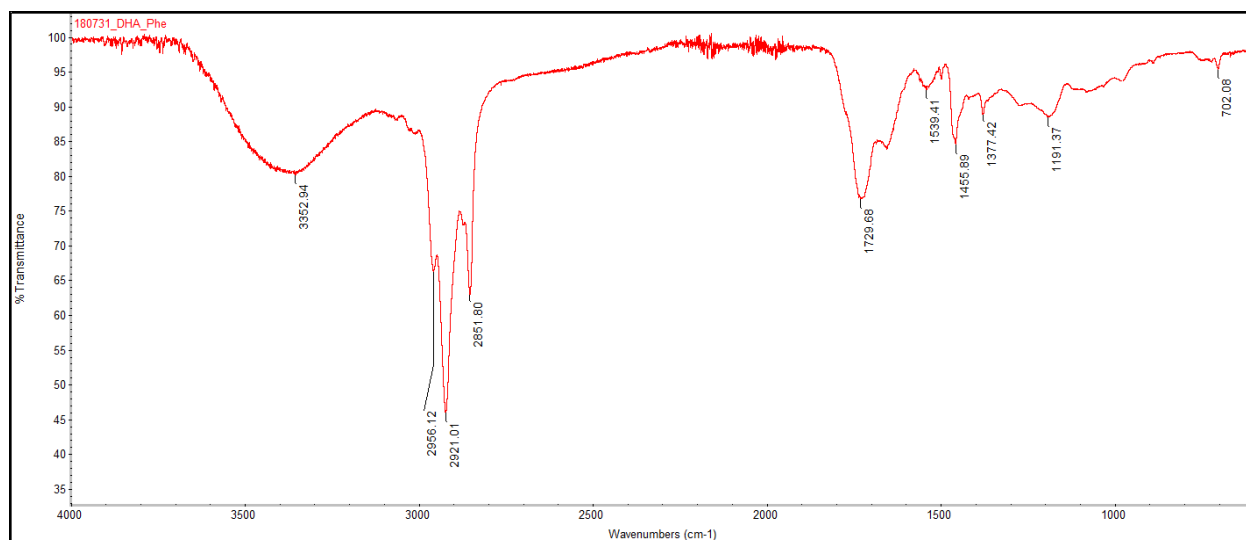
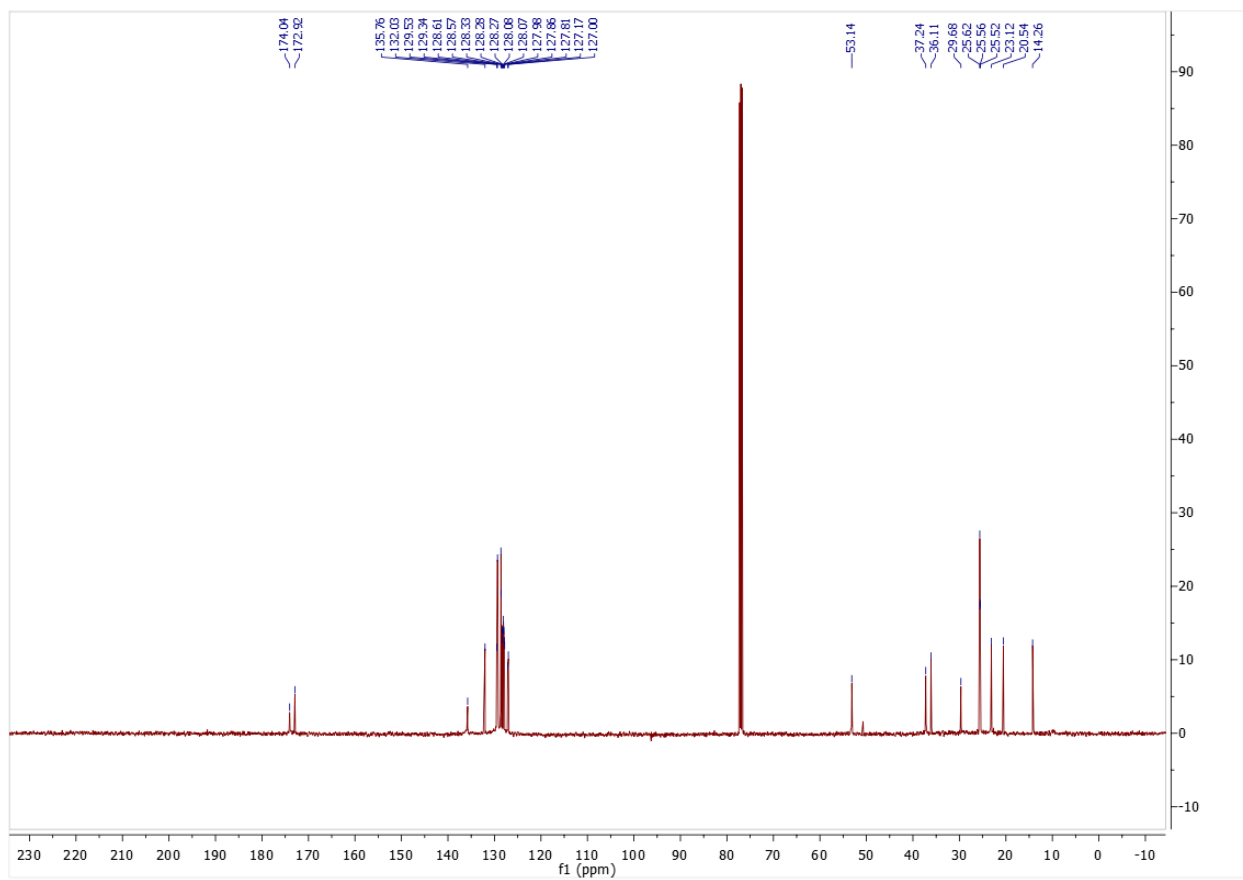
Eicosapentaenoyl valine: ^1H NMR (300 MHz, CDCl_3) δ 0.97 (m, 9H), 1.73 (m, 3H), 2.07 (m, 4H), 2.24 (m, 4H), 2.83 (m, 7H), 4.58 (m, 1H), 5.37 (m, 9H), 5.97 (d, 1H). ^{13}C NMR (500 MHz, CDCl_3): δ 14.26, 17.69, 19.02, 20.54, 25.52, 25.61, 26.63, 30.88, 35.99, 56.97, 126.99, 127.85, 128.07, 128.17, 128.18, 128.27, 128.57, 128.84, 128.95, 129.47, 132.04, 173.53, 174.68. HRMS (ESI) m/z calculated for $\text{C}_{26}\text{H}_{39}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 400.2857, found: 400.2852. IR shown below.



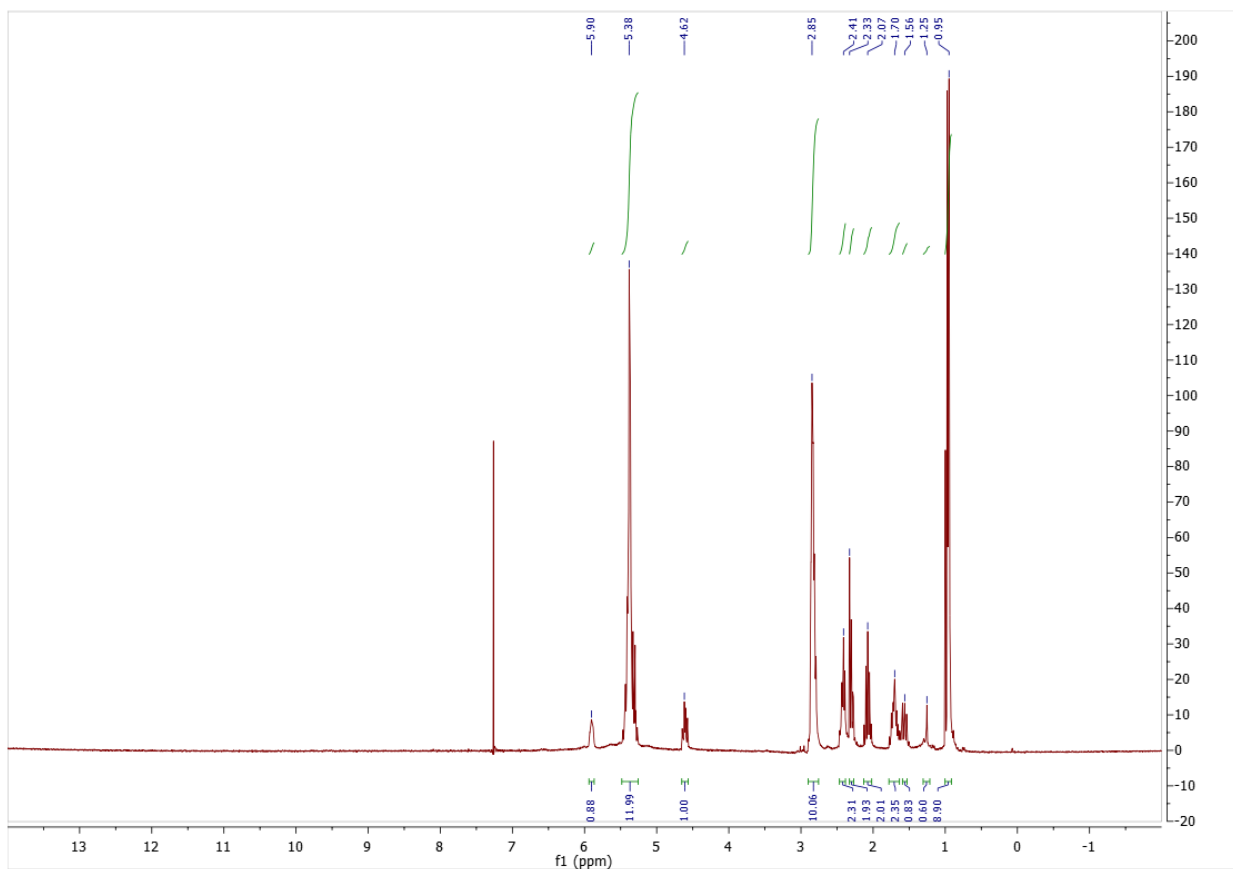
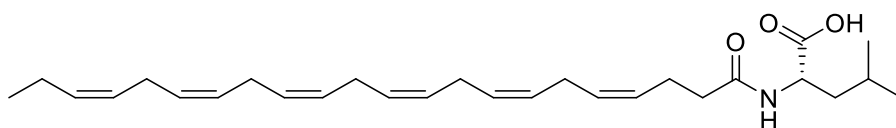


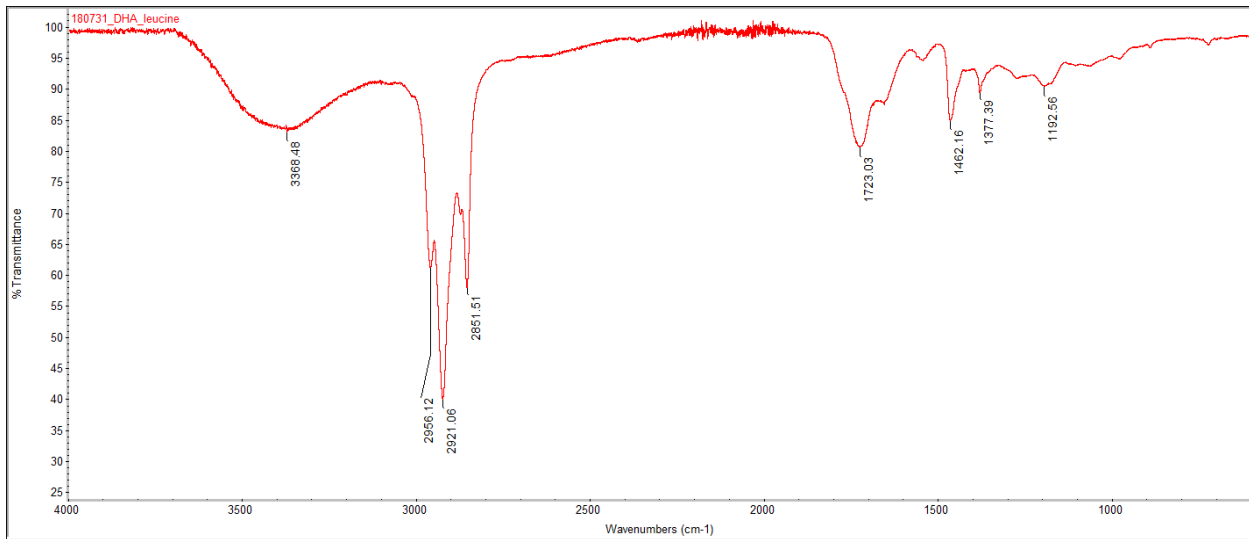
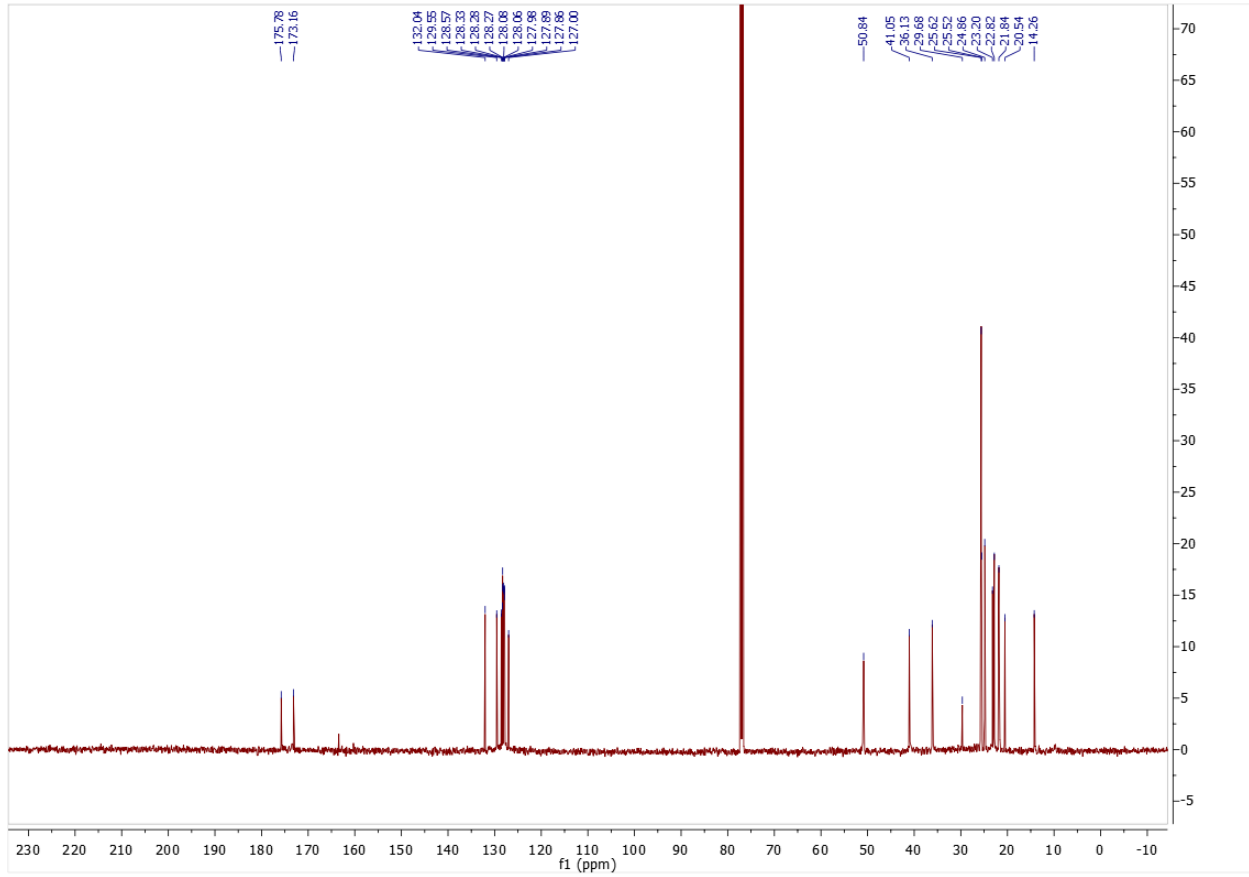
Docosahexaenoyl phenylalanine: ^1H NMR (300 MHz, CDCl_3) δ 0.97 (t, 3H), 2.07 (m, 2H), 2.25 (m, 2H), 2.35 (m, 2H), 2.83 (m, 10H), 3.13 (m, 1H), 3.23 (m, 1H), 4.85 (m, 1H), 5.38 (m, 12H), 5.90 (d, 1H), 7.17 (m, 2H), 7.28 (m, 3H). ^{13}C NMR (500 MHz, CDCl_3): δ 14.26, 20.54, 23.12, 25.52, 25.56, 25.62, 29.68, 36.11, 37.24, 53.14, 127.00, 127.17, 127.81, 127.86, 127.98, 128.07, 128.08, 128.27, 128.28, 128.33, 128.57, 128.61, 129.34, 129.53, 132.03, 135.76, 172.92, 174.04. HRMS (ESI) m/z calculated for $\text{C}_{31}\text{H}_{41}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 474.3014, found: 474.3012. IR shown below.



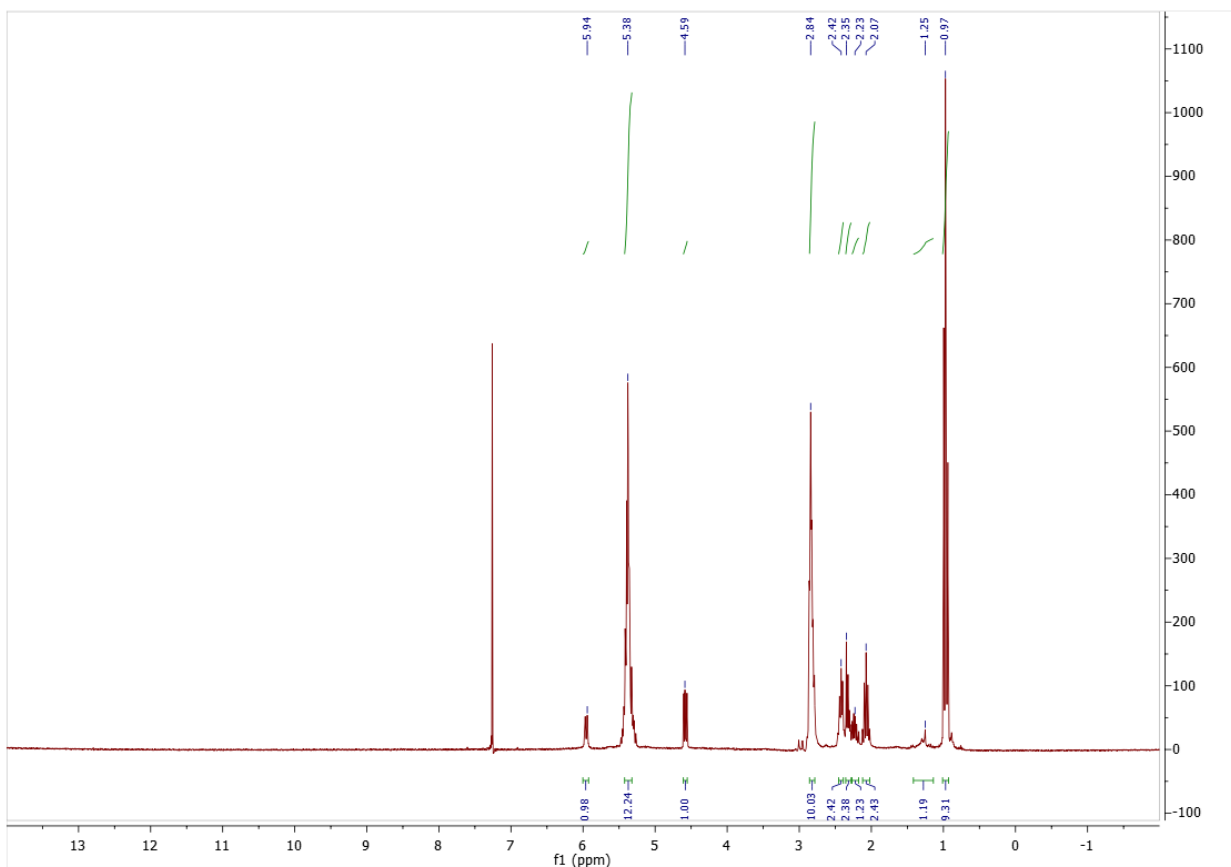
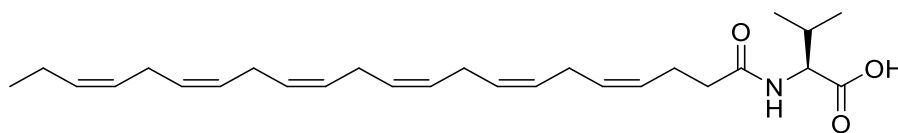


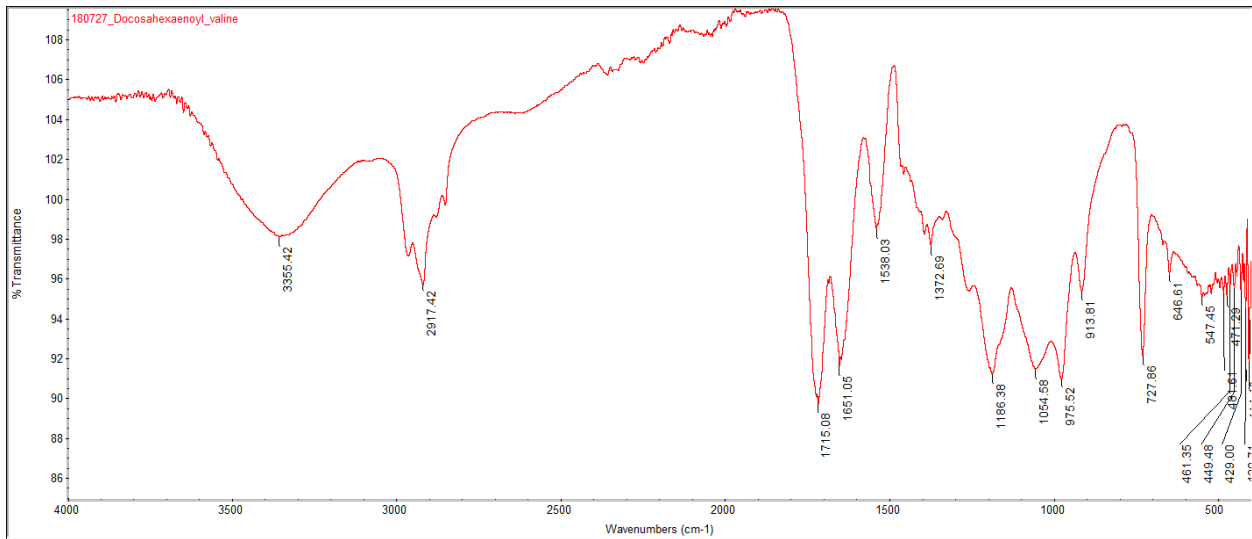
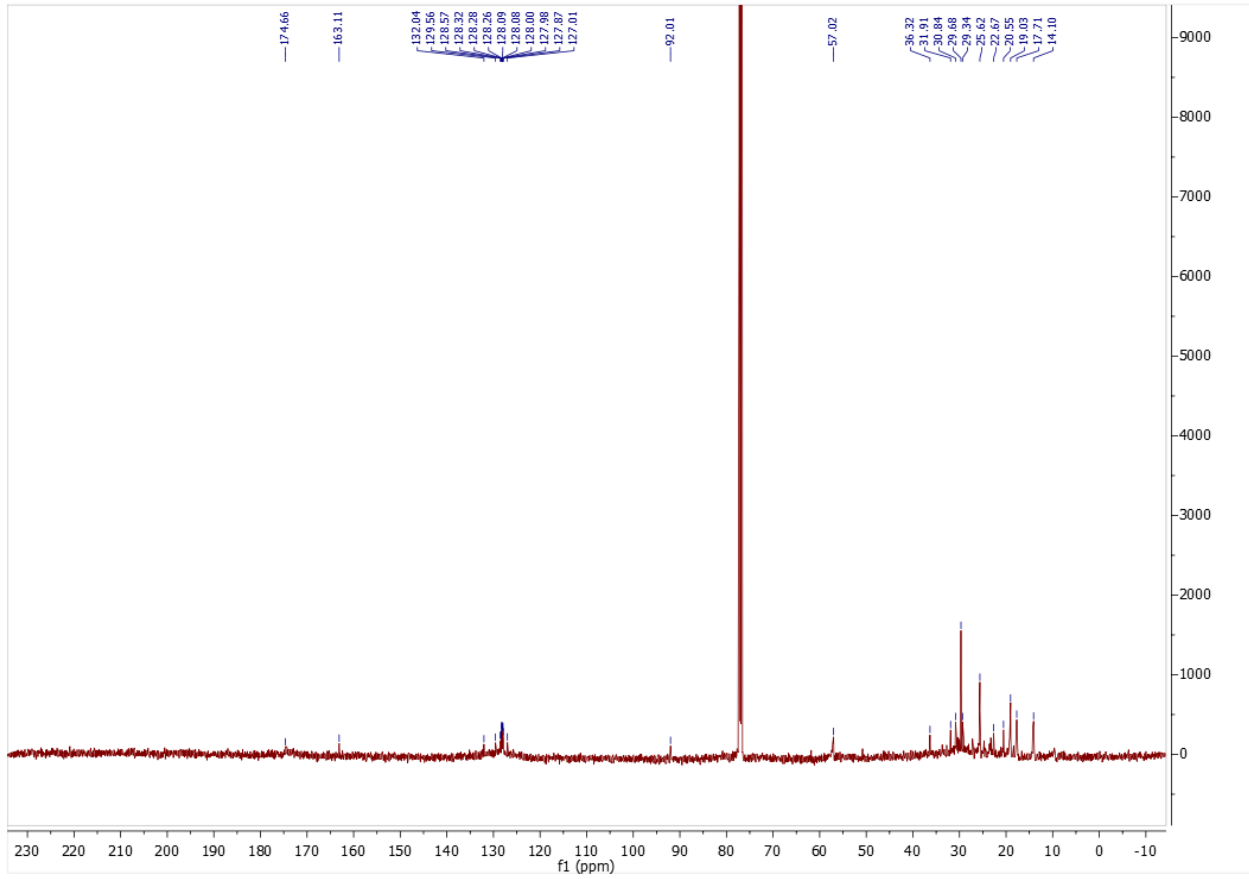
Docosahexaenoyl leucine: ^1H NMR (300 MHz, CDCl_3) δ 0.95 (m, 9H), 1.25 (m, 1H), 1.56 (m, 1H), 1.70 (m, 2H), 2.07 (m, 2H), 2.33 (m, 2H), 2.41 (m, 2H), 2.85 (m, 10H), 4.62 (m, 1H), 5.38 (m, 12H), 5.90 (d, 1H). ^{13}C NMR (500 MHz, CDCl_3): δ 14.26, 20.54, 21.84, 22.82, 23.20, 24.86, 25.52, 25.62, 29.68, 36.13, 41.05, 50.84, 127.00, 127.86, 127.89, 127.98, 128.06, 128.08, 128.27, 128.28, 128.33, 128.57, 129.55, 132.04, 173.16, 175.78. HRMS (ESI) m/z calculated for $\text{C}_{28}\text{H}_{43}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 440.3170, found: 440.3170. IR shown below.



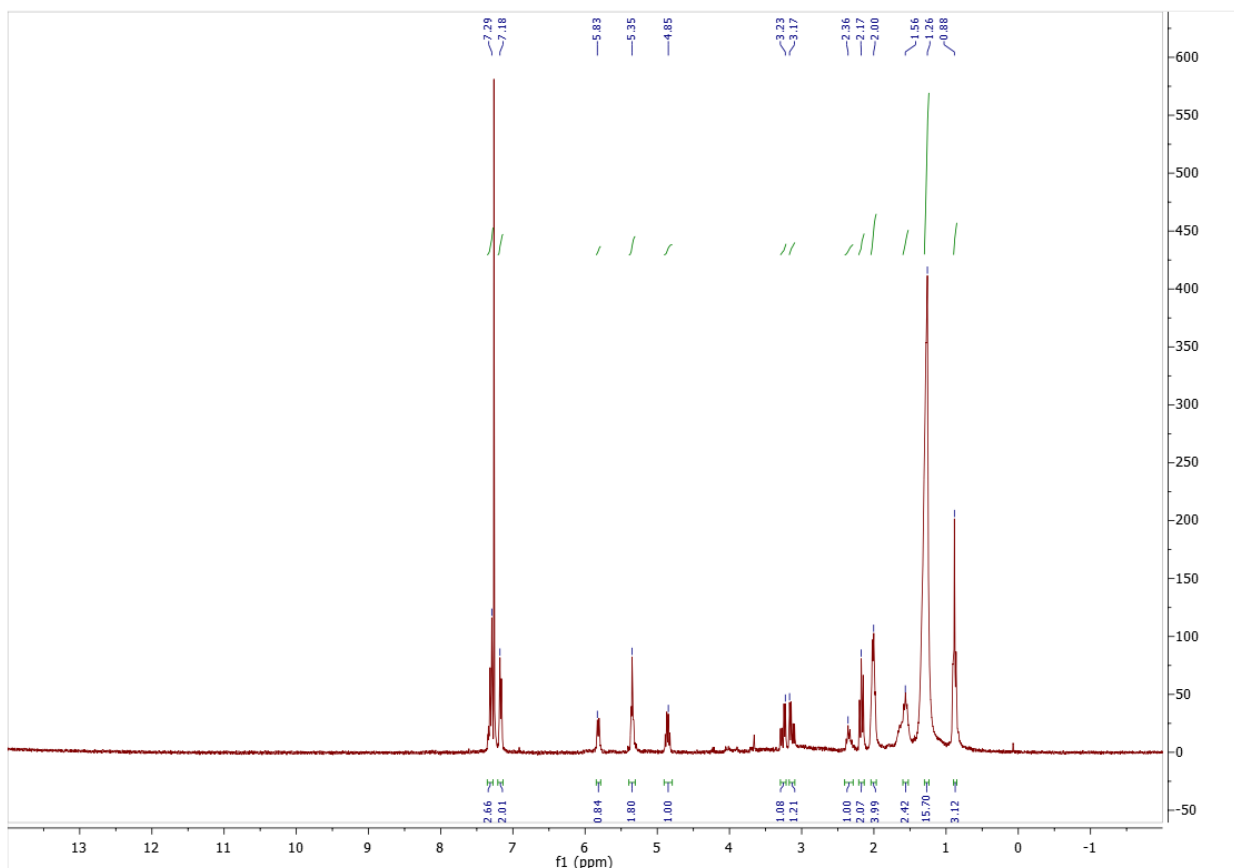
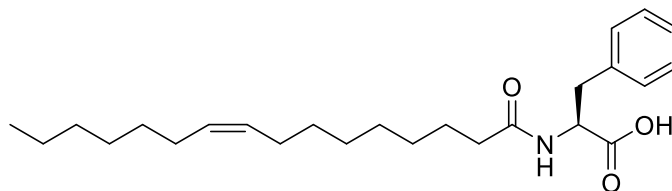


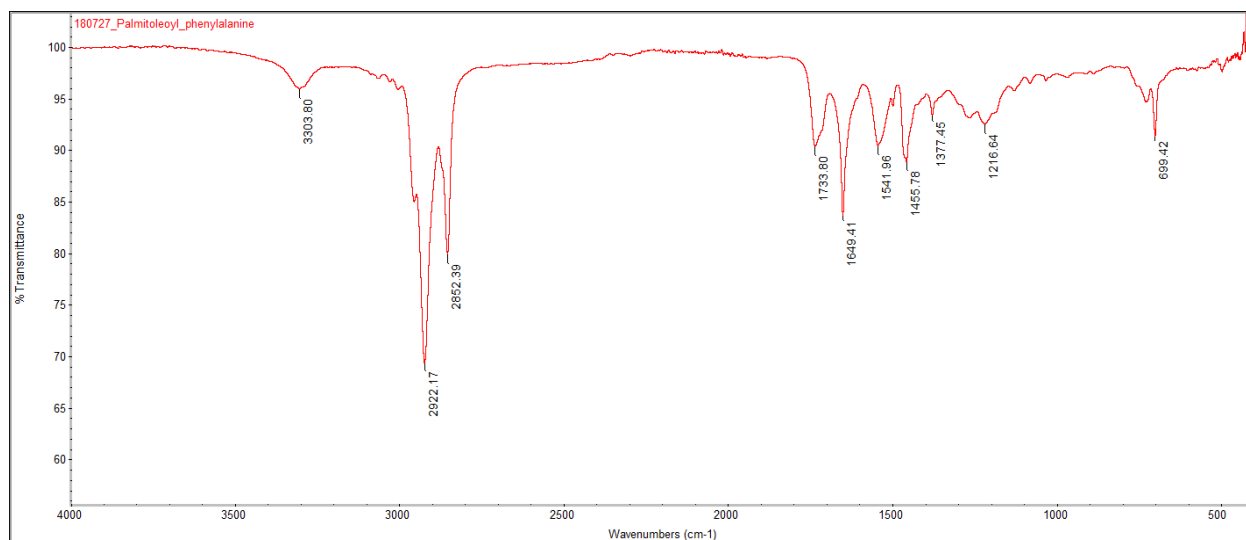
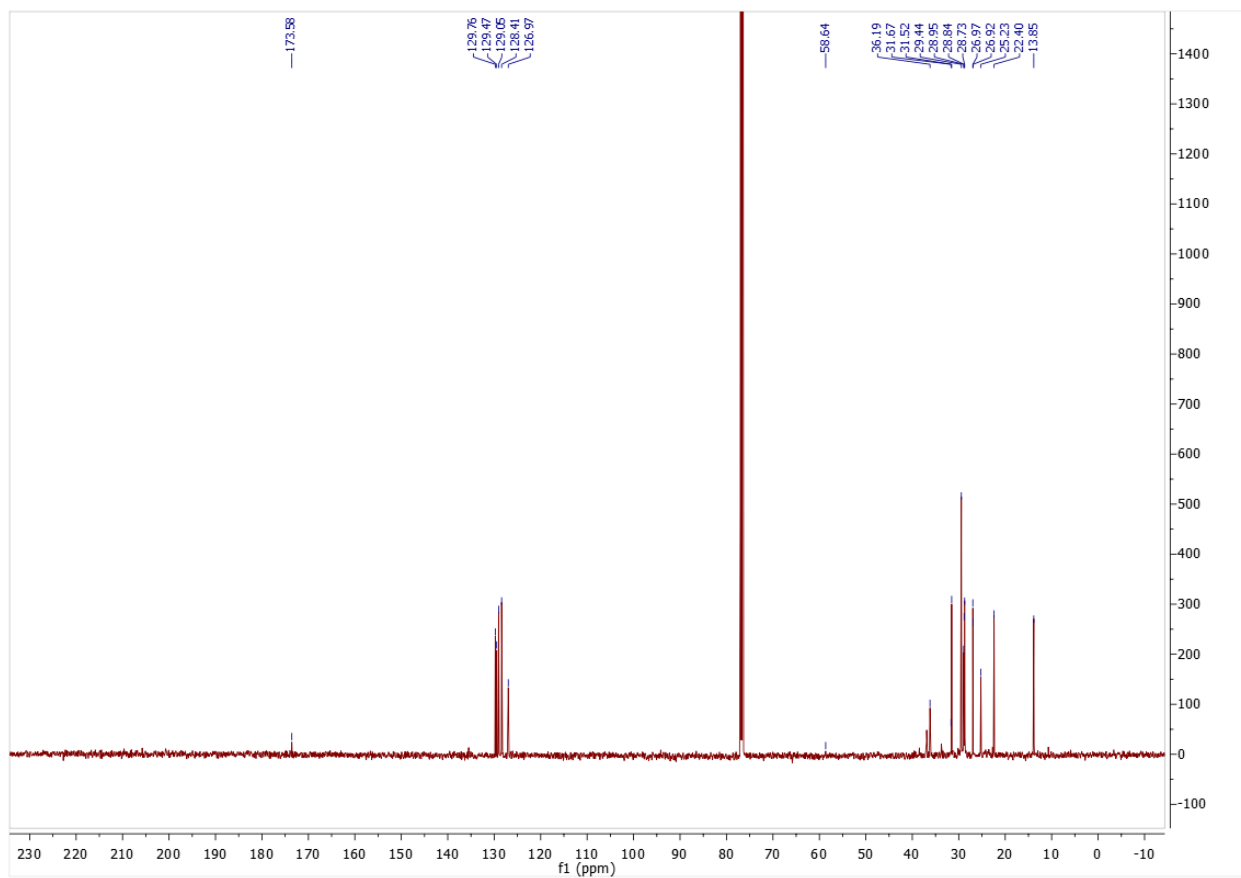
Docosahexaenoyl valine: ^1H NMR (300 MHz, CDCl_3) δ 0.97 (m, 9H), 1.25 (m, 1H), 2.07 (m, 2H), 2.23 (m, 1H), 2.35 (m, 2H), 2.42 (m, 2H), 2.84 (m, 10H), 4.59 (m, 1H), 5.38 (m, 12H), 5.94 (d, 1H). ^{13}C NMR (500 MHz, CDCl_3): δ 14.10, 17.71, 19.03, 20.55, 22.67, 25.62, 29.34, 29.68, 30.84, 31.91, 36.32, 57.02, 92.01, 127.01, 127.87, 127.98, 128.00, 128.08, 128.09, 128.26, 128.28, 128.32, 128.57, 129.56, 132.04, 163.11, 174.66. HRMS (ESI) m/z calculated for $\text{C}_{27}\text{H}_{41}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 426.3014, found: 426.3019. IR shown below.



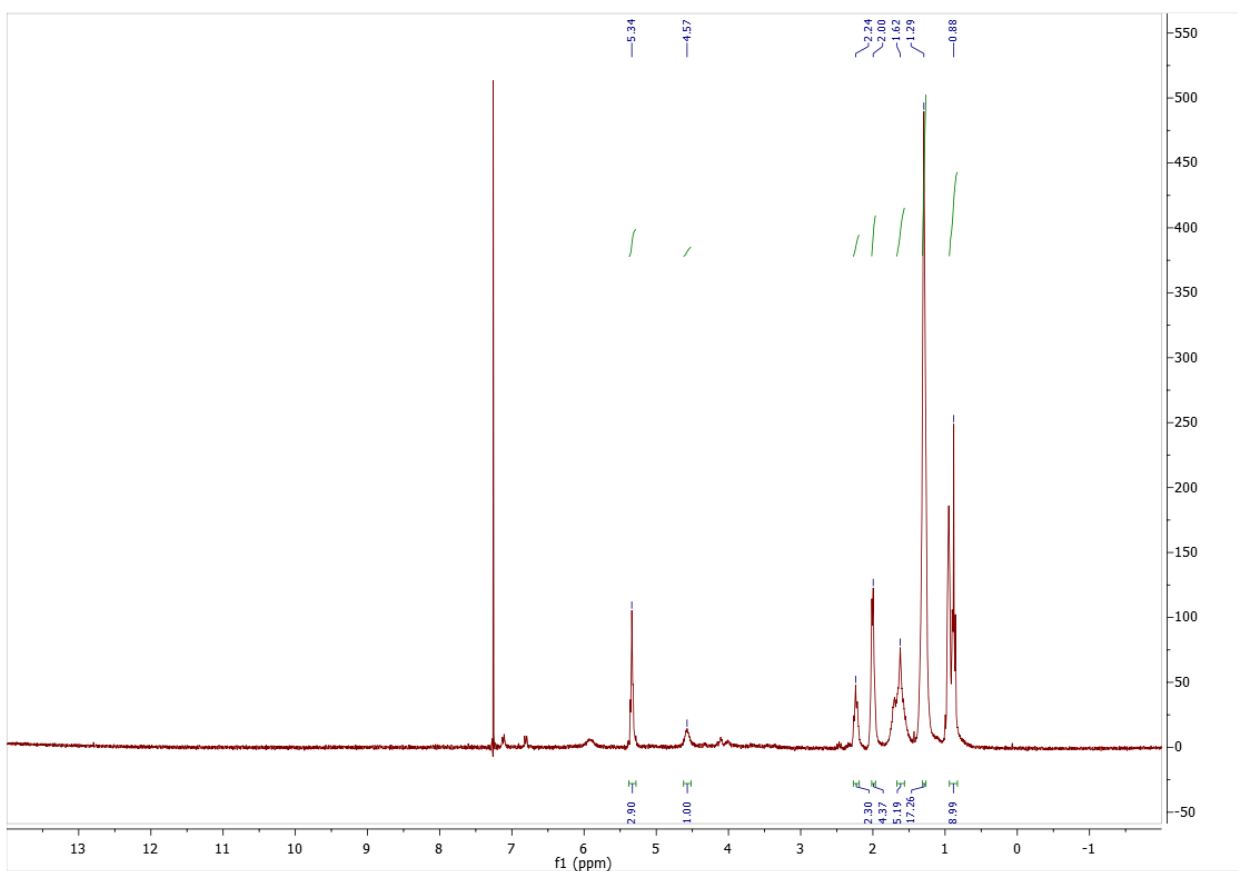
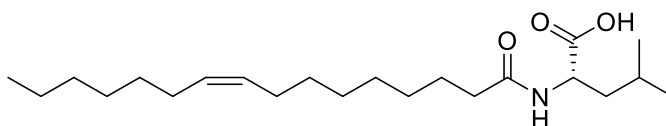


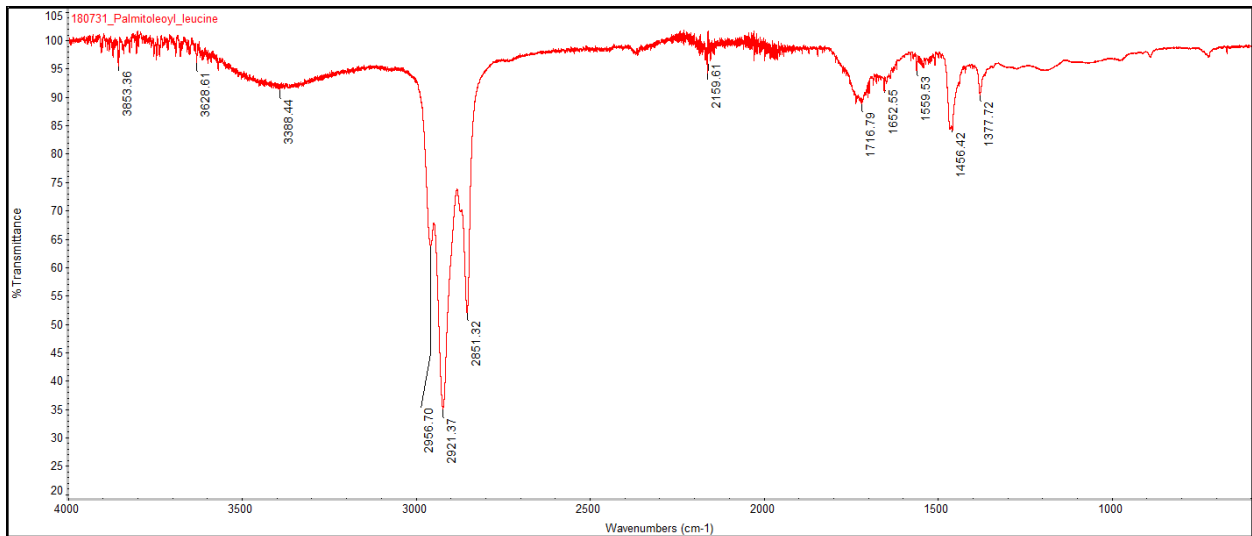
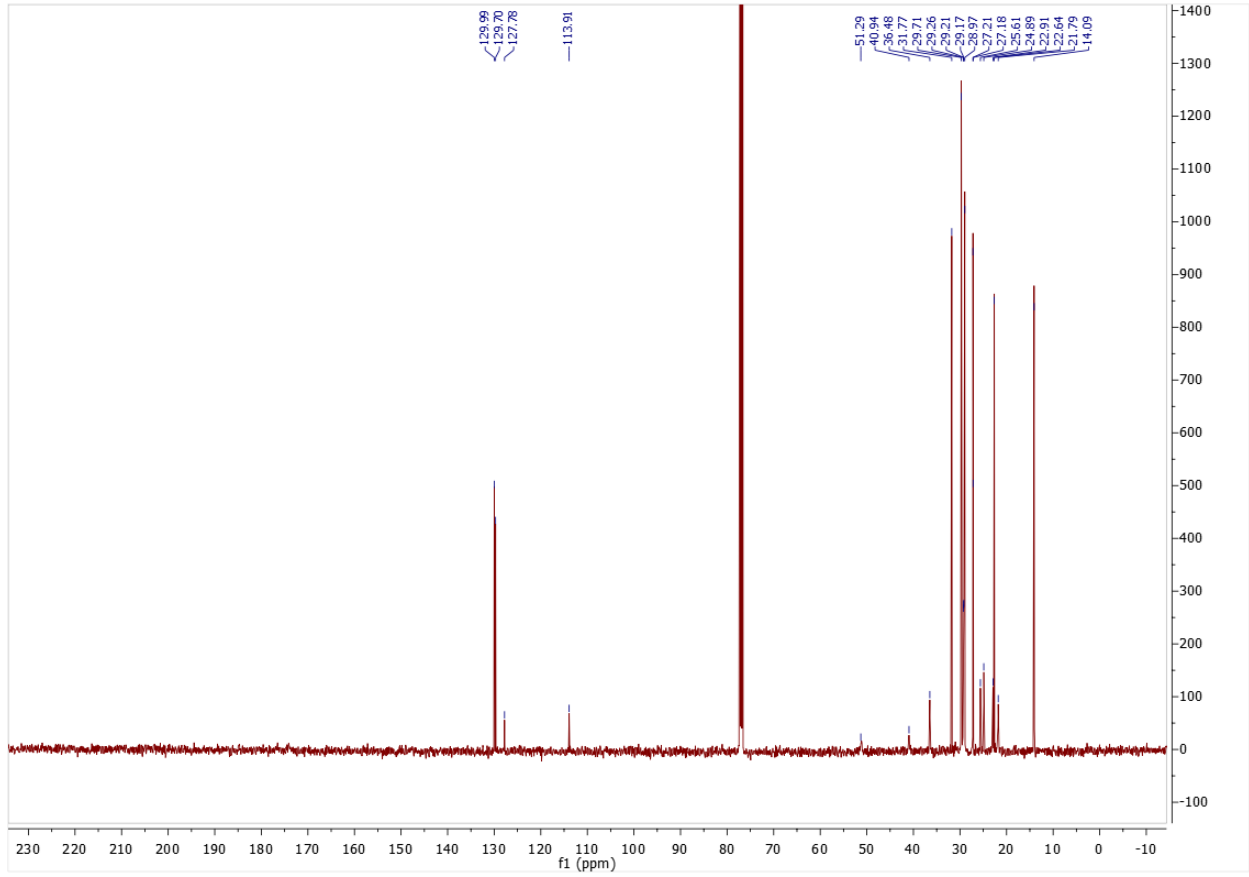
Palmitoleoyl phenylalanine: ^1H NMR (300 MHz, CDCl_3) δ 0.88 (t, 3H), 1.26 (m, 16H), 1.56 (m, 2H), 2.00 (m, 4H), 2.17 (m, 2H), 2.36 (m, 1H), 3.17 (dd, 1H), 3.23 (dd, 1H), 4.85 (m, 1H), 5.35 (m, 2H), 5.83 (d, 1H), 7.18 (m, 2H), 7.29 (m, 3H). ^{13}C NMR (500 MHz, CDCl_3): δ 13.85, 22.40, 25.23, 26.92, 26.97, 28.84, 28.95, 29.44, 31.52, 31.67, 36.19, 58.64, 126.97, 128.41, 129.05, 129.47, 129.76, 173.58. HRMS (ESI) m/z calculated for $\text{C}_{25}\text{H}_{39}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 400.2857, found: 400.2857. IR shown below.



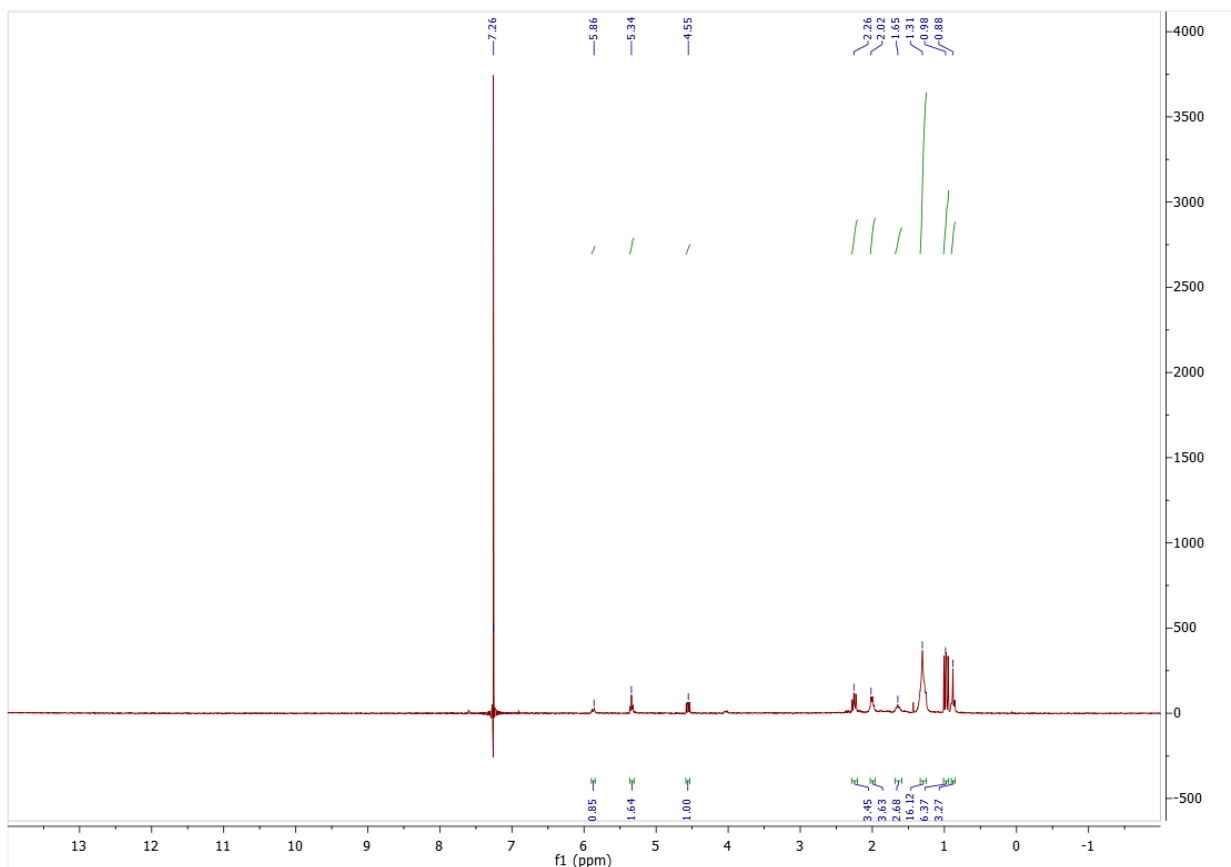
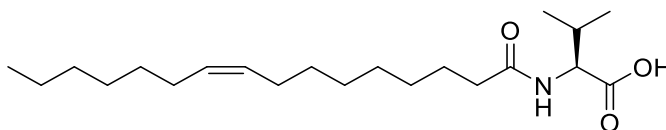


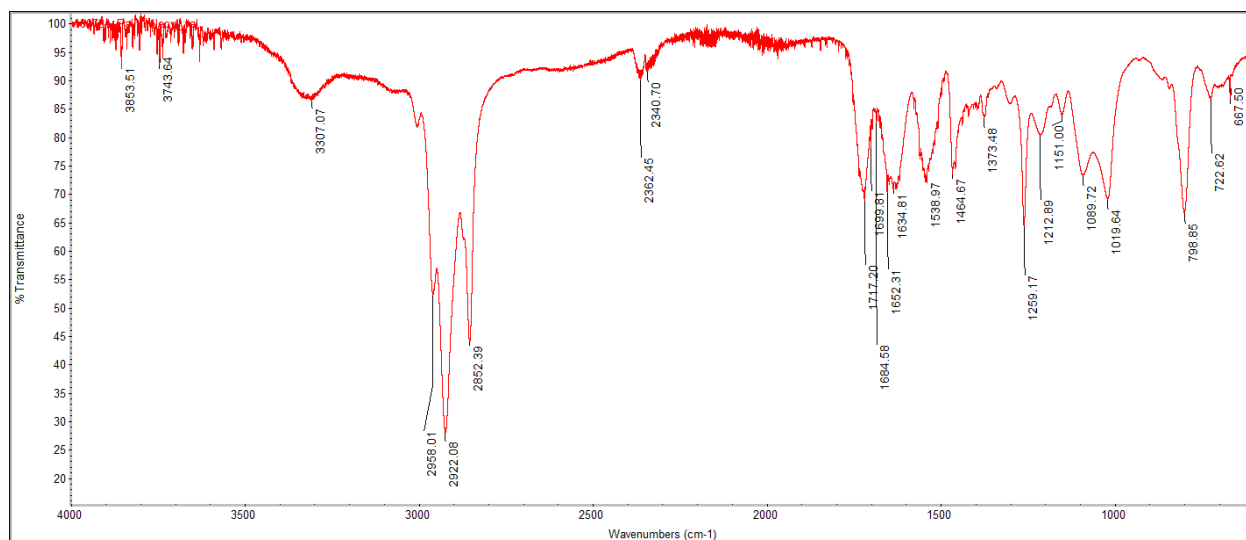
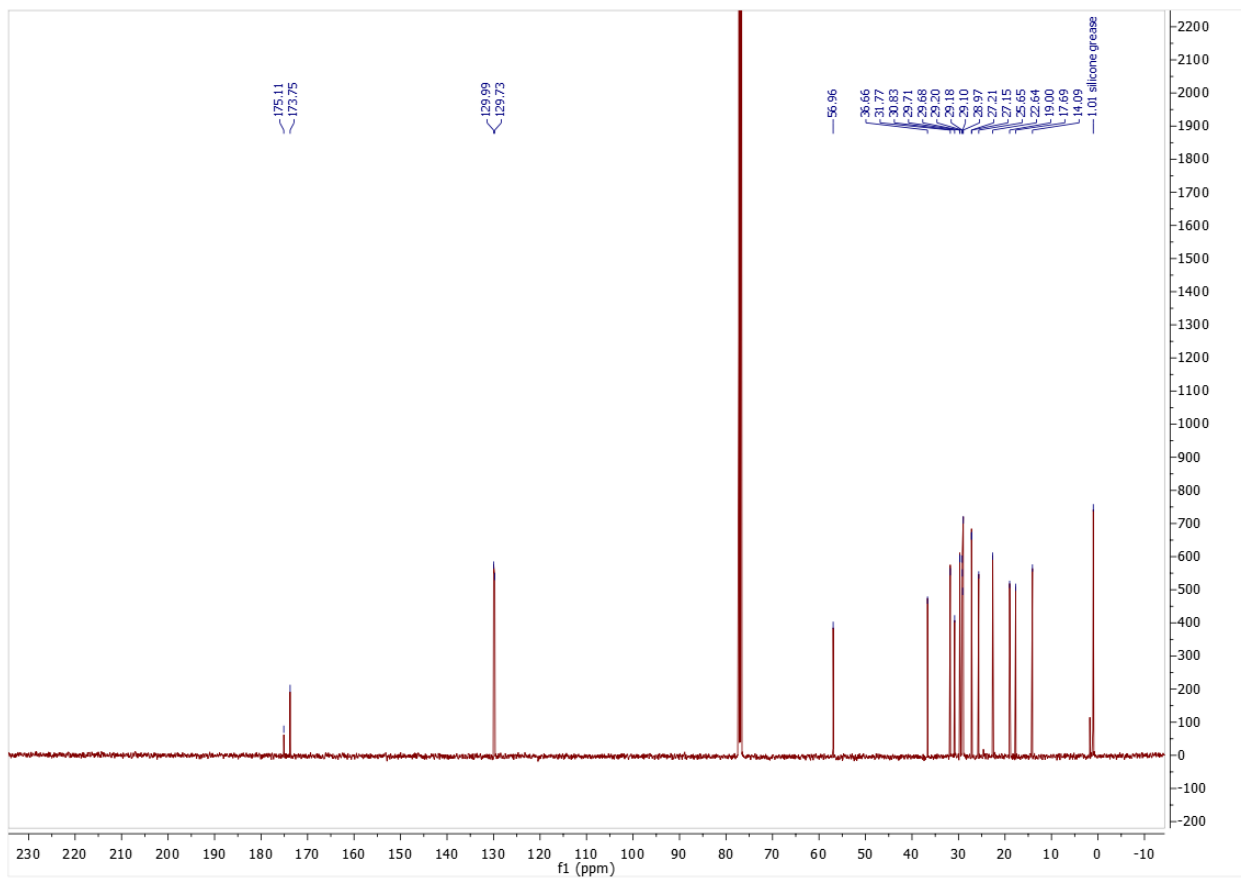
Palmitoleoyl leucine: ^1H NMR (300 MHz, CDCl_3) δ 0.88 (m, 9H), 1.29 (m, 17H), 1.62 (m, 5H), 2.00 (m, 4H), 2.24 (m, 2H), 4.57 (m, 1H), 5.34 (m, 3H). ^{13}C NMR (500 MHz, CDCl_3): 14.09, 21.79, 22.64, 22.91, 24.89, 25.61, 27.18, 27.21, 28.97, 29.17, 29.21, 29.26, 29.71, 31.77, 36.48, 40.94, 51.29, 113.91, 172.78, 129.70, 129.99. HRMS (ESI) m/z calculated for $\text{C}_{22}\text{H}_{41}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 366.3014, found: 366.3017. IR shown below.



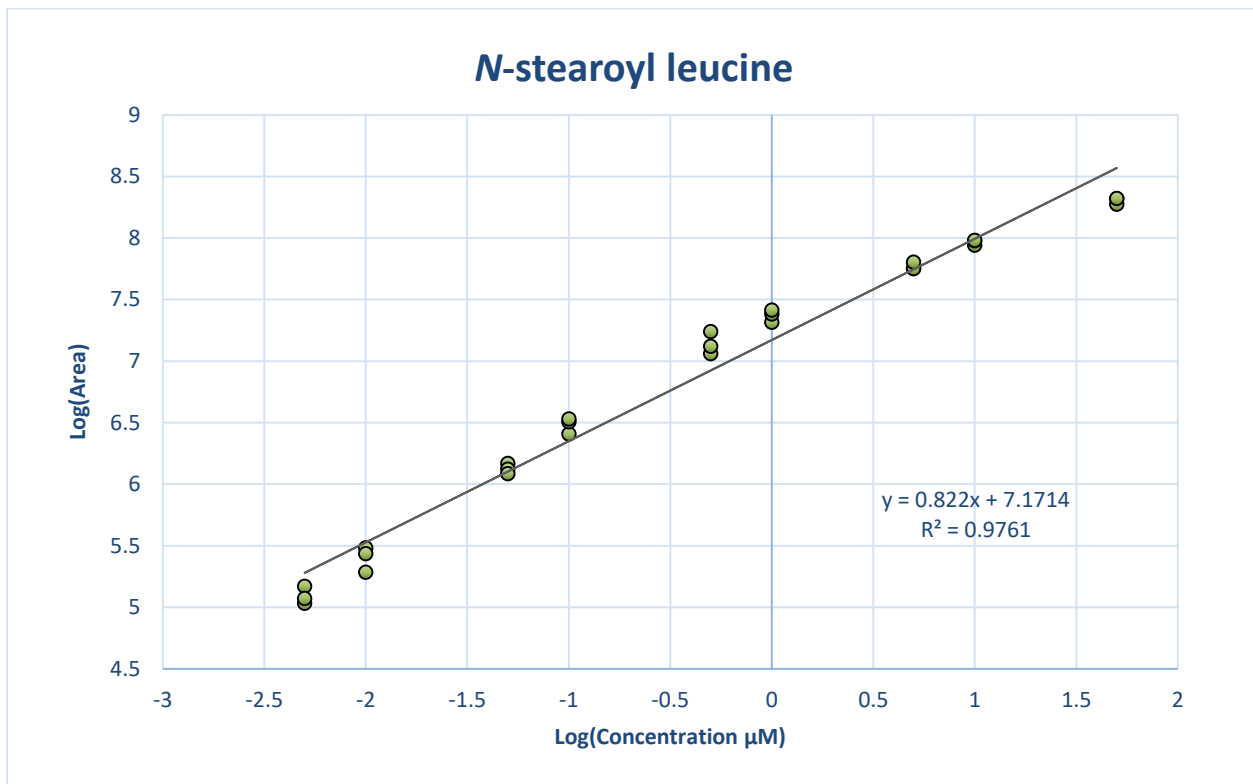
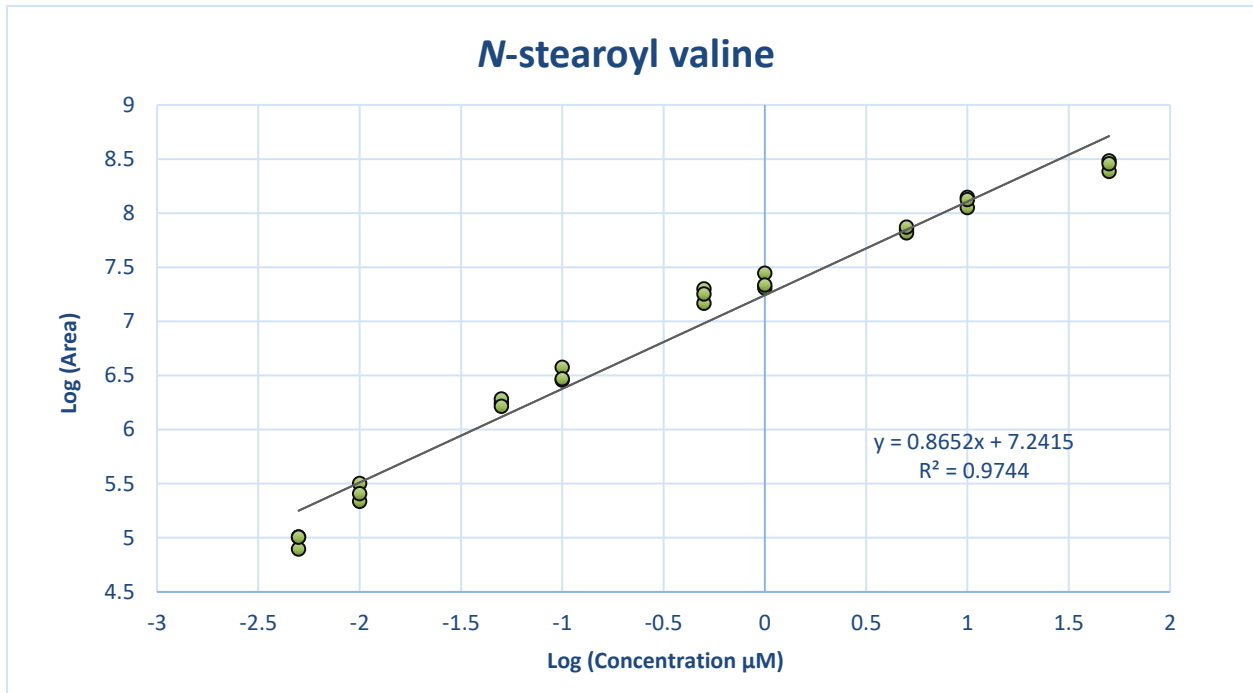


Palmitoleoyl valine: ^1H NMR (300 MHz, CDCl_3) δ 0.88 (m, 3H), 0.98 (m, 6H), 1.31 (m, 16H), 1.65 (m, 3H), 2.02 (m, 4H), 2.26 (m, 3H), 4.55 (m, 1H), 5.34 (m, 2H), 5.86 (d, 1H). ^{13}C NMR (500 MHz, CDCl_3): δ 14.09, 17.69, 19.00, 22.64, 25.65, 27.15, 28.97, 29.10, 29.18, 29.20, 29.68, 29.71, 30.83, 31.77, 36.66, 129.73, 129.99, 173.75, 175.11. HRMS (ESI) m/z calculated for $\text{C}_{21}\text{H}_{39}\text{NO}_3$ $[\text{M}-\text{H}]^-$ 352.2857, found: 352.2861. IR shown below.

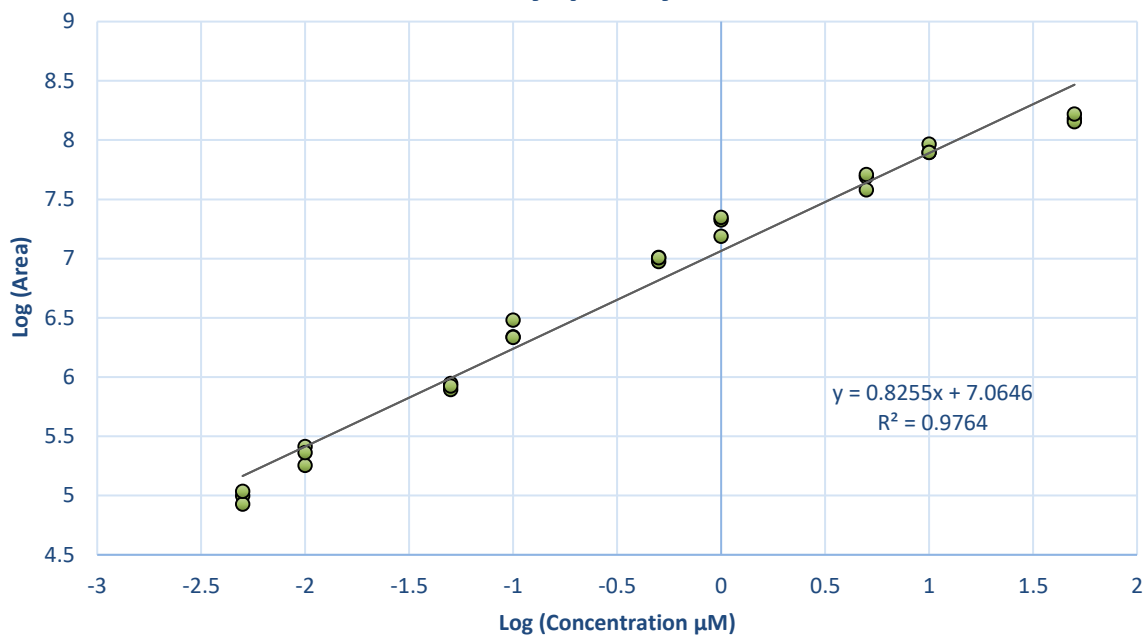




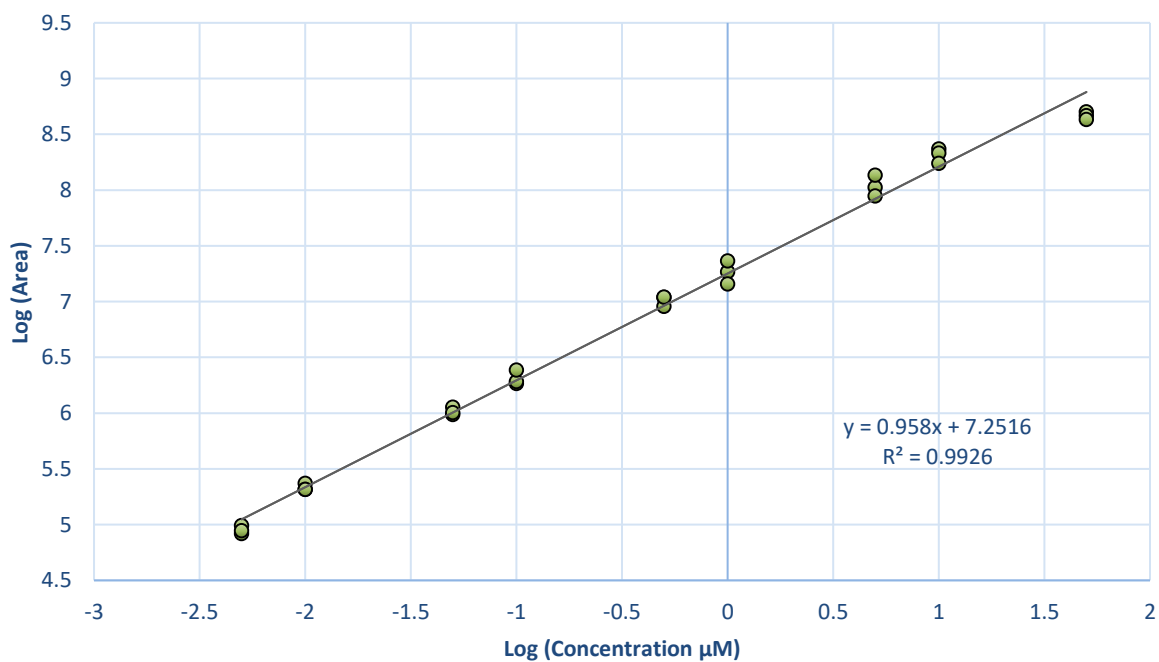
B2: Calibration plots for N-Acyl amino acid standards



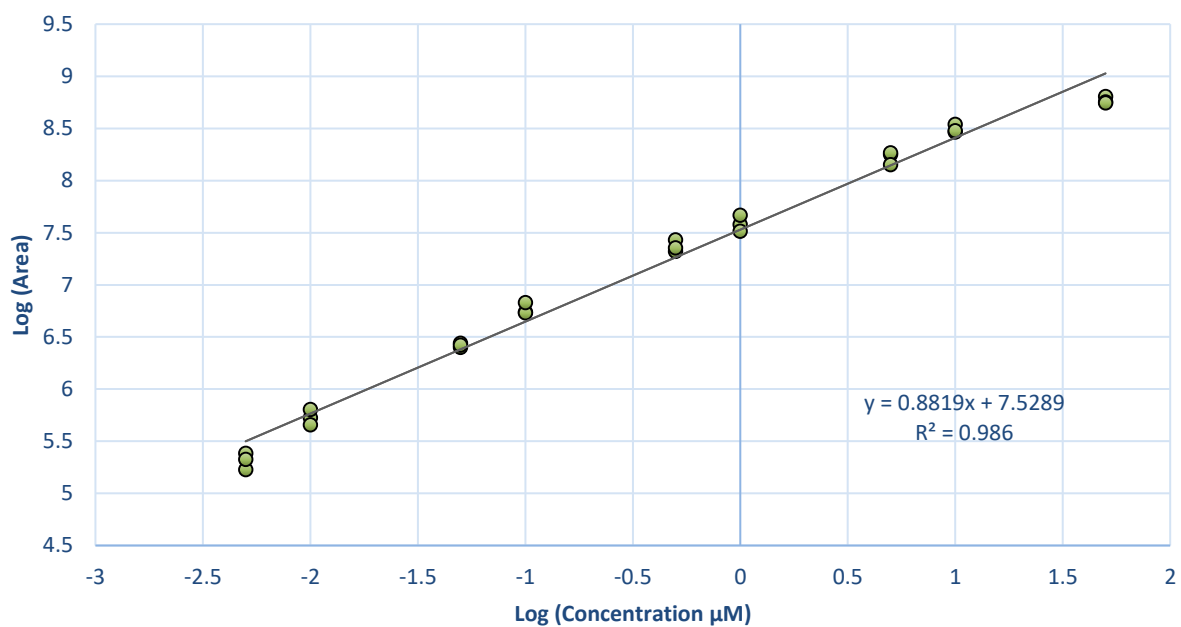
N-stearoyl phenylalanine



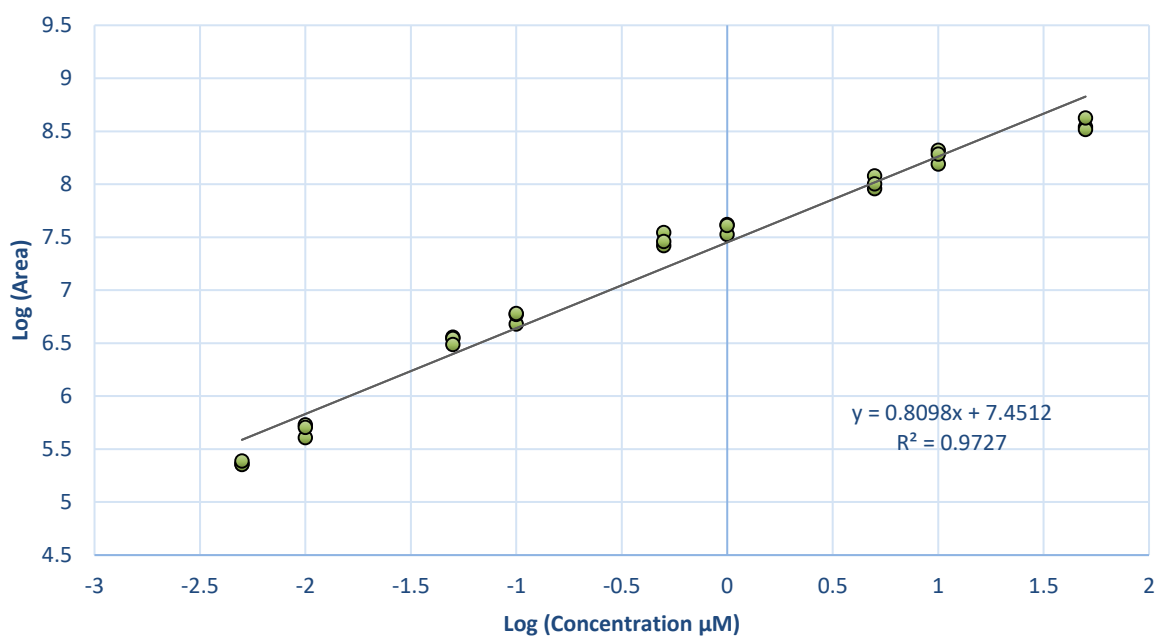
N-pamitoleoyl valine



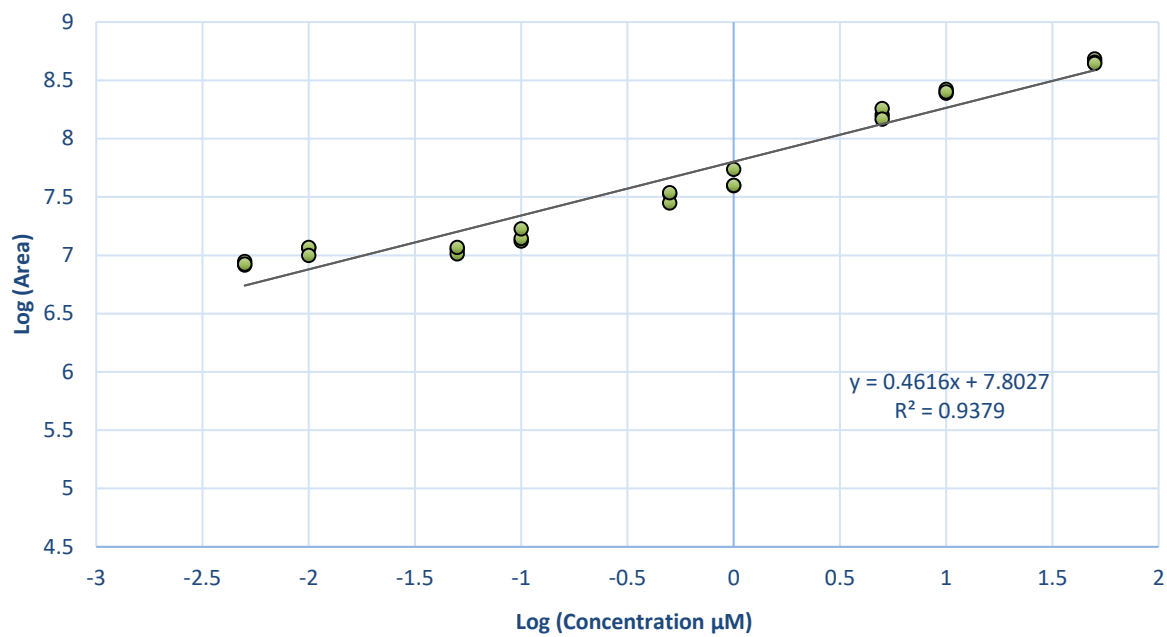
***N*-palmitoyl valine**



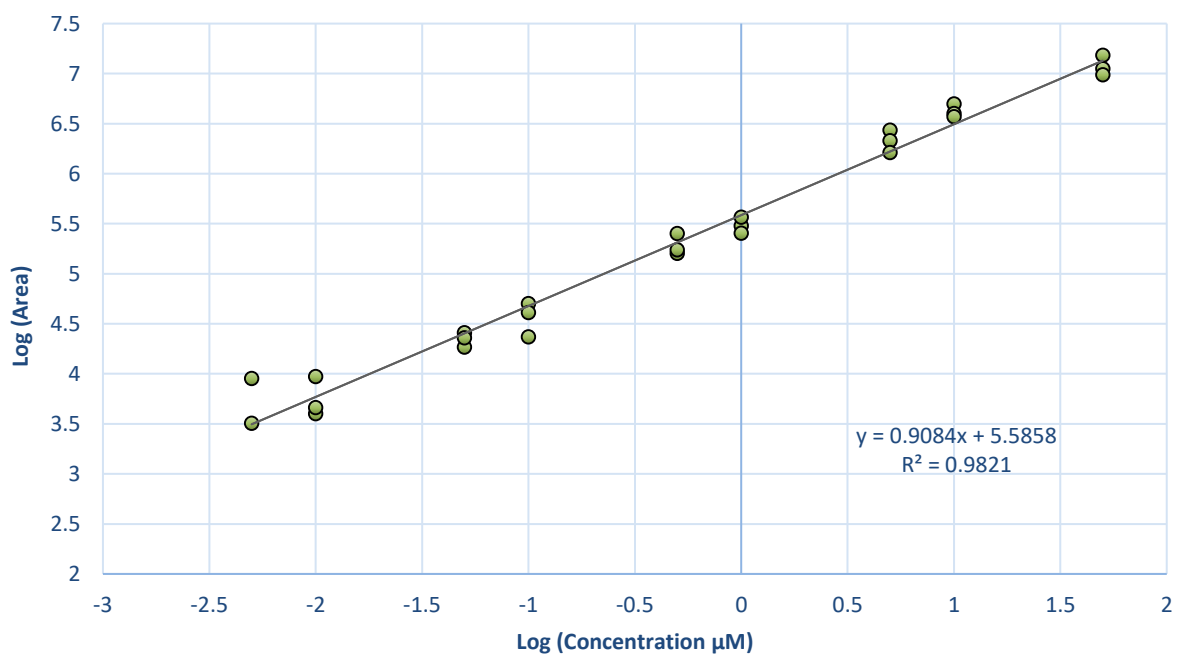
***N*-palmitoyl phenylalanine**



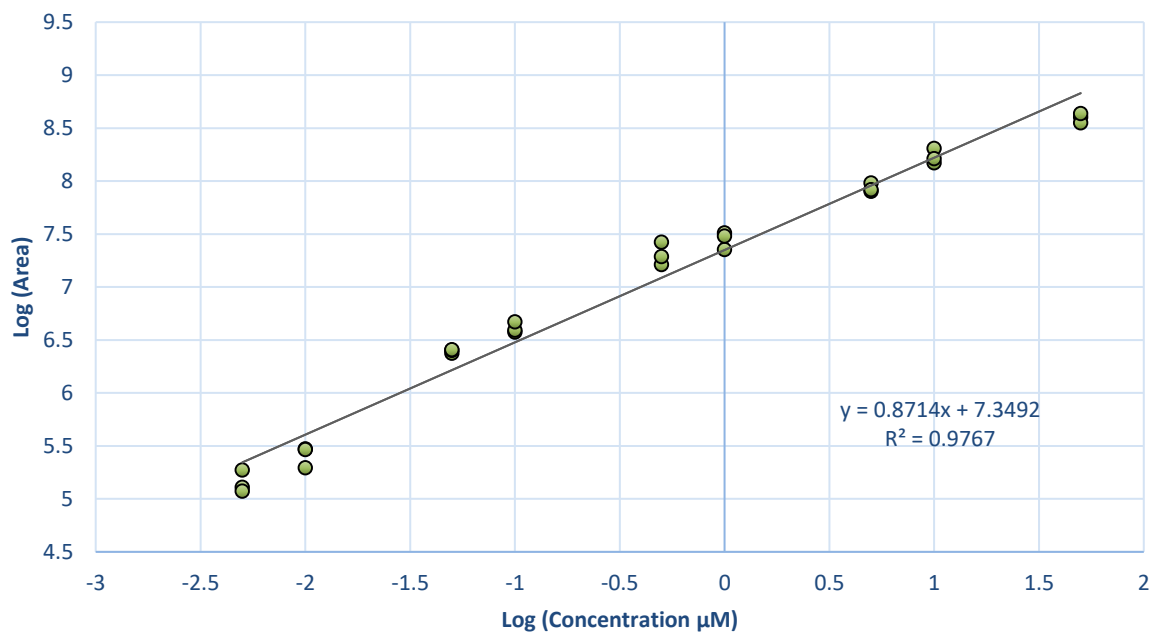
N-palmitoyl leucine



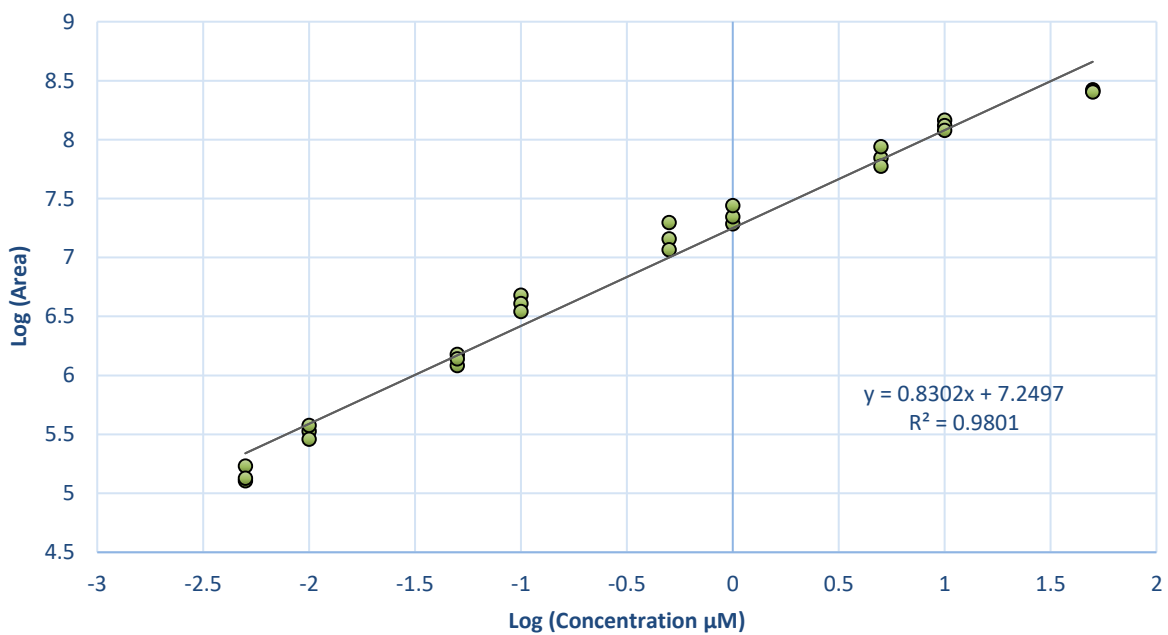
N-palmitoleyl leucine



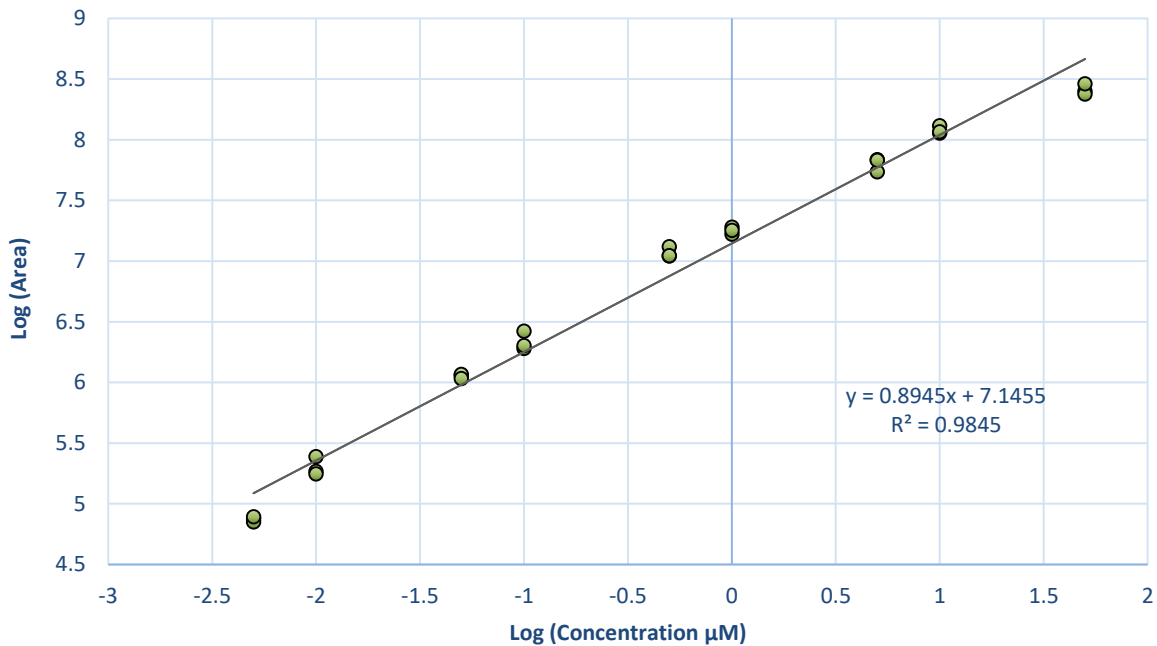
N-oleoyl valine



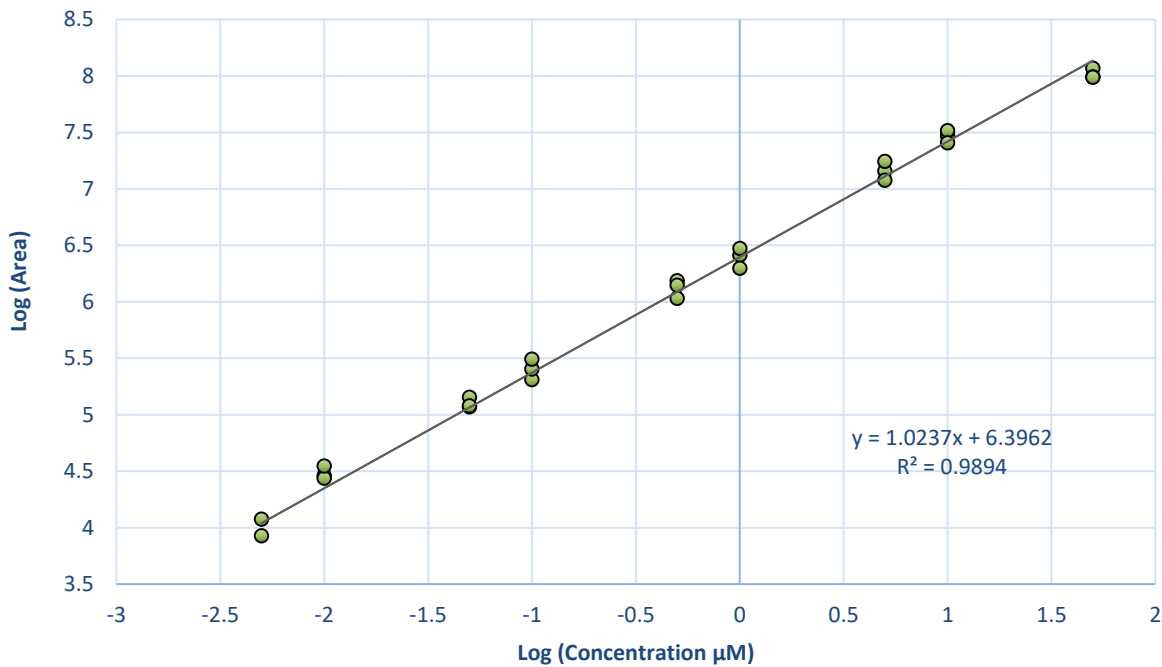
N-oleoyl phenylalanine



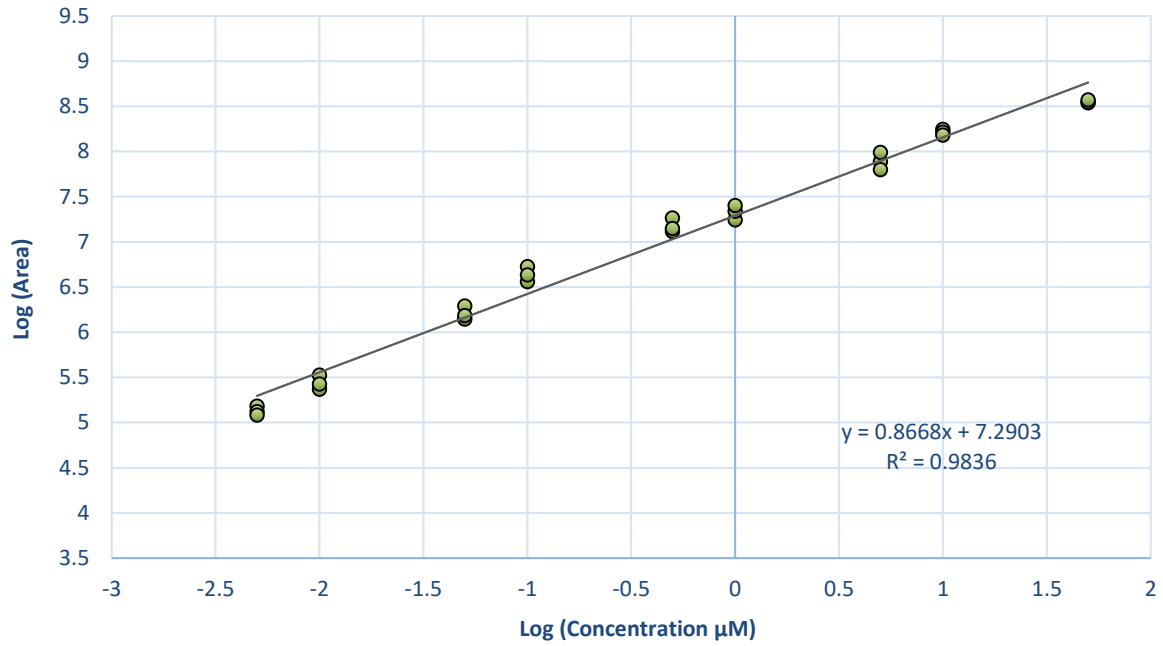
N-oleoyl leucine



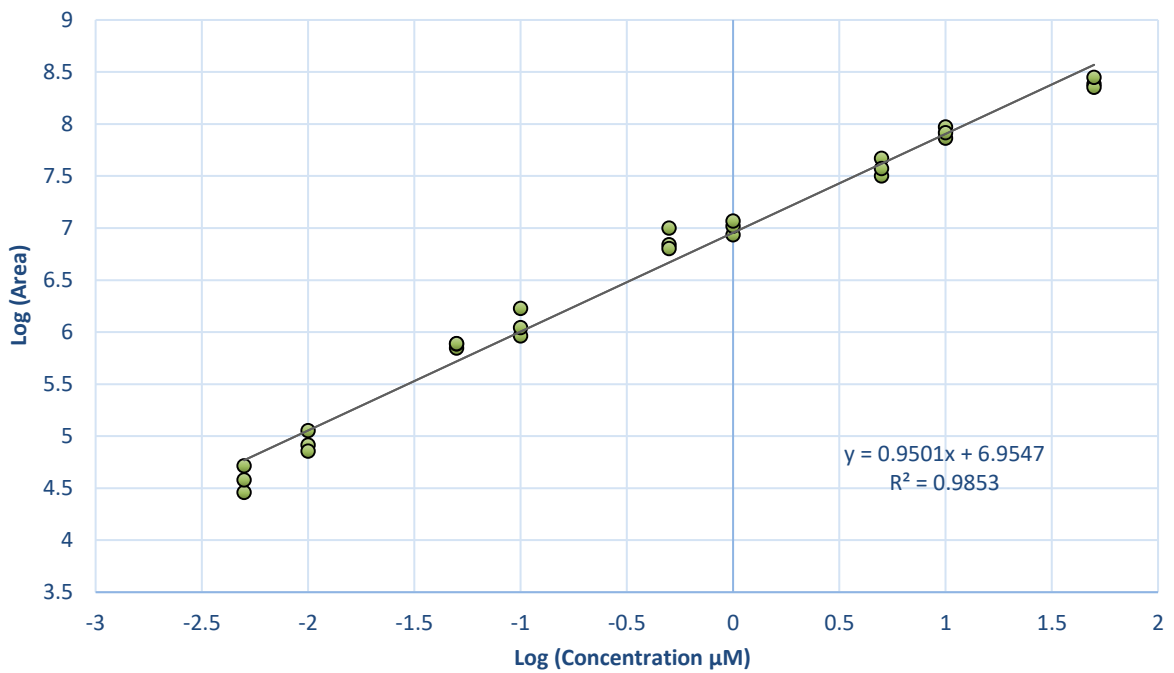
N-linoleoyl valine



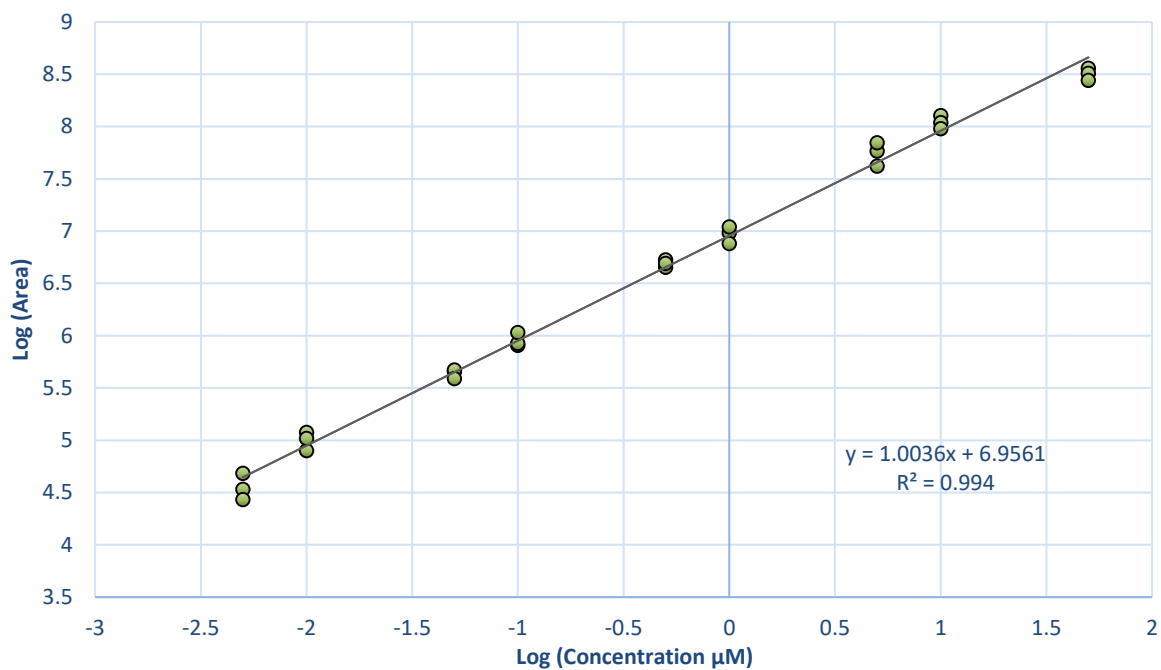
N-linoleoyl phenylalanine



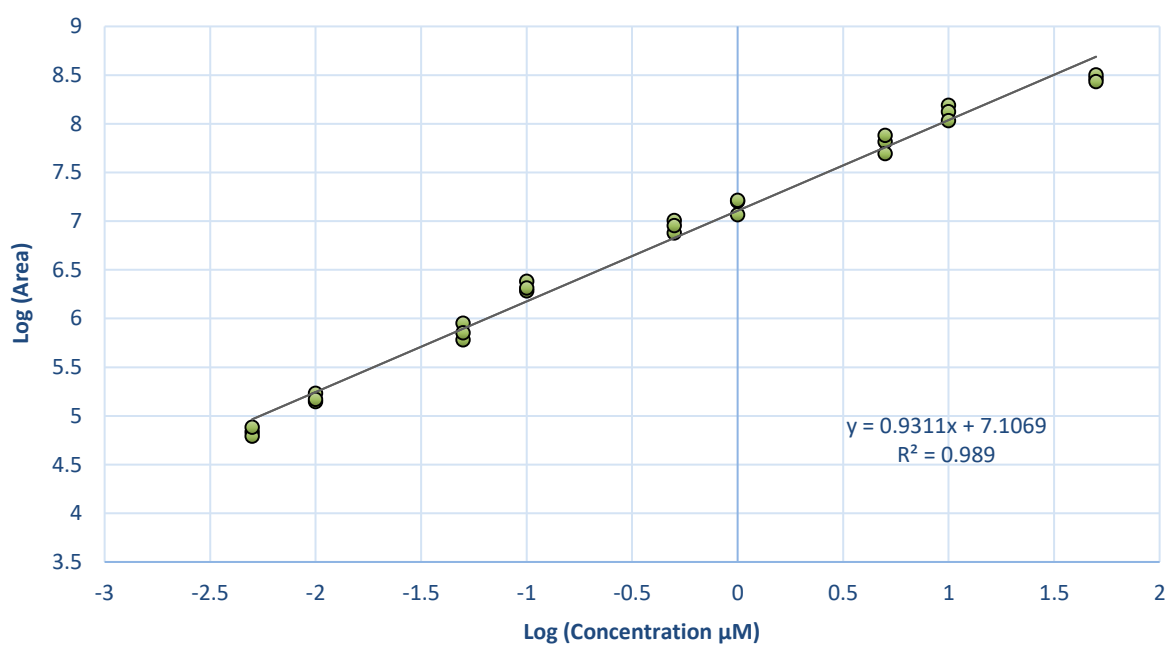
N-linoleoyl leucine



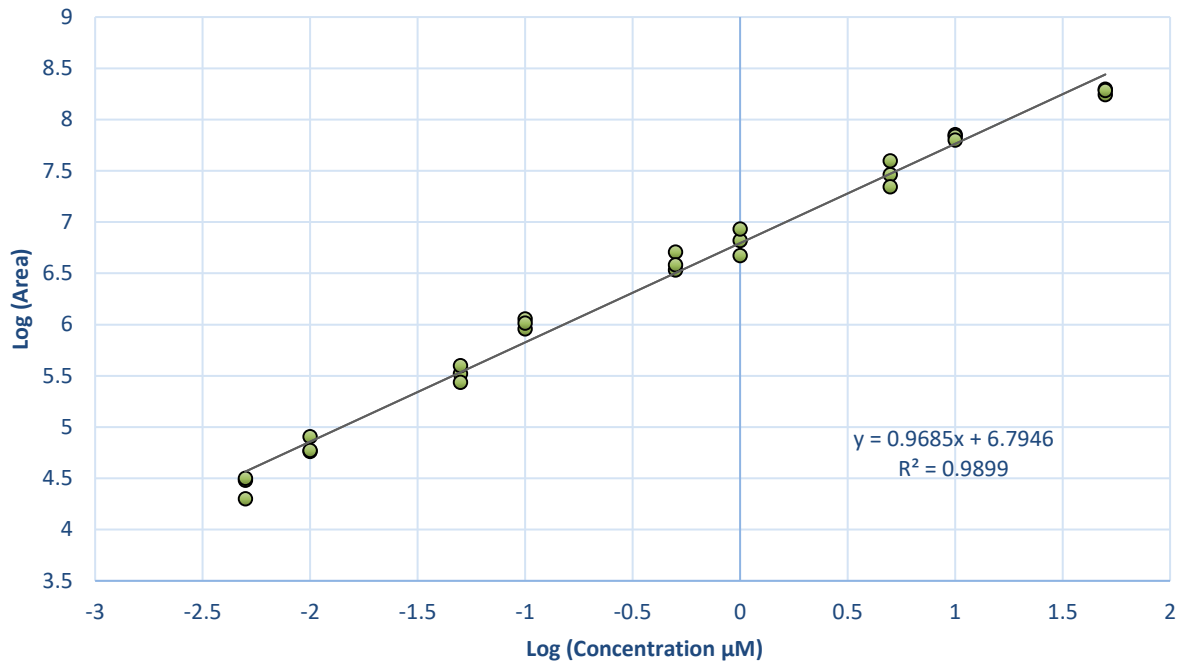
N-eicosapentaenoyl valine



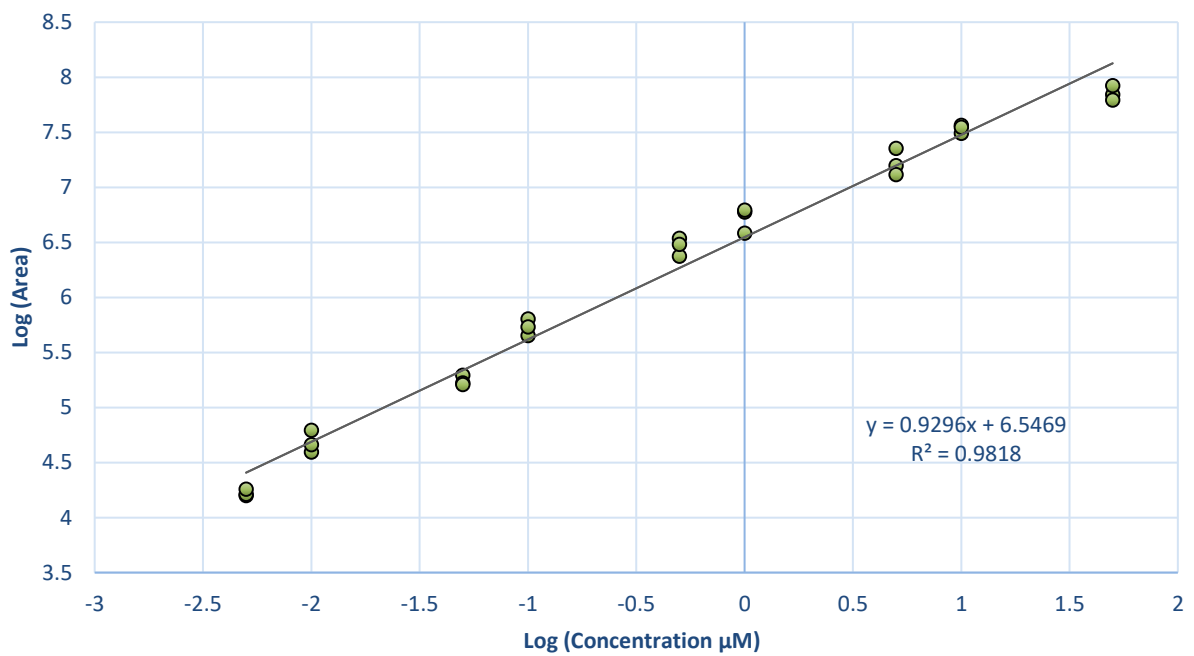
N-eicosapentaenoyl phenylalanine



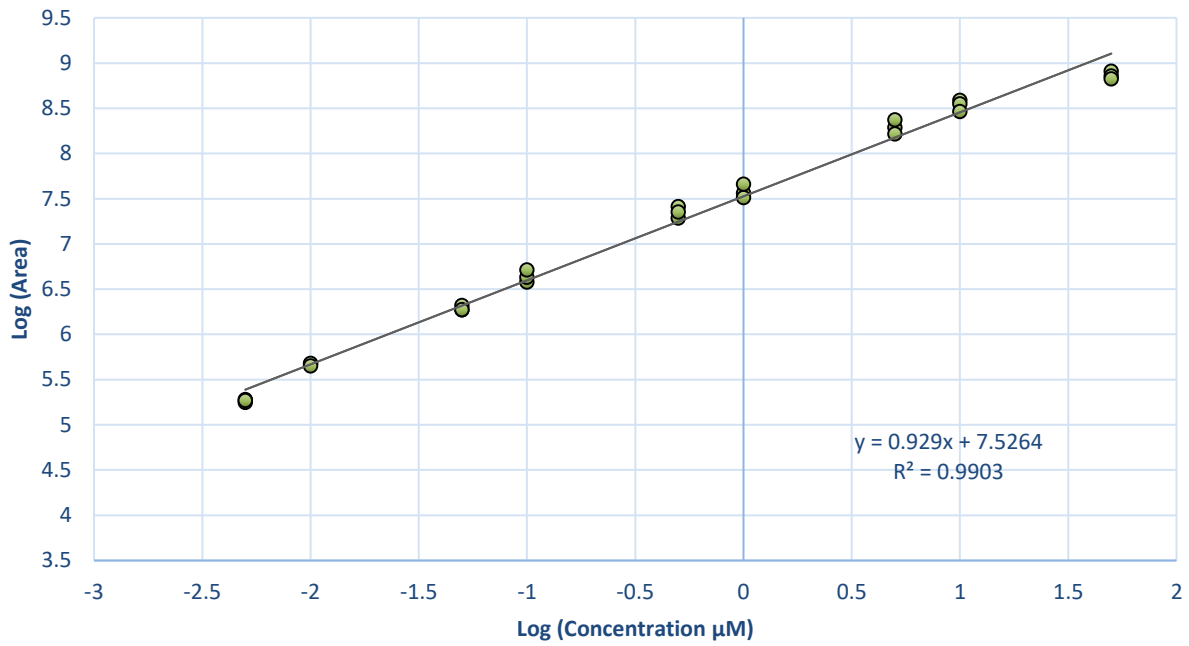
N-eicosapentaenoyl leucine



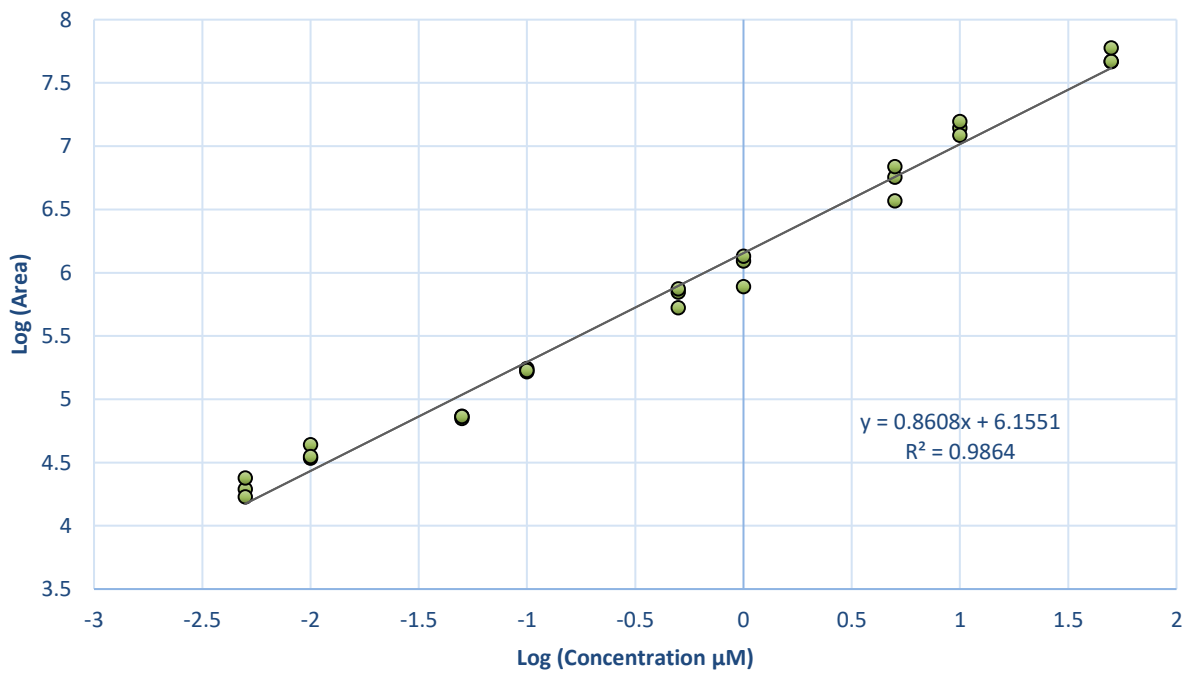
N-eicosanoyl phenylalanine



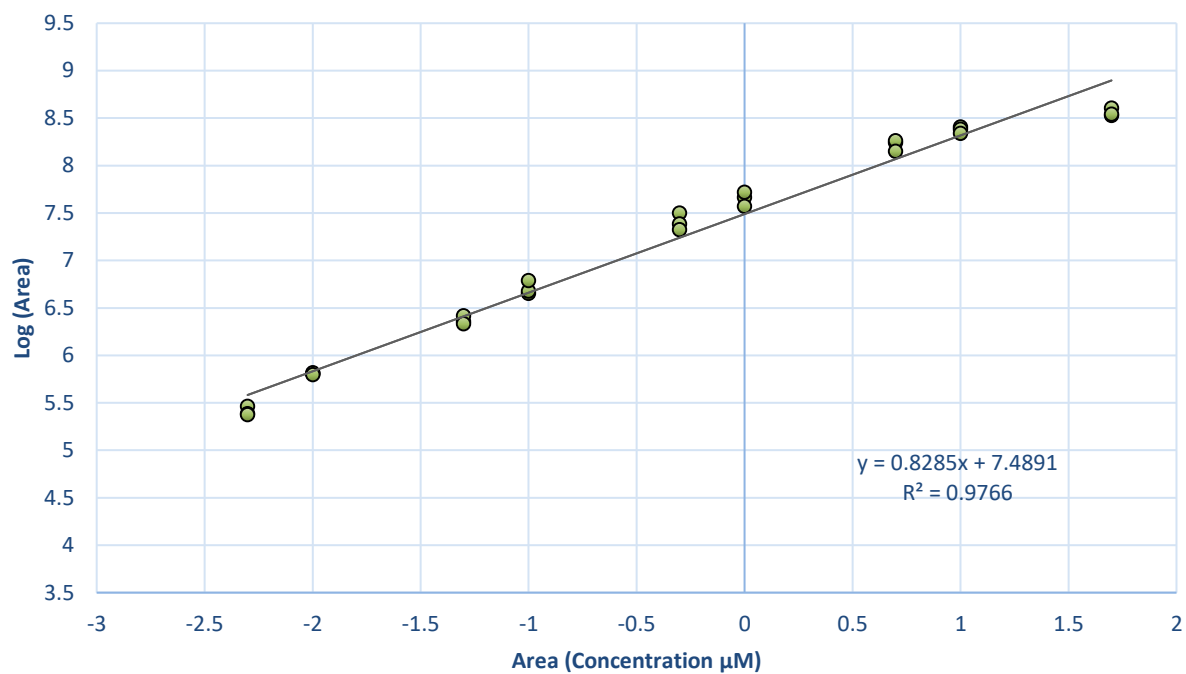
N-dodecanoyl valine



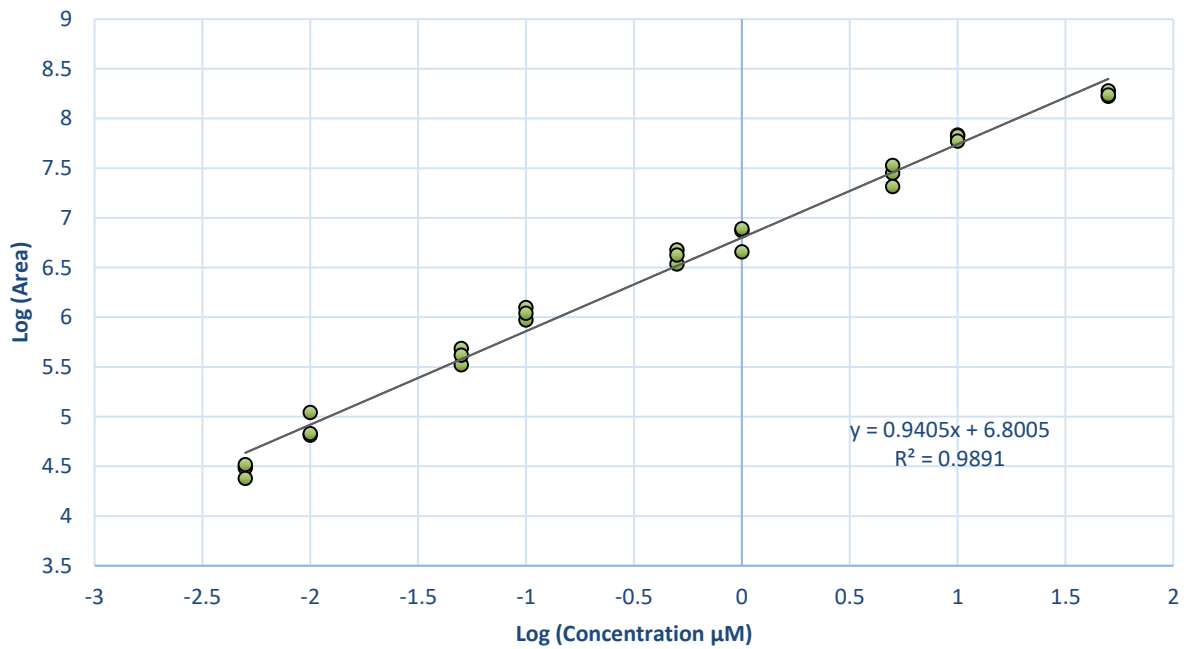
N-dodecanoyl phenylalanine



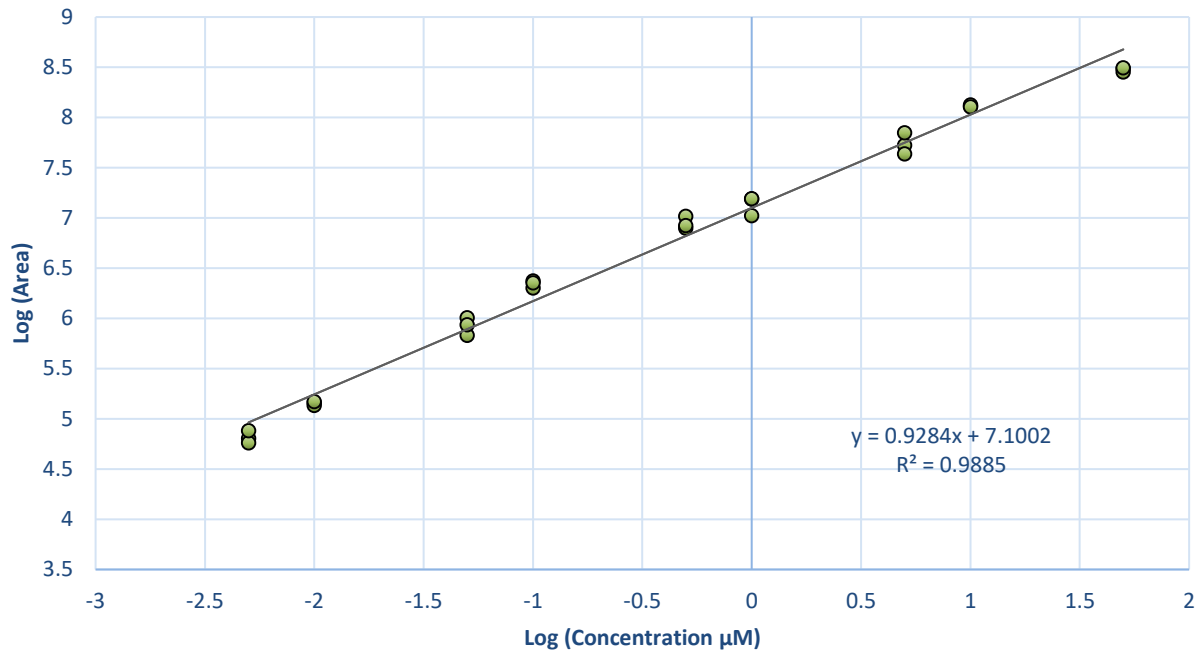
N-dodecanoyl leucine



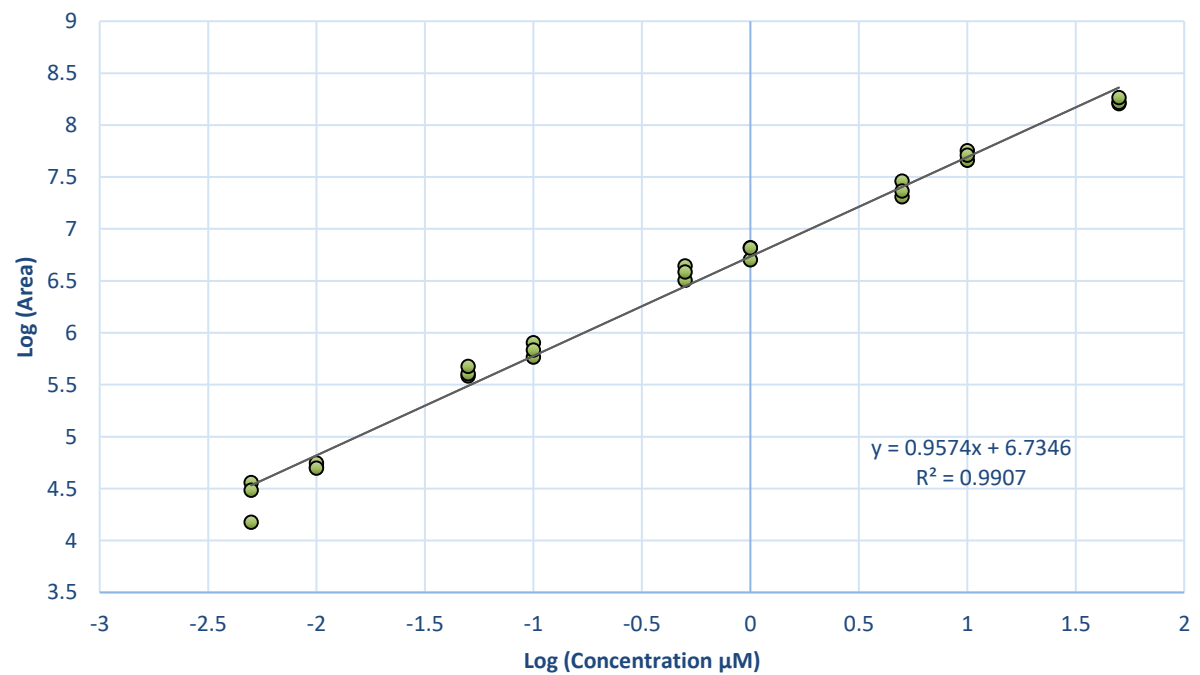
N-docosahexaenoyl phenylalanine



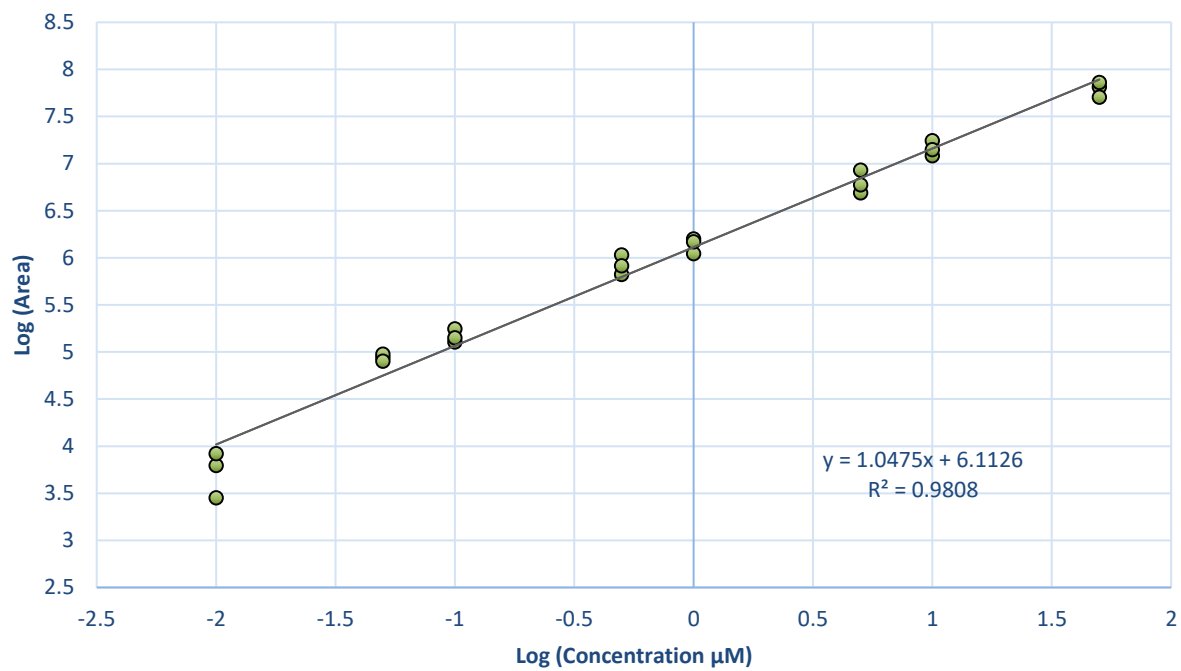
N-docosahexaenoyl leucine



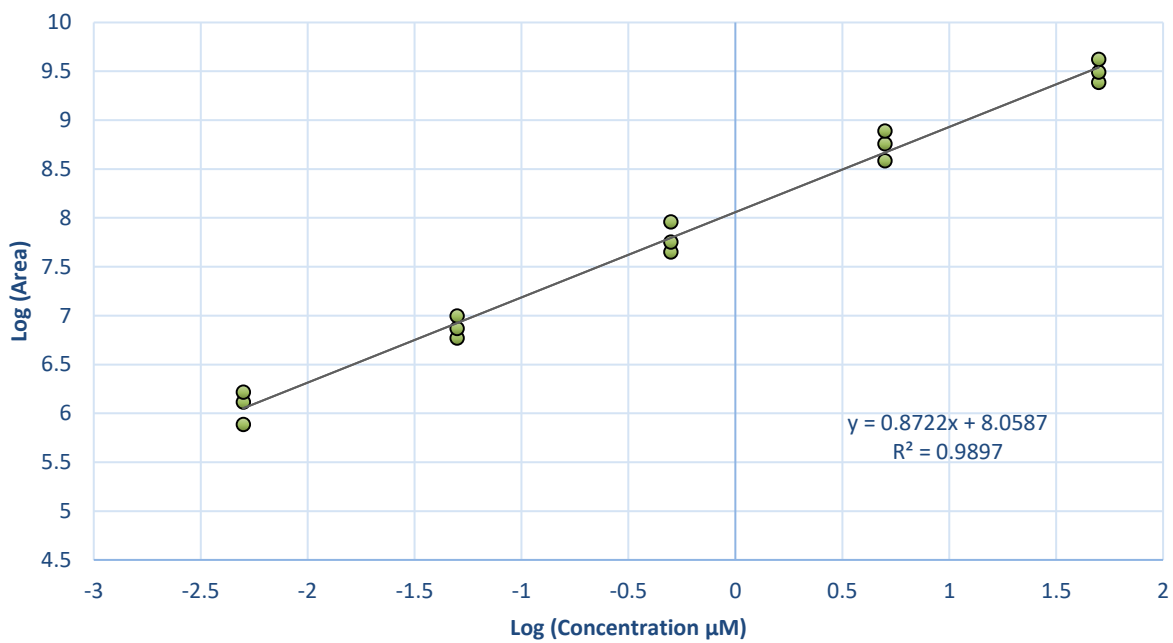
N-palmitoleoyl phenylalanine



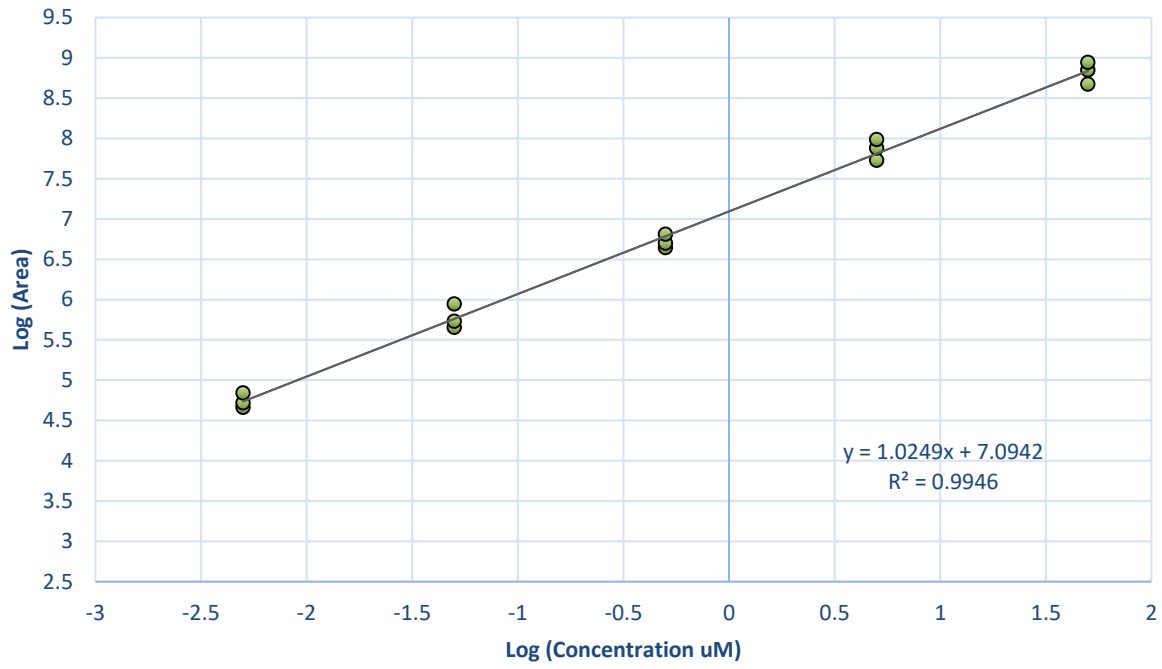
N-docosahexaenoyl valine



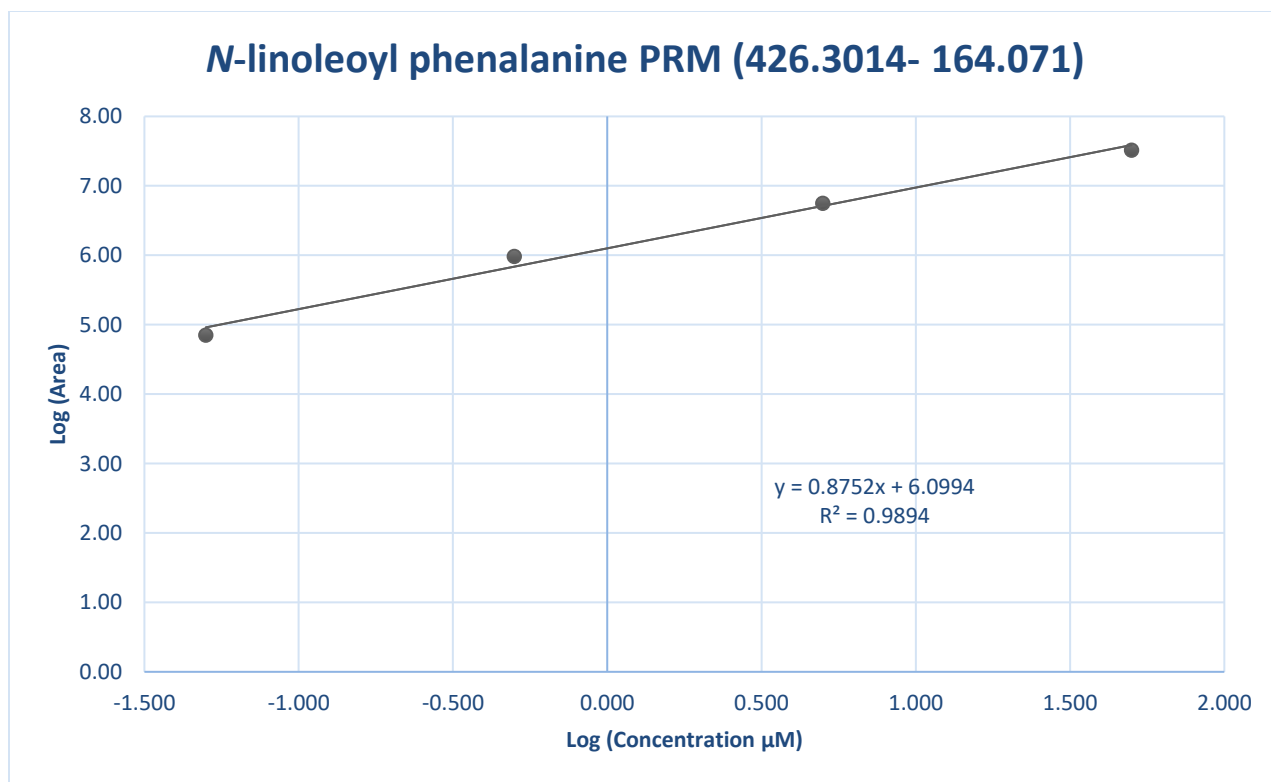
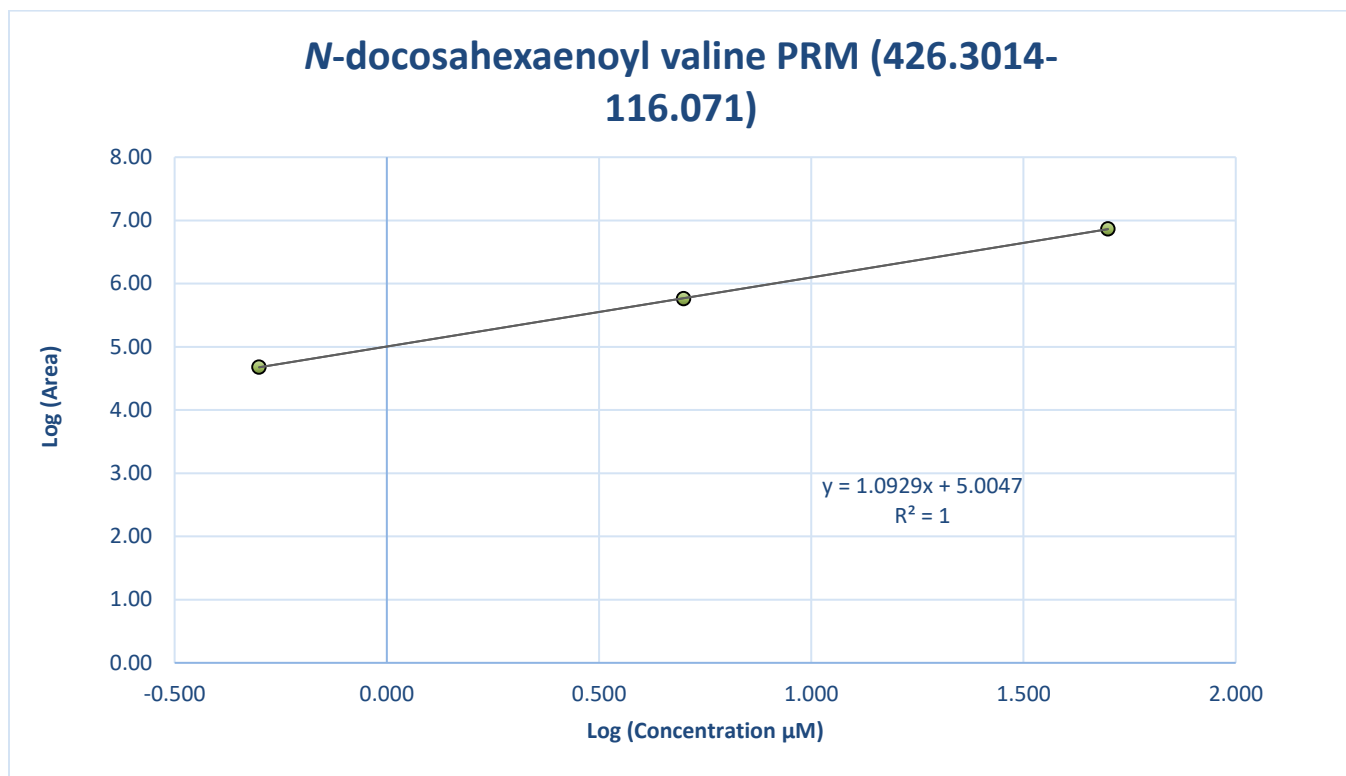
N-eicosanoyl leucine



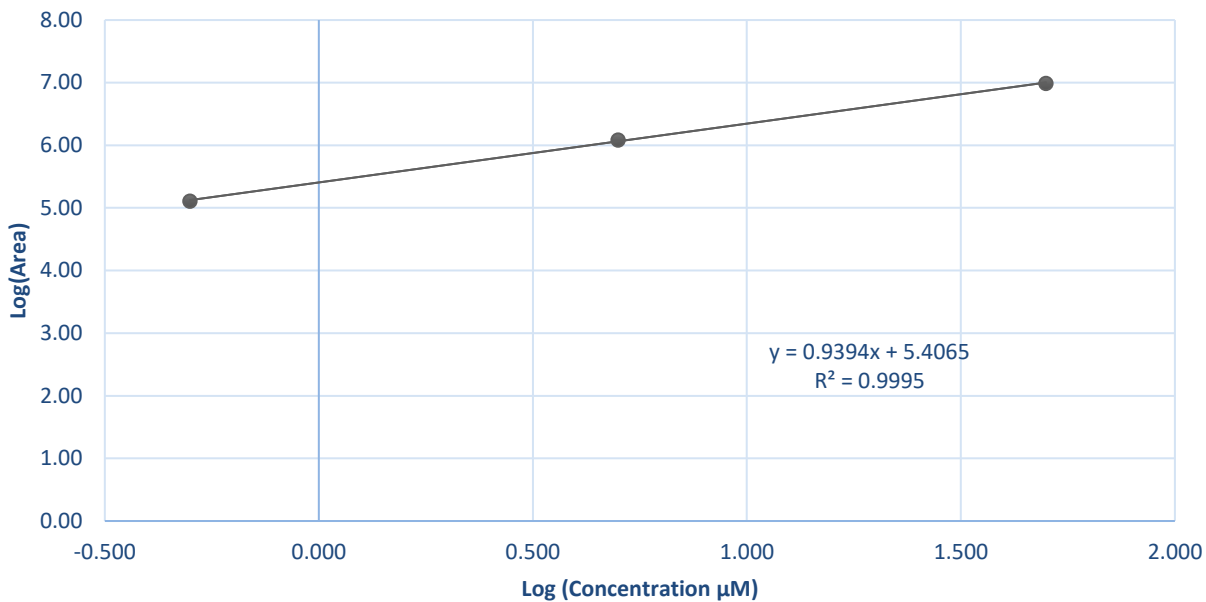
N-eicosanoyl valine



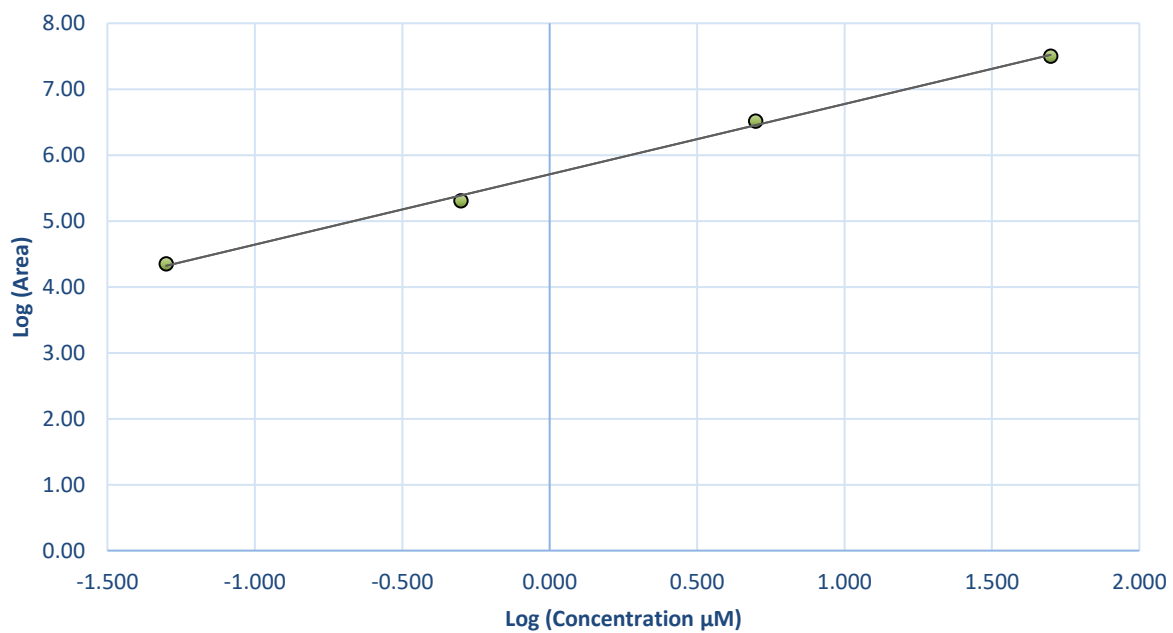
B3: MS/MS Fragmentation Calibration Curves



Palmitoleoyl Phenalanine PRM (400.2857- 164.071)



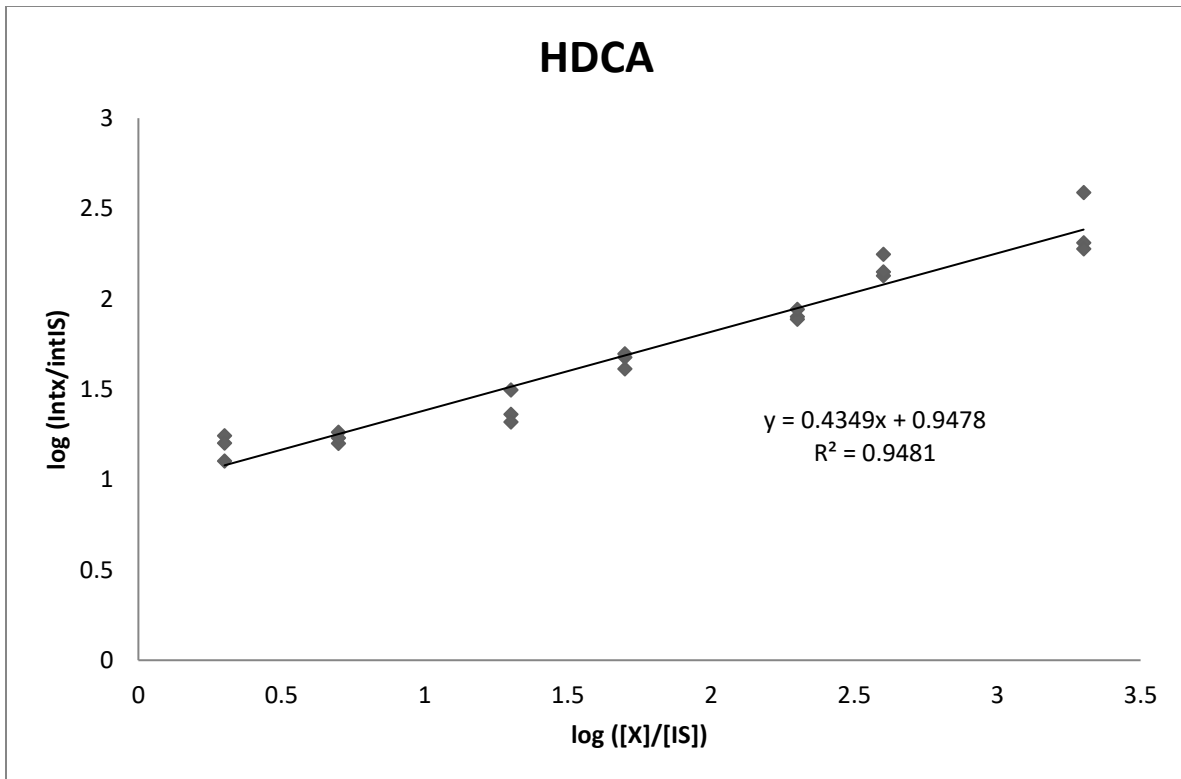
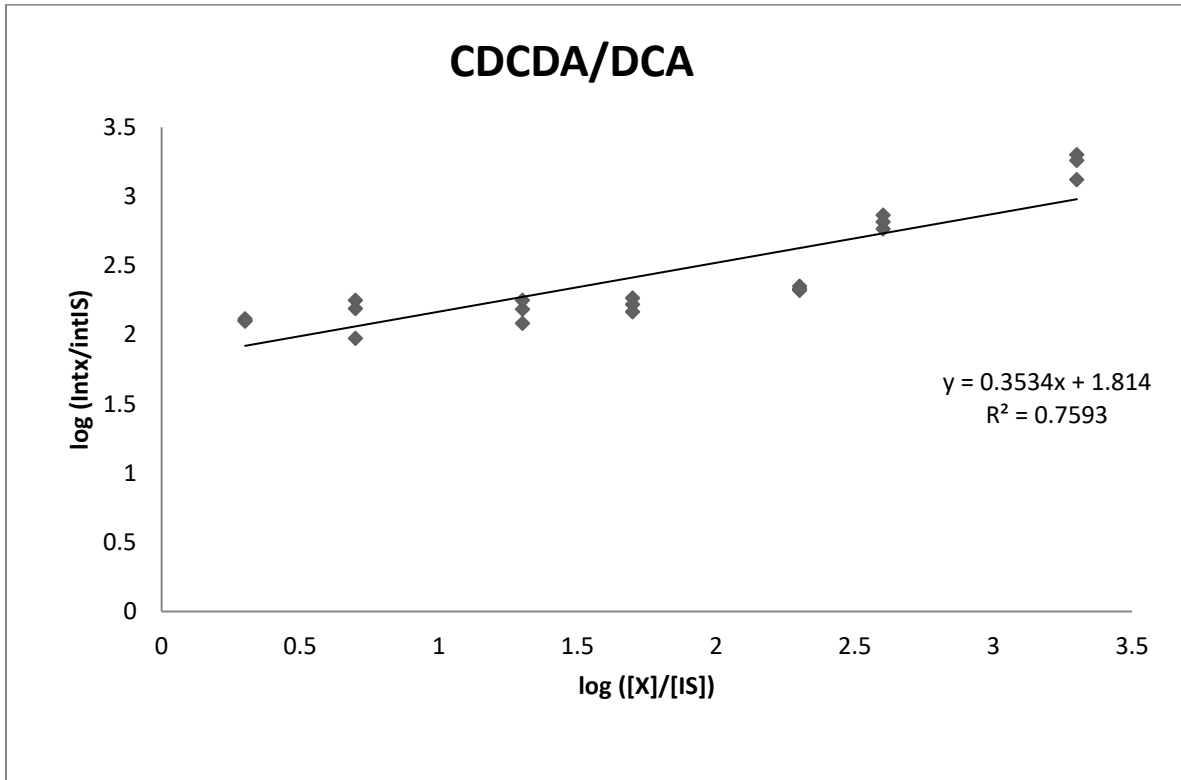
N-eicosapentaenoyl valine PRM (400.2857-116.071)

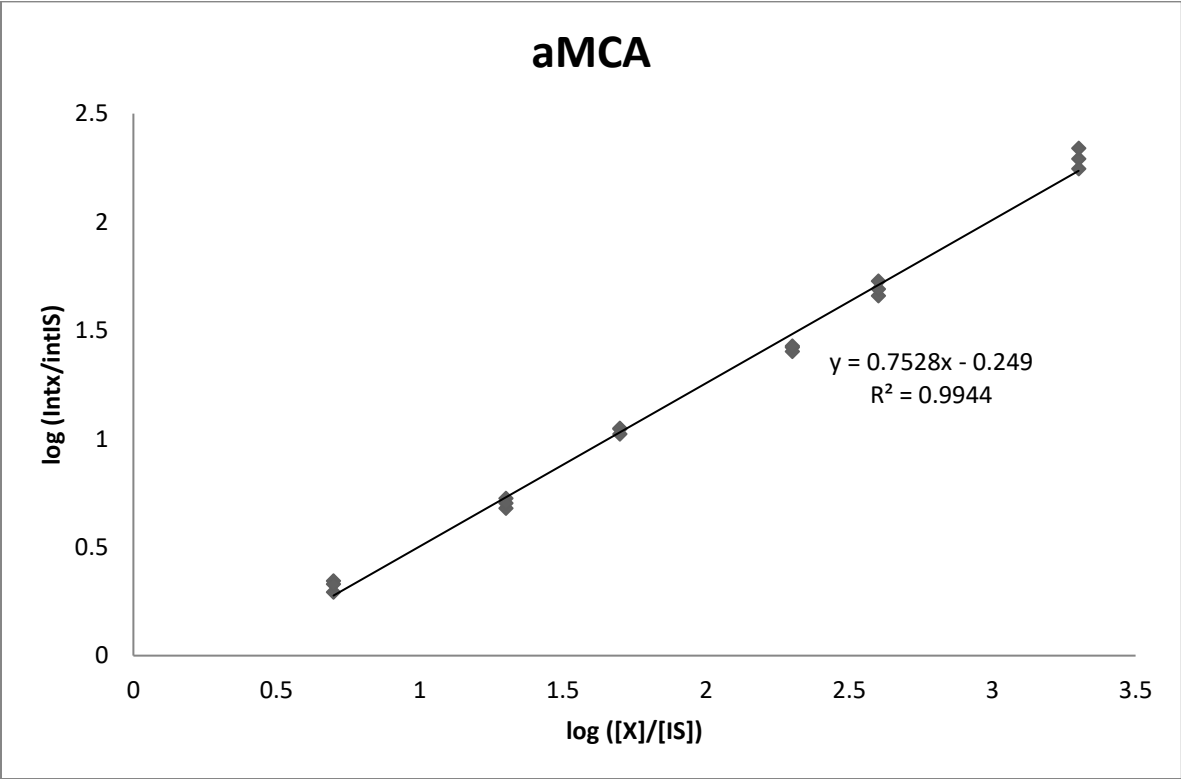
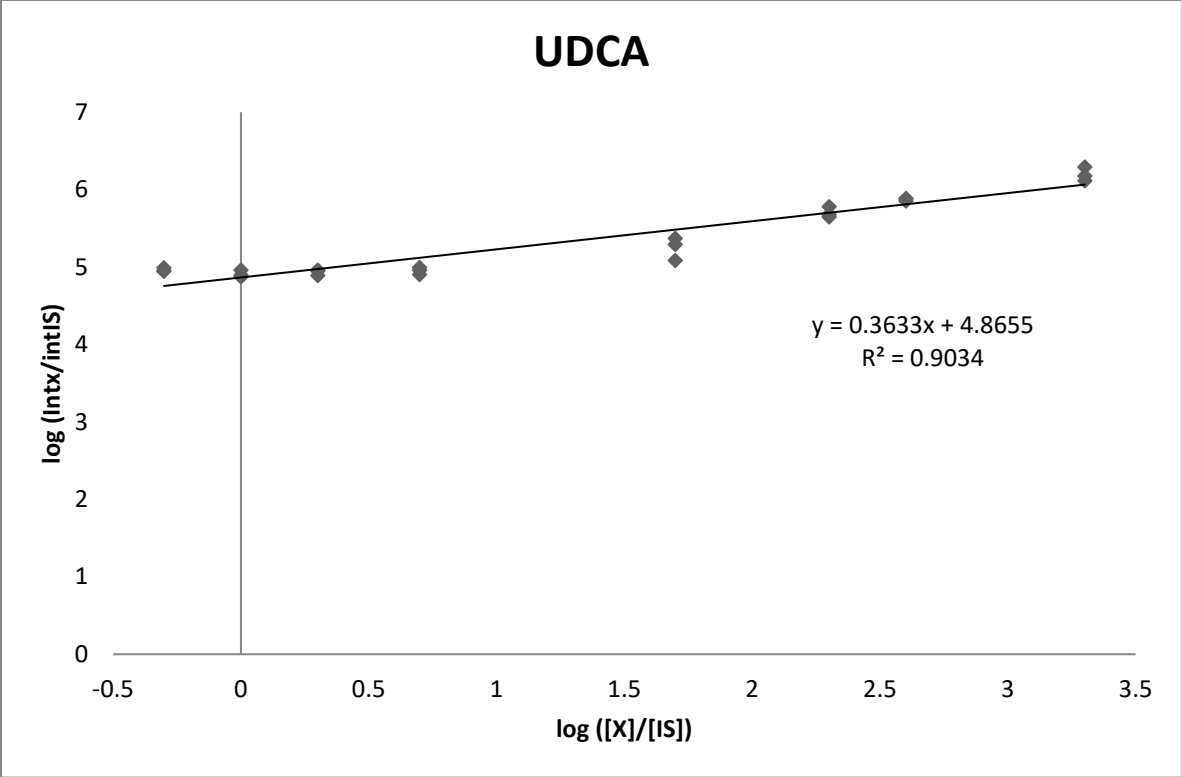


B4: Raw calibration data for N-acyl amino acids

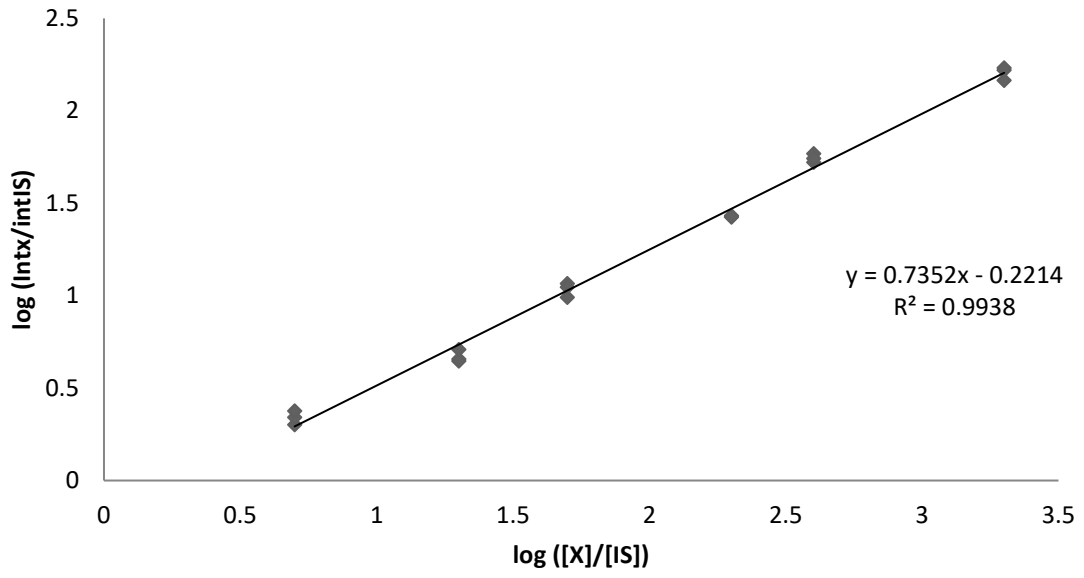
log (concentration µM)	stearoyl valine	stearoyl leucine	stearoyl phenylalanine	palmitoyl valine	palmitoyl leucine	palmitoyl phenylalanine	palmitoyl leucine	palmitoyl leucine	oleoyl valine	oleoyl phenylalanine	oleoyl leucine	linoleoyl valine	linoleoyl phenylalanine	linoleoyl leucine	eicosapentaenoyl valine	eicosapentaenoyl phenylalanine	eicosapentaenoyl leucine	icosanoyl phenylalanine	dodecanoyl valine	dodecanoyl phenylalanine	dodecanoyl leucine	docosahexaenoyl phenylalanine	docosahexaenoyl leucine	palmitoleoyl phenylalanine	docosahexaenoyl valine	eicosanoyl valine	eicosanoyl leucine
2.0000	4.6683	8.3895	8.2976	8.8337	8.8632	8.6429	8.7059	7.3937	8.7569	8.5393	8.5611	8.2531	8.6464	8.5679	8.6419	8.6279	8.4039	8.0660	9.0149	7.9436	8.6452	8.3649	8.5941	8.3664	8.0884		
2.0000	8.5302	8.3988	8.2612	8.8525	8.9104	8.6343	8.7197	7.4774	8.7807	8.5528	8.5469	8.3328	8.6544	8.6282	8.7095	8.6255	8.4320	8.0786	9.0808	7.9724	8.6576	8.4060	8.6137	8.4530	8.1317		
2.0000	8.5376	8.4195	8.2687	8.8258	8.8593	8.5936	8.7057	7.3593	8.7780	8.5468	8.5400	8.2360	8.6459	8.5708	8.6969	8.6212	8.4486	8.0906	9.0785	8.0636	8.6586	8.4356	8.6126	8.3843	8.0509		
1.6990	4.8483	8.3201	8.1816	8.7018	8.8056	8.5405	8.6844	7.1829	8.6053	8.4142	8.4010	8.0686	8.5403	8.3850	8.5589	8.4678	8.2969	7.8429	8.9084	7.6681	8.6043	8.2793	8.4840	8.2076	7.8135	8.6760	9.3855
1.6990	8.3856	8.2769	8.1550	8.6681	8.7549	8.5166	8.6569	7.0501	8.5512	8.4094	8.3762	7.9919	8.5469	8.3541	8.5074	8.5008	8.2432	7.7929	8.8582	7.6692	8.5301	8.2245	8.4515	8.2194	7.7067	8.8479	9.4923
1.6990	8.4564	8.3223	8.2199	8.6343	8.7453	8.6255	8.6442	6.9885	8.6385	8.4034	8.4601	7.9877	8.5725	8.4488	8.4405	8.4340	8.2845	7.9231	8.8243	7.7758	8.5420	8.2347	8.4937	8.2642	7.8600	8.9448	9.6228
1.0000	8.0531	7.9442	7.9663	8.3683	8.5393	8.1907	8.3911	6.6967	8.1714	8.1661	8.0544	7.4716	8.2435	7.8554	8.1048	8.1901	7.8524	7.4919	8.5875	7.1433	8.4951	7.8342	8.1129	7.6906	7.0919		
1.0000	8.1461	7.9785	7.8946	8.3317	8.4656	8.3202	8.4235	6.6022	8.3077	8.1197	8.1154	7.5144	8.2138	7.9715	8.0369	8.1257	7.8369	7.5641	8.5497	7.1950	8.3827	7.8243	8.1255	7.7522	7.2430		
1.0000	8.1271	7.9831	7.8953	8.2404	8.4793	8.2831	8.4016	6.5691	8.2109	8.0771	8.0664	7.4094	8.1809	7.9177	7.9799	8.0328	7.8012	7.5443	8.4655	7.0871	8.3395	7.7729	8.1073	7.7279	7.1489		
0.6990	7.8476	7.7520	7.6999	8.0252	8.2513	8.0778	8.2032	6.4349	7.9810	7.8456	7.8357	7.1598	7.8977	7.6718	7.7643	7.8145	7.4627	7.1962	8.2862	6.7546	8.2405	7.4505	7.7219	7.4607	6.9297	7.7280	8.5837
0.6990	7.8195	7.7523	7.5796	8.1336	8.2651	7.9574	8.2595	6.3299	7.9030	7.7733	7.7358	7.2444	7.8003	7.5009	7.8453	7.6949	7.3436	7.1143	8.3736	6.5691	8.2621	7.3153	7.6360	7.3094	6.6992	7.8020	8.7592
0.6990	7.8710	7.8063	7.7084	7.9486	8.1520	8.0025	8.1678	6.2123	7.9191	7.9387	7.8313	7.0778	7.9913	7.5705	7.6220	7.8799	7.5961	7.3520	8.2150	6.8375	8.1490	7.5293	7.8465	7.3666	6.7445	7.9858	8.8880
0.0000	7.3118	7.3162	7.1885	7.2679	7.5774	7.5250	7.5976	5.4778	7.3544	7.2853	7.2226	6.4139	7.2413	6.9364	6.9857	7.0647	6.6740	6.5842	7.5630	5.8899	7.6654	6.6595	7.0240	6.7014	6.0456		
0.0000	7.4472	7.3833	7.3262	7.1576	7.5127	7.6206	7.5997	5.4051	7.5110	7.3466	7.2790	6.2977	7.1369	7.0171	6.8821	7.2030	6.8201	6.7750	7.5120	6.0926	7.5718	6.8710	7.1898	6.8200	6.2010		
0.0000	7.3365	7.3164	7.3488	7.3660	7.6689	7.6120	7.7382	5.5648	7.4801	7.4392	7.2520	6.4728	7.4017	7.0764	7.0412	7.2156	6.9325	6.7924	7.6593	6.1309	7.1186	6.8894	7.1921	6.8164	6.1703		
-0.3010	7.1673	7.0699	6.9743	7.0389	7.4301	7.4211	7.5280	5.2032	7.2127	7.1592	7.0420	6.1871	7.1152	6.8397	6.7248	6.8802	6.5337	6.3728	7.4148	5.7223	6.8989	6.5351	6.8989	6.5057	5.8227	6.6446	7.6517
-0.3010	7.3010	7.2410	7.0109	6.9583	7.3188	7.5436	7.4495	5.2382	7.4231	7.2969	7.1181	6.0316	7.2635	6.9995	6.6540	6.8078	6.7065	6.5368	7.2876	5.8472	7.3869	6.6765	7.0149	6.6439	6.0289	6.7012	7.7517
-0.3010	7.2529	7.1218	7.0092	7.0393	7.3542	7.4610	7.5365	5.4007	7.2867	7.0654	6.1481	7.1460	6.8027	6.6917	6.9537	6.5836	6.4820	7.3524	5.8717	7.3247	6.6255	6.9245	6.5863	5.9134	6.8137	7.9576	
-1.0000	6.4594	6.4082	6.4814	6.2651	6.7366	6.6813	7.1214	4.3708	6.5736	6.6804	6.2795	5.2104	6.7354	5.9647	6.0811	6.3835	6.0543	5.8048	6.5767	5.2177	6.6527	6.0978	6.3720	5.7667	5.1075		
-1.0000	6.5762	6.5080	6.3390	6.2829	6.7302	6.7713	7.1434	4.7005	6.5930	6.6095	6.4235	5.4645	6.5998	6.2387	5.9354	6.2830	5.9593	5.6517	6.6329	5.2439	6.6771	5.9731	6.3039	5.9050	5.2439		
-1.0000	6.4706	6.5311	6.3361	6.3870	6.8306	6.7815	7.2269	4.6131	6.6716	6.5426	6.3030	5.4935	6.6339	6.0437	6.0301	6.3123	6.0133	5.7301	6.7132	5.2274	6.7893	6.0416	6.3547	5.8347	5.1521		
-1.3010	6.2465	6.1688	5.9461	6.0521	6.4008	6.5560	7.0492	4.4099	6.3757	6.1783	6.0574	5.1544	6.2896	5.8452	5.6561	5.9523	5.5214	5.2927	6.2707	4.8666	6.3573	5.6846	6.0077	5.5844	4.9385	5.6599	6.7698
-1.3010	6.2842	6.1224	5.8945	5.9875	6.4408	6.5448	7.0117	4.2673	6.4019	6.0842	6.0675	5.0744	6.1461	5.8959	5.6721	5.9785	5.4355	5.2214	6.3211	4.8472	6.4163	5.9222	5.8318	5.6053	4.9784	5.7326	6.8679
-1.3010	6.2152	6.0688	5.9244	6.0055	6.4207	6.4876	7.0684	4.3589	6.4084	6.1392	6.0325	5.0772	6.1826	5.8883	5.6986	5.8537	5.5892	5.2070	6.2744	4.8632	6.3323	5.6171	6.0283	5.6771	4.9023	5.9471	6.9948
-2.0000	5.5371	5.2849	5.4150	5.3707	5.7256	5.6087	7.0607	3.6017	5.2926	5.5282	5.2683	4.4601	5.5239	4.9160	4.8997	5.2329	4.9075	4.7932	5.6723	4.6407	5.8061	5.0423	5.1574	4.7052	3.7956		
-2.0000	5.5027	5.4830	5.3534	5.3151	5.8044	5.7289	7.0661	3.9727	5.4710	5.5995	5.3887	4.4370	5.3990	5.0230	5.0754	5.1478	4.5937	5.6787	4.5343	4.5443	5.8166	4.8119	5.1346	4.7464	3.9201		
-2.0000	5.4084	5.3372	5.3622	5.3174	5.6579	5.7053	7.0001	3.6617	5.4652	5.5760	5.2466	4.5456	5.4360	4.8576	5.0169	5.1700	4.7710	4.6614	5.6503	4.5460	5.7971	4.8308	5.1700	4.6962	3.4515		
-2.3010	5.0100	5.1707	4.9980	4.9924	5.3842	5.3602	6.9461	3.5042	5.2733	5.2598	4.8478	3.9278	5.1908	4.4601	4.6841	4.8343	4.4988	4.1991	5.2793	4.2983	5.4633	4.4851	4.8063	4.5592	#NUM!	4.6621	5.8855
-2.3010	4.8984	5.0350	4.9268	4.9188	5.2275	5.3554	6.9167	3.9527	5.1105	5.1073	4.8498	3.8979	5.1200	4.7160	4.5295	4.4982	4.2123	5.2485	4.2390	4.3860	4.3787	4.7615	4.8812	4.4845	#NUM!	4.7211	6.1187
-2.3010	5.0099	5.0722	5.0367	4.9456	5.3254	5.3891	6.9243	3.5772	5.0722	5.1275	4.8927	4.0752	5.0833	4.3802	4.4334	4.4842	4.2587	5.2680	4.3783	5.3760	4.5181	4.8812	4.4845	#NUM!	4.8416	6.2180	
-3.0000	4.3712	4.4479	4.1948	4.7515	4.8572	4.8572	4.9543	4.4554	4.5452	4.3834	4.3508	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	3.1101	4.4563	4.0238	4.7147	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	
-3.0000	4.4919	4.4953	4.0325	3.8678	4.7177	4.8024	6.8155	3.6236	4.5994	3.9070	4.0930		4.0779	3.8611				3.1616	4.5278	4.0013	4.7520	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	
-3.0000	3.8692		4.2101	3.3965	4.7957	4.7453	6.6721			4.6286			4.2717	#NUM!	3.7833				4.6327	3.9295	4.8904	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	

B5: Calibration plots for bile acid standards

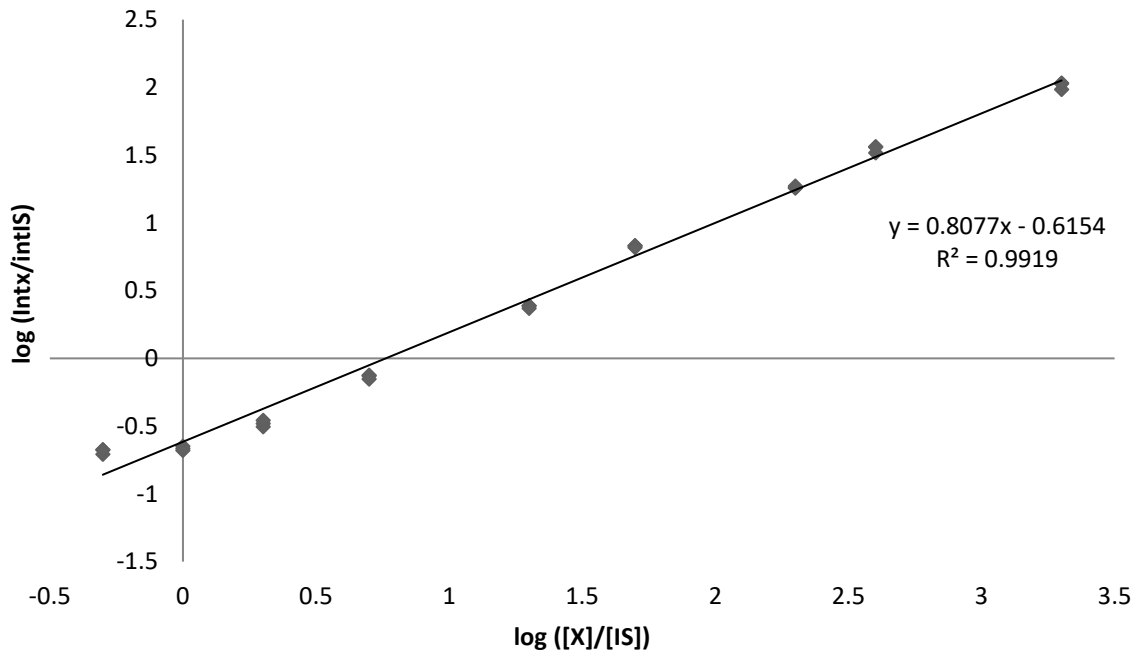


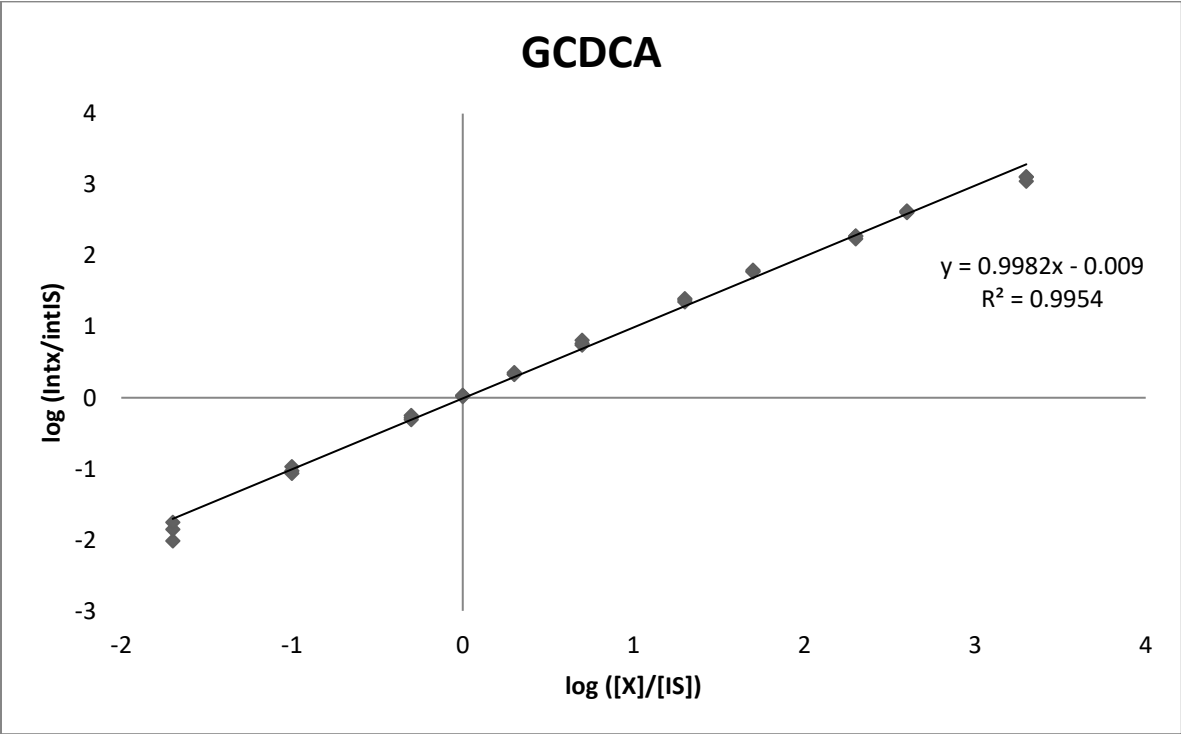
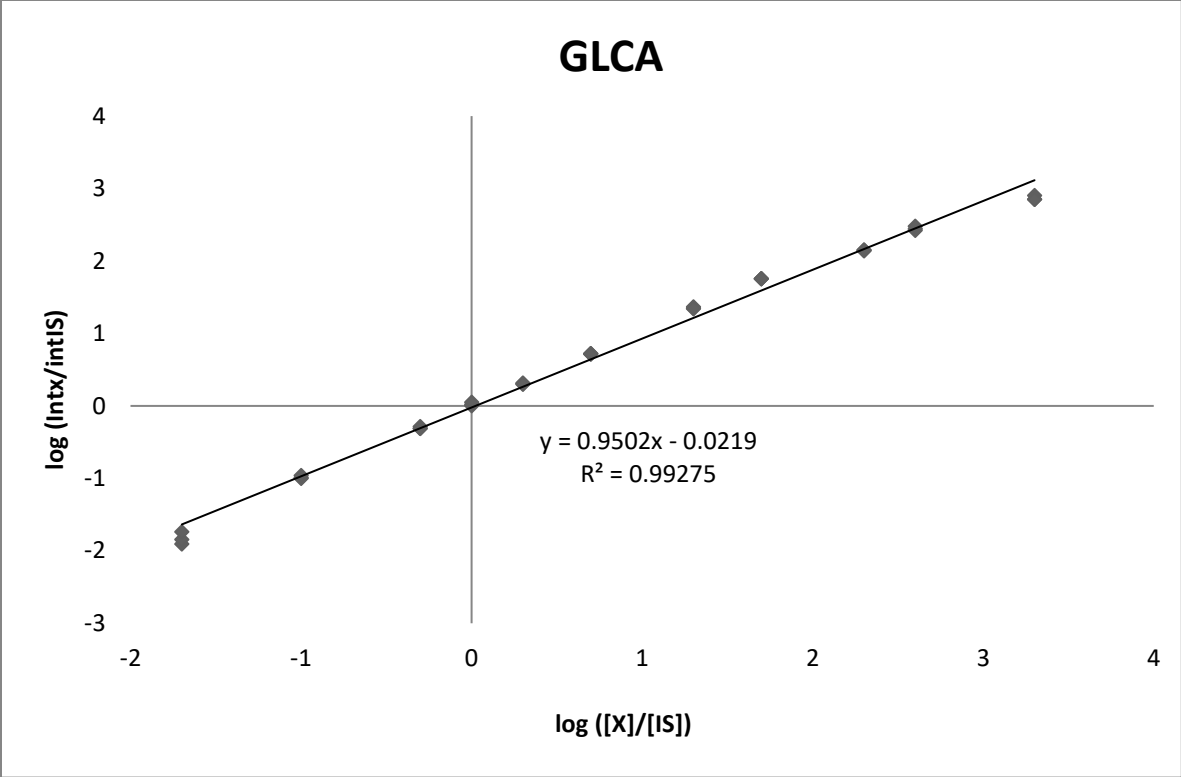


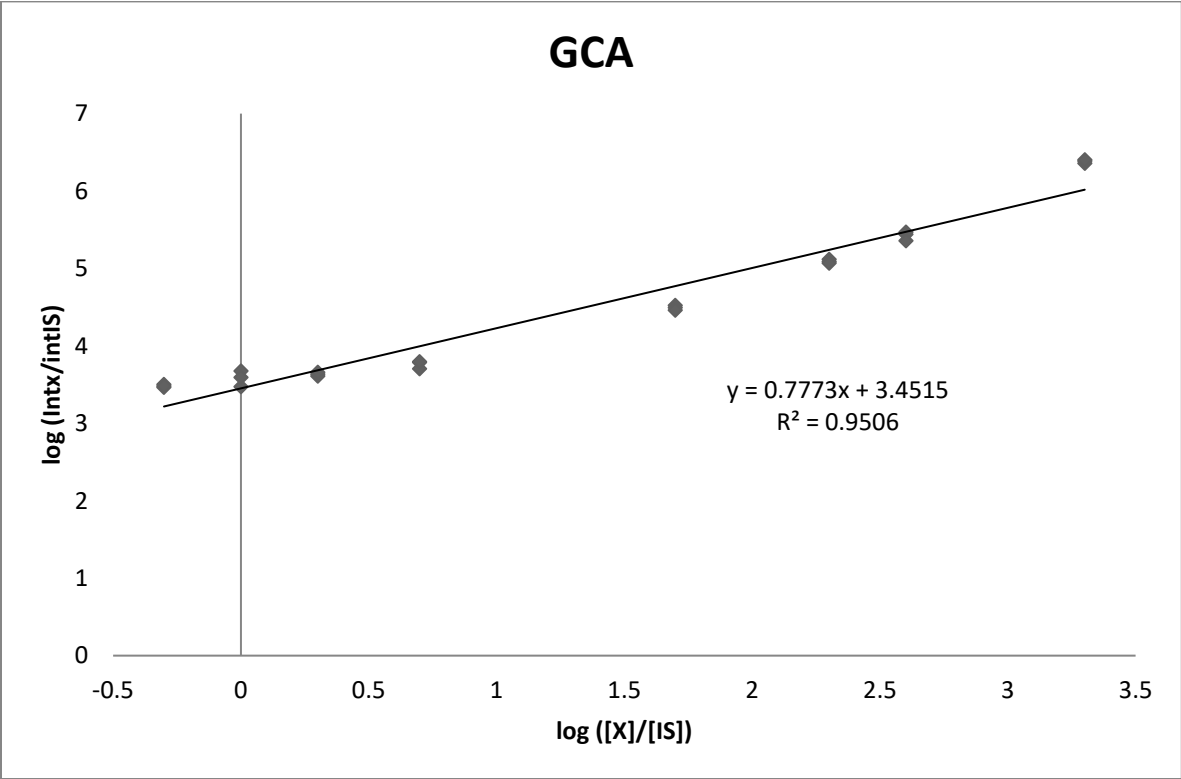
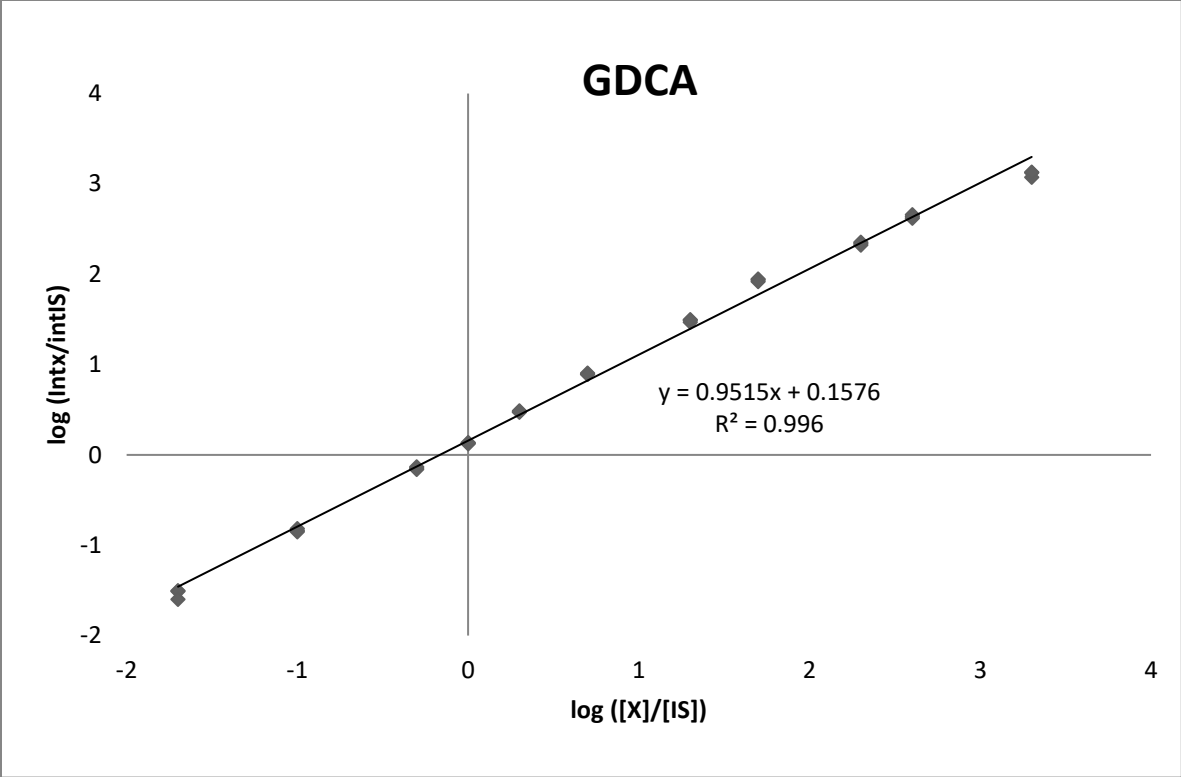
bMCA

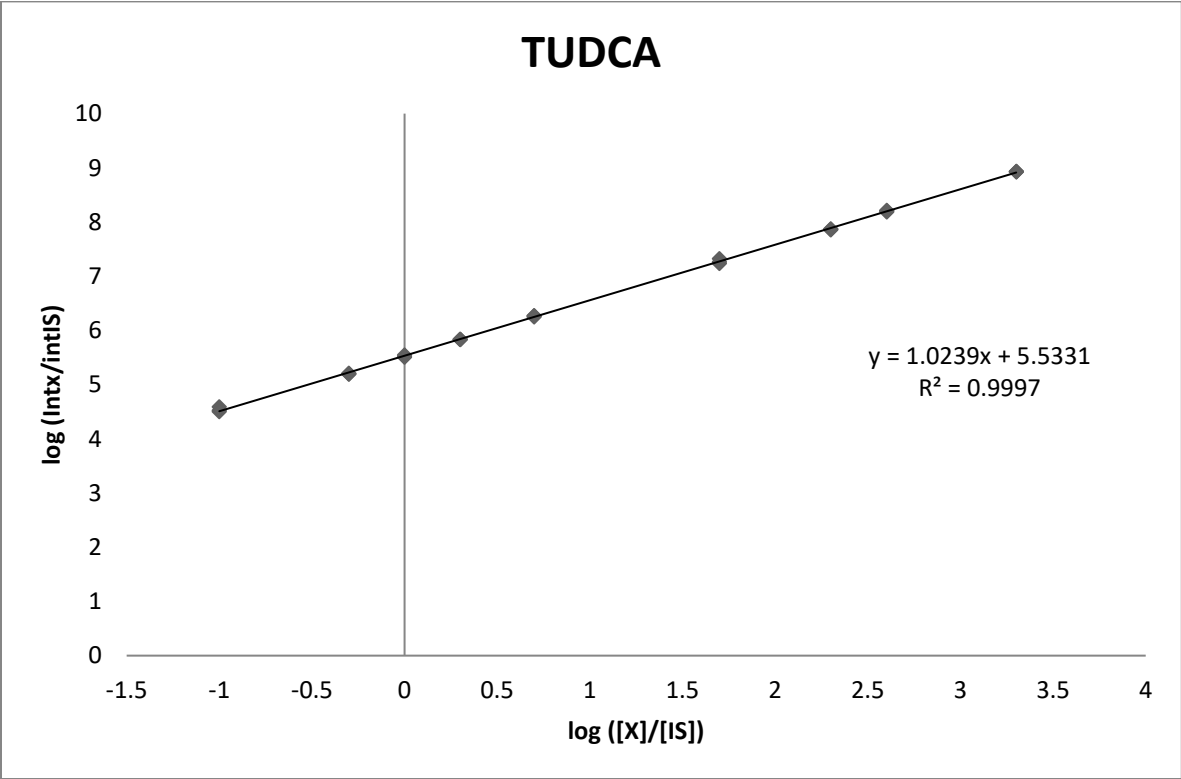
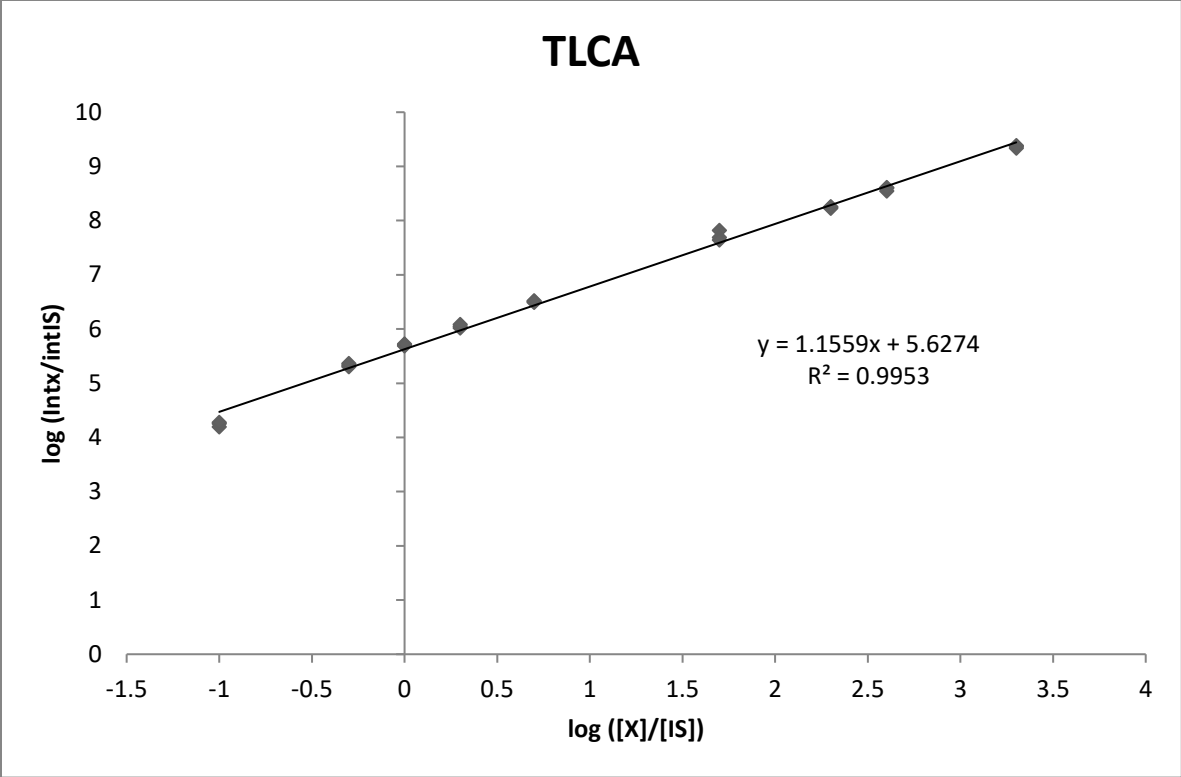


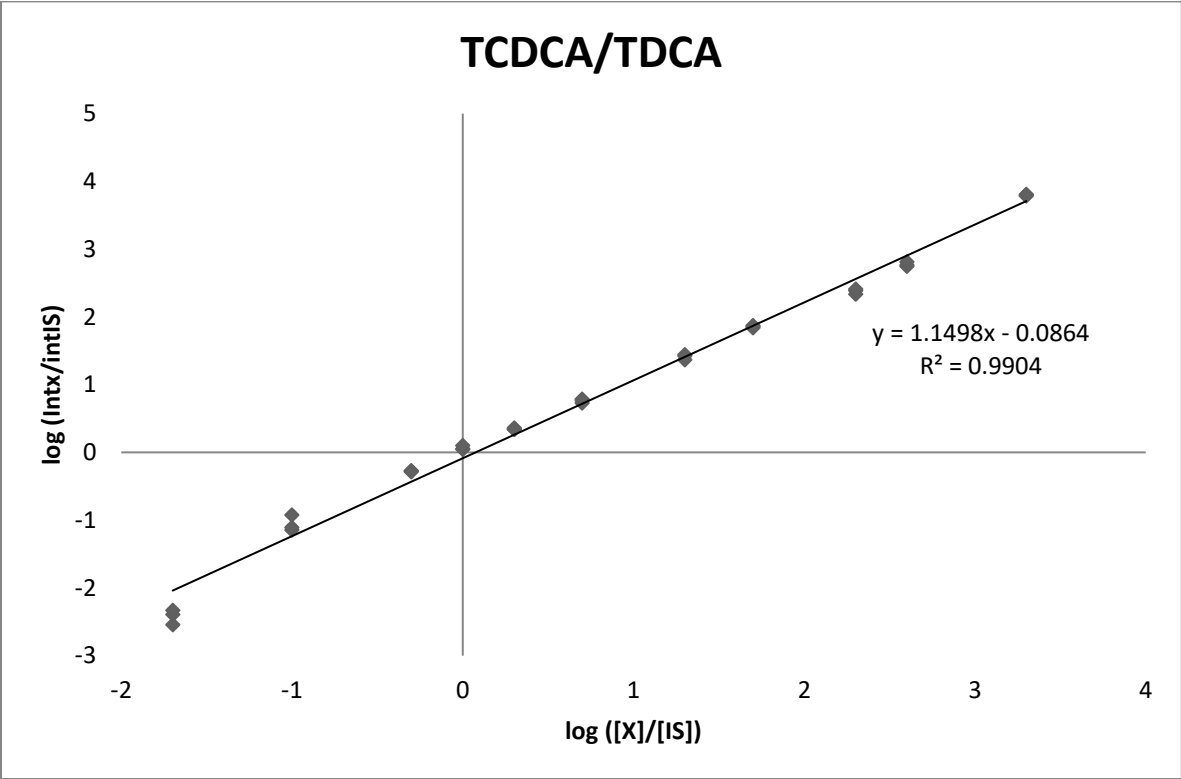
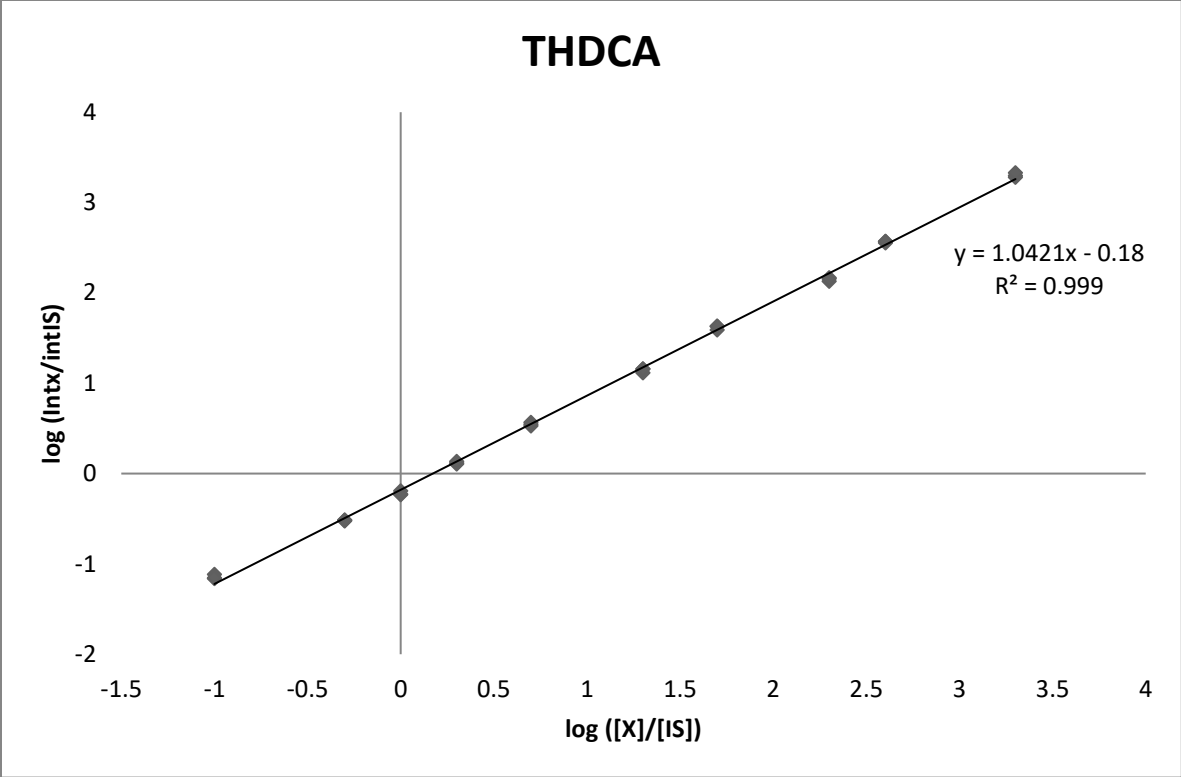
CA

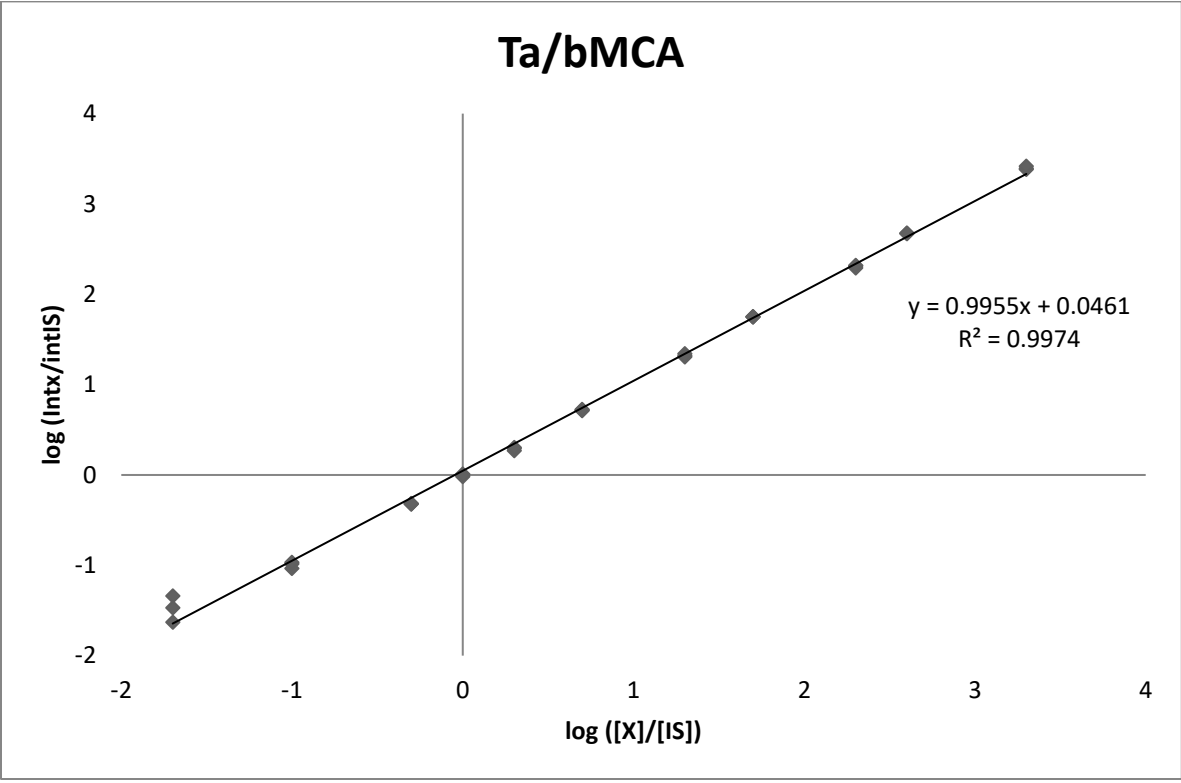
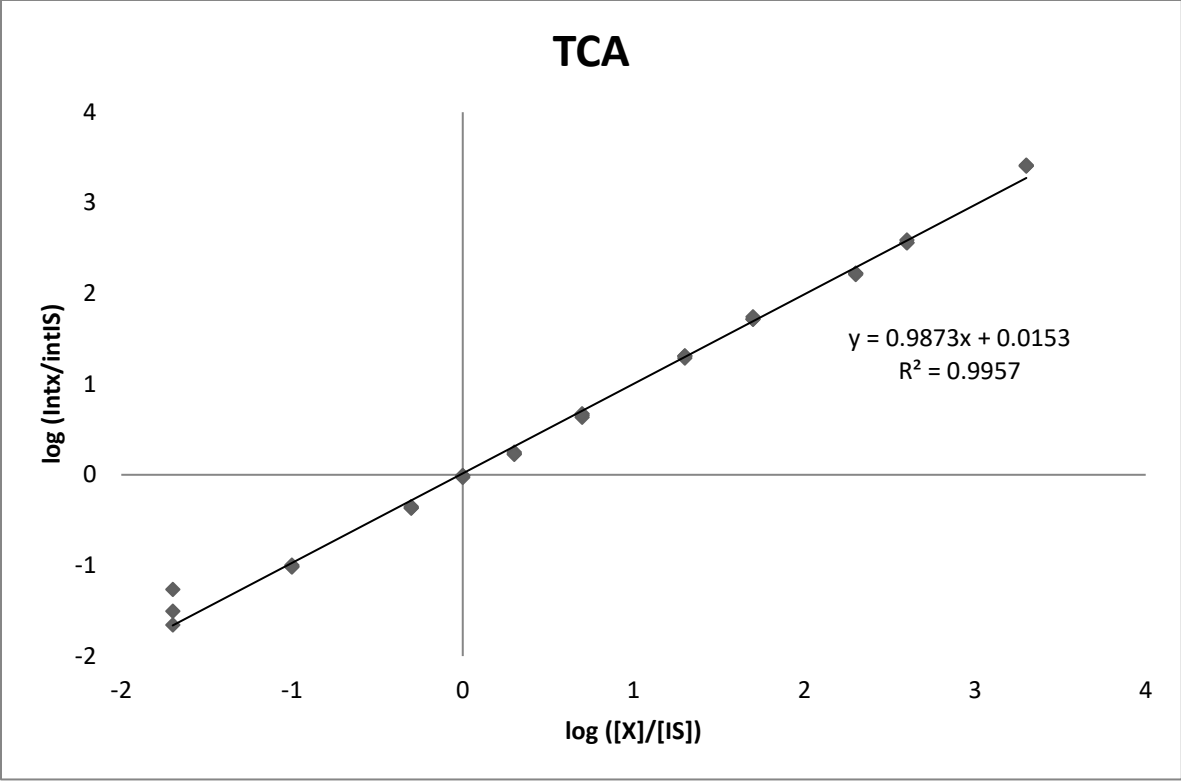




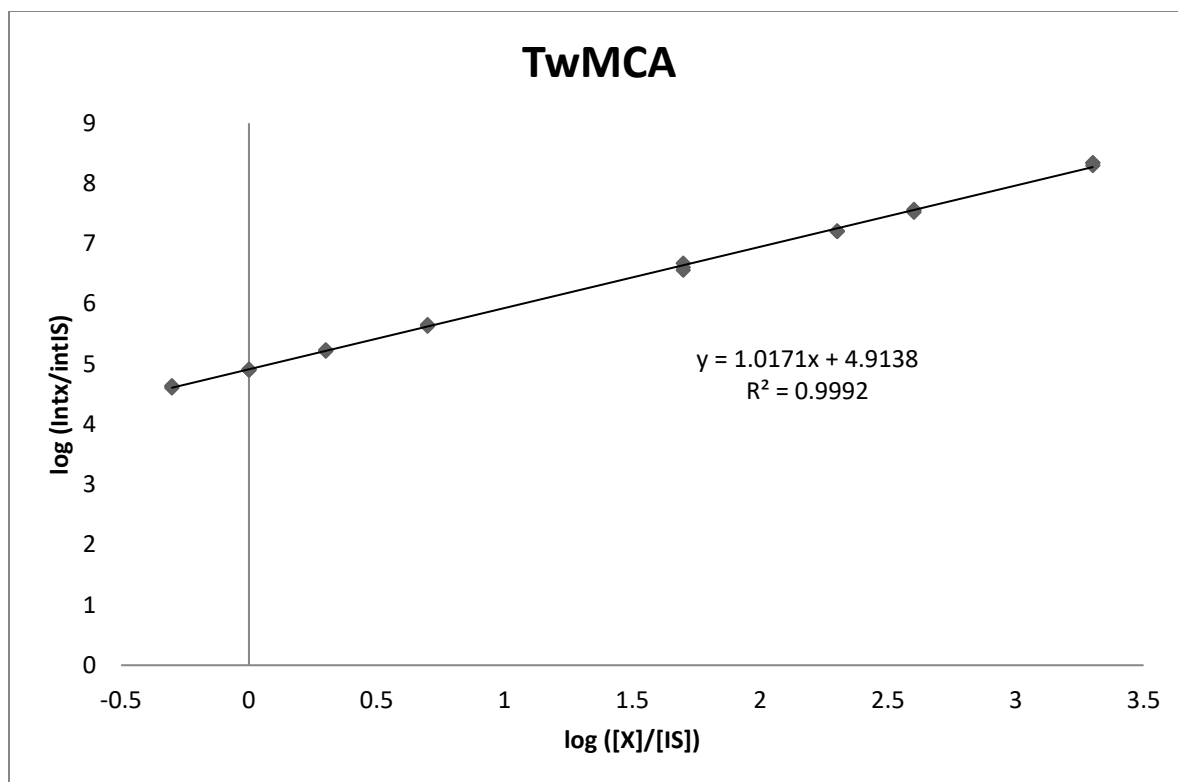








TwMCA



B7: Raw calibration data for bile acids

log (Conc/	CDCA/DCA	HDCA	UDCA	aMCA	bMCA	CA	GLCA	GCDCA	GDCA	GCA	TLCA	TUDCA	THDCA	TCDCa/TD	TCA	Ta/bMCA	TwMCA
-1.69897	2.061549	1.291834	4.936562	0.557784	0.44174	-0.63837	-1.90671	-2.01447	-1.50285	3.53364	3.215055	6.089721	0.387581	-2.54432	-1.26514	-0.21055	5.177999
-1.69897	1.934273	1.192529	5.050733	0.464156	0.478596	-0.64295	-1.84404	-1.85193	-1.5103	3.554461	2.487105	5.255419	-0.37261	-2.39106	-1.50504	-0.96956	4.525954
-1.69897	2.285369	1.311113	5.054098	0.453399	0.411575	-0.68337	-1.73597	-1.75337	-1.59839	3.4661	3.246484	4.585176	-1.03536	-2.33081	-1.6575	-1.15481	4.054612
-1	2.139919	1.260654	4.97654	0.447699	0.408374	-0.65725	-0.99022	-1.06475	-0.83844	3.426025	4.196224	4.587472	-1.11613	-1.14542	-1.01895	-1.02584	4.103345
-1	2.136675	1.348237	5.077999	0.366911	0.342868	-0.67755	-0.99872	-1.02699	-0.81513	3.475145	4.255996	4.500273	-1.16386	-1.10686	-0.99664	-1.00639	3.923778
-1	2.13634	1.342496	5.047725	0.493127	0.395866	-0.62696	-0.95915	-0.96881	-0.85275	3.402518	4.274449	4.528088	-1.15646	-0.92576	-1.01062	-1.04785	4.087695
-0.30103	2.042518	1.279993	4.947524	0.360121	0.355176	-0.67326	-0.31237	-0.30997	-0.16178	3.463542	5.302292	5.197278	-0.52626	-0.27156	-0.36704	-0.62428	4.602338
-0.30103	2.110533	1.117402	4.988151	0.340047	0.337218	-0.70732	-0.28532	-0.24893	-0.14642	3.501813	5.336804	5.190742	-0.51463	-0.27182	-0.36333	-0.60343	4.638714
-0.30103	2.13543	1.244465	4.949313	0.38261	0.33868	-0.67418	-0.31352	-0.28299	-0.13176	3.480111	5.362707	5.208728	-0.51711	-0.28486	-0.34987	-0.6074	4.624786
0	2.131737	1.353962	4.959916	0.276967	0.277103	-0.66488	0.021573	0.032635	0.128177	3.474185	5.686838	5.505079	-0.23522	0.051319	-0.03065	-0.34075	4.91738
0	2.087328	1.189093	4.894921	0.310986	0.346211	-0.64658	0.000705	0.03134	0.121568	3.677419	5.703423	5.524322	-0.19237	0.043465	-0.00854	-0.32931	4.903131
0	2.061806	1.183272	4.874971	0.260664	0.264302	-0.67824	0.04904	0.017522	0.135058	3.593949	5.723721	5.544178	-0.22667	0.100916	-0.01179	-0.32757	4.903316
0.30103	2.101633	1.103246	4.891796	0.238356	0.277716	-0.50393	0.314118	0.327314	0.477105	3.634097	6.018661	5.834332	0.104604	0.336439	0.225349	-0.04248	5.21263
0.30103	2.115152	1.242678	4.949001	0.25853	0.250795	-0.47936	0.306791	0.351731	0.473531	3.610388	6.048766	5.837545	0.117413	0.35984	0.248697	-0.01497	5.230111
0.30103	2.099049	1.201706	4.959151	0.224052	0.265198	-0.45658	0.296437	0.338196	0.488345	3.660664	6.082633	5.827793	0.135205	0.345169	0.230831	-0.03535	5.2387
0.69897	1.973486	1.199988	4.963023	0.293514	0.301404	-0.12845	0.723296	0.807719	0.904135	3.7886	6.492577	6.248372	0.52887	0.730159	0.634922	0.339955	5.62705
0.69897	2.18937	1.229179	4.993761	0.34462	0.34157	-0.12215	0.720468	0.739901	0.894193	3.706718	6.509929	6.268306	0.567198	0.783449	0.655849	0.359375	5.64767
0.69897	2.249319	1.261812	4.903756	0.329944	0.377103	-0.15172	0.712065	0.765249	0.890711	3.797861	6.519846	6.271982	0.544317	0.766842	0.672346	0.370009	5.653961
1.30103	2.184107	1.496463		0.680031	0.657059	0.389174	1.349483	1.3769	1.468179				1.117188	1.368283	1.287538	0.948162	
1.30103	2.250375	1.361118		0.725994	0.708941	0.389539	1.368836	1.39626	1.489709				1.156322	1.442045	1.308211	0.968263	
1.30103	2.082861	1.319449		0.702676	0.646884	0.370791	1.341555	1.349806	1.49849				1.158959	1.421628	1.313905	0.969932	
1.69897	2.265211	1.611716	5.08424	1.048165	0.991443	0.83145	1.756037	1.795311	1.945096	4.522144	7.821931	7.320776	1.590848	1.8379	1.715736	1.429013	6.677097
1.69897	2.220159	1.675844	5.368491	1.046407	1.064684	0.832127	1.755291	1.773257	1.937405	4.494728	7.691655	7.266153	1.631912	1.871604	1.719076	1.427255	6.608406
1.69897	2.165547	1.695461	5.293018	1.022788	1.046615	0.820109	1.755111	1.779075	1.917807	4.460812	7.641858	7.235828	1.629221	1.859117	1.741862	1.433964	6.561709
2.30103	2.320513	1.900512	5.781278	1.40221	1.425285	1.260996	2.141254	2.235969	2.355227	5.118507	8.22841	7.857564	2.129927	2.33649	2.209283	1.940939	7.200616
2.30103	2.350462	1.887456	5.678831	1.422139	1.431333	1.259375	2.147057	2.273461	2.323841	5.088863	8.24303	7.866914	2.169094	2.38653	2.223995	1.934994	7.211098
2.30103	2.333381	1.942191	5.648778	1.427742	1.427181	1.272111	2.151875	2.275303	2.340012	5.067379	8.256082	7.872256	2.156586	2.409216	2.223926	1.941661	7.203625
2.60206	2.864654	2.128362	5.878563	1.659828	1.768156	1.563377	2.445453	2.626792	2.658258	5.470269	8.543031	8.187047	2.544402	2.74746	2.556907	2.318533	7.568549
2.60206	2.766666	2.149522	5.887471	1.690519	1.720627	1.519126	2.422387	2.604906	2.618967	5.440759	8.609311	8.216684	2.572629	2.766128	2.591306	2.367786	7.545447
2.60206	2.815619	2.24615	5.852139	1.726876	1.742778	1.561501	2.478119	2.621104	2.637534	5.359096	8.600249	8.200935	2.567428	2.812132	2.561516	2.348022	7.521068
3.30103	3.121772	2.309614	6.292235	2.290989	2.219269	1.986101	2.90532	3.049464	3.072469	6.357246	9.385569	8.926366	3.278438	3.809477	3.415052	3.011474	8.297135
3.30103	3.301912	2.276663	6.113176	2.245725	2.23263	2.026941	2.849728	3.109934	3.123496	6.38486	9.336742	8.934414	3.297277	3.786352	3.405443	2.998876	8.345405
3.30103	3.260079	2.587632	6.174343	2.340085	2.165064	2.032869	2.850502	3.10864	3.129724	6.403824	9.364468	8.926782	3.327562	3.796863	3.410608	3.066655	8.340015

B8: Pairing of bile acids with appropriate internal standards

Abbr.	m/z (-)	IS used
CDCA/DCA	391.285	dCDCA
HDCA	391.285	dCDCA
UDCA	391.285	dCDCA
aMCA	407.300	dCA
bMCA	407.300	dCA
CA	407.281	dCA
GLCA	432.311	dGCDCA
GCDCA	448.307	dGCDCA
GDCA	448.307	dGCDCA
GCA	464.300	dGCA
TLCA	482.300	dCDCA
TUDCA	498.300	GCDCA-d4
THDCA	498.289	GCDCA-d4
TCDCA/TDCA	498.289	GCDCA-d4
TCA	514.284	dTCA-d4
TaMCA/TbMCA	514.284	dTCA-d4
TwMCA	514.300	dTCA-d4

Abbr.	m/z (-)
dCA	411.3
dCDCA	395.3
dGCDCA	452.3
dGCA	468.3
dTCA-d4	518.3

B9: Raw data for bile acids detected in biological samples

Specimen	BC16	BC5	BC4	BC9	BC6	BC1	BC11	BC8	BC12	BC7	BC2	BC13	BC14	BC15	BC10
Fat Content	10% Fat	10% Fat	10% Fat	10% Fat	10% Fat	10% Fat	10% Fat	10% Fat	10% Fat	10% Fat	10% Fat	10% Fat	10% Fat	10% Fat	10% Fat
chenodeoxycholate(CDCA)/ deoxycholate(DCA)	8.600	3.501	0.078	3.323	0.003	0.471	0.143	0.030	39.551	3.353	50.349	20.561	0.004	0.000	20.285
Hyodeoxycholate (HDCA)	0.021	0.045	0.001	0.139	0.002	0.003	0.001	0.000	3.295	0.086	8.454	0.218	0.000	0.000	0.554
Ursodeoxycholate (UDCA)	0.044	0.004	0.001	0.022	0.002	0.003	0.002	0.000	1.501	0.003	3.462	0.063	0.000	0.000	0.248
cholate (CA)	2491.320	503.046	121.694	1359.863	34.620	327.572	162.876	192.048	4434.263	734.088	5651.631	2382.554	103.868	163.172	3985.769
beta-muricholate (bMCA)	63.400	105.653	1.496	39.796	2.036	4.368	4.139	10.445	12.764	8.945	21.225	21.930	2.774	4.144	27.177
alpha-muricholate (aMCA)	66.356	51.215	9.281	54.018	1.657	10.728	5.243	10.038	17.899	14.526	26.456	37.564	2.521	7.055	26.517
glycochenodeoxycholate (GCDCA)	0.052	0.025	0.039	0.015	0.024	0.012	0.019	0.023	0.002	0.043	0.026	0.075	0.007	0.063	0.004
glycodeoxycholic acid (GDCA)	0.161	0.015	0.041	0.040	0.026	0.019	0.017	0.016	0.010	0.058	0.106	0.109	0.027	0.072	0.017
glycocholate (GCA)	283.552	318.322	201.624	146.063	99.842	38.530	85.818	251.803	28.004	324.460	160.877	344.878	98.682	1015.406	29.746
tauroolithocholate (TLCA)	0.394	0.053	0.162	0.073	0.101	0.107	0.114	0.162	0.057	0.273	0.217	0.430	0.124	0.195	0.081
taurochenodeoxycholate (TCDCA)	275.126	25.978	50.033	39.130	40.653	40.993	35.745	56.967	17.644	80.087	163.205	222.734	86.556	99.430	35.780
taurohyodeoxycholate (THDCA)	514.890	44.554	88.360	95.533	69.119	71.489	54.798	74.688	41.015	125.254	396.908	364.018	187.378	158.715	85.844
tauroursodoxycholate(TUDCA)/ taurodeoxycholate (TDCA)	64.290	5.781	15.248	12.002	8.997	9.288	8.983	13.561	6.640	18.411	45.461	44.947	31.300	22.280	11.650
taurocholate (TCA)	1917.063	213.712	206.913	248.259	197.388	173.874	203.005	386.484	147.183	500.203	1236.998	2648.459	542.156	643.811	221.964
tauro-muricholate	2207.587	220.932	247.279	293.528	201.492	174.274	213.636	397.257	158.094	461.973	1279.290	2189.039	491.306	484.890	213.196

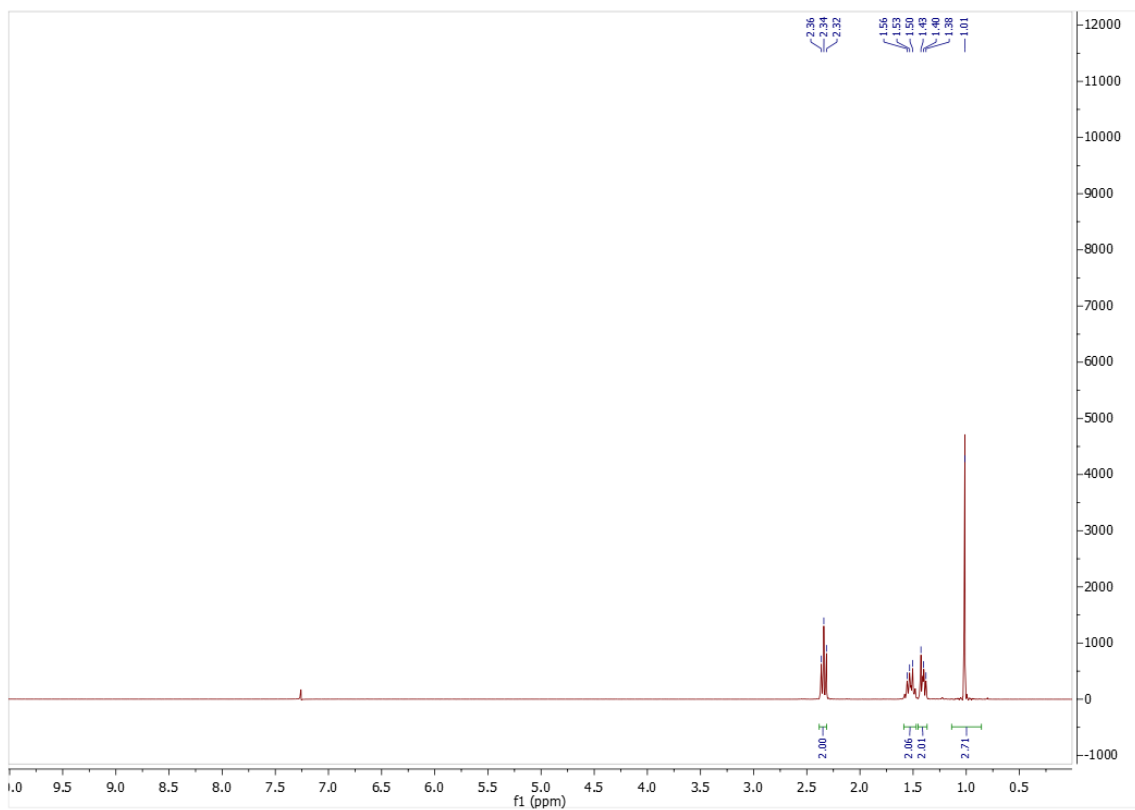
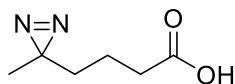
B10: Raw data for N-acyl amino acids detected in biological samples

raw areas	Tcanola20	Tcanola29	Tcanola39	Tcanola48	Tcanola59	Tfish6	Tfish8	Tfish15	Tfish21	Tfish49	Tflaxseed	Tflaxseed	Tflaxseed	Tflaxseed	Tflaxseed	Tlard4	Tlard13	Tlard14	Tlard41	Tlard60	
docosahexaenoyl leucine	0.00E+00	0.00E+00	0.00E+00	6.91E+03	2.48E+03	2.91E+05	1.52E+05	1.35E+05	4.08E+05	3.67E+04	8.52E+03	5.88E+03	3.37E+03	0.00E+00	4.57E+03	0.00E+00	7.55E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00
docosahexaenoyl phenylalanine	3.07E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.93E+05	1.84E+04	0.00E+00	1.35E+05	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.39E+04	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
linoleoyl phenylalanine	2.21E+04	3.44E+04	1.92E+04	5.73E+03	2.94E+04	3.54E+03	6.19E+03	1.80E+03	0.00E+00	0.00E+00	6.11E+03	0.00E+00	4.42E+03	3.34E+03	1.47E+04	1.64E+04	3.81E+05	5.19E+05	3.00E+03	5.73E+04	
dodecanoyl leucine	1.70E+05	1.91E+05	1.32E+05	2.05E+05	1.31E+05	1.82E+05	1.28E+05	1.37E+05	2.03E+05	7.63E+04	1.15E+05	1.00E+05	6.71E+04	6.76E+04	1.16E+05	1.86E+05	9.93E+05	8.86E+05	5.18E+04	1.88E+05	
dodecanoyl valine	1.68E+04	1.60E+04	4.59E+03	4.37E+04	1.06E+04	5.46E+04	0.00E+00	1.08E+04	3.03E+04	3.38E+03	0.00E+00	1.02E+04	0.00E+00	6.73E+03	3.57E+04	2.02E+05	3.26E+04	2.70E+04	3.58E+04	3.05E+04	
linoleoyl leucine	2.33E+05	3.80E+05	2.55E+05	2.49E+05	2.32E+05	1.25E+05	6.57E+04	6.13E+04	1.21E+05	4.69E+04	1.15E+05	7.52E+04	8.00E+04	2.03E+04	1.46E+05	5.70E+05	3.15E+06	4.36E+06	1.25E+04	5.82E+05	
linoleoyl valine	5.08E+04	5.02E+04	6.58E+04	4.56E+04	4.07E+04	2.50E+04	3.64E+04	2.48E+04	2.33E+04	2.81E+04	1.72E+04	1.90E+04	1.37E+04	1.49E+04	4.46E+04	4.09E+04	1.63E+05	4.01E+05	1.38E+04	5.73E+04	
oleoyl leucine	3.15E+05	3.34E+05	2.62E+05	2.22E+05	2.01E+05	1.54E+05	6.30E+04	5.38E+04	9.24E+04	2.46E+04	1.52E+05	8.76E+04	7.41E+04	2.07E+04	1.56E+05	5.47E+05	1.70E+06	2.73E+06	1.22E+04	5.22E+05	
oleoyl phenylalanine	2.93E+04	7.01E+04	4.85E+04	3.46E+04	2.70E+04	7.01E+03	2.30E+03	1.89E+03	5.80E+03	0.00E+00	1.29E+04	3.13E+03	0.00E+00	2.82E+03	1.95E+04	3.68E+04	2.50E+05	2.24E+05	0.00E+00	6.12E+04	
oleoyl valine	5.15E+04	5.38E+04	4.72E+04	3.36E+04	4.11E+04	6.51E+03	4.65E+03	5.67E+03	5.93E+03	1.55E+03	1.64E+04	5.90E+03	6.84E+03	4.32E+03	1.77E+04	4.20E+04	3.56E+05	2.91E+05	0.00E+00	6.91E+04	
palmitoleyl leucine	8.25E+04	1.04E+05	4.27E+04	3.66E+04	5.00E+04	3.68E+05	1.66E+05	9.61E+04	3.24E+05	4.66E+04	3.62E+04	3.42E+04	2.01E+04	7.18E+03	4.74E+04	1.52E+05	6.10E+05	1.15E+06	8.72E+03	1.72E+05	
palmitoyl leucine	1.87E+05	1.73E+05	1.45E+05	1.50E+05	1.13E+05	2.29E+05	1.34E+05	1.19E+05	1.43E+05	8.24E+04	1.77E+05	7.95E+04	6.70E+04	4.52E+04	1.34E+05	3.70E+05	1.09E+06	1.24E+06	6.66E+04	2.92E+05	
palmitoyl phenylalanine	1.42E+04	4.20E+04	1.68E+04	2.50E+04	3.93E+03	2.42E+04	1.44E+04	3.14E+03	1.07E+04	0.00E+00	1.64E+04	6.77E+03	1.30E+04	8.15E+03	9.35E+03	1.23E+04	1.57E+05	1.93E+05	2.32E+03	4.64E+04	
palmitoyl valine	5.77E+04	6.86E+04	6.14E+04	8.48E+04	5.88E+04	4.27E+04	2.24E+04	4.76E+04	4.94E+04	2.14E+04	3.87E+04	3.53E+04	2.80E+04	1.93E+04	4.44E+04	1.01E+05	4.63E+05	8.25E+05	3.45E+04	8.12E+04	
pamilitoleyl valine	8.22E+04	1.70E+05	8.53E+04	8.61E+04	7.91E+04	4.61E+04	4.05E+04	2.86E+04	6.07E+04	1.31E+04	2.37E+04	2.75E+04	1.72E+04	1.61E+04	5.90E+04	9.74E+04	6.94E+05	1.26E+06	8.01E+03	1.37E+05	
docosahexaenoyl valine	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	8.28E+03	3.18E+03	1.59E+03	1.32E+04	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	6.17E+03	0.00E+00	0.00E+00	
palmitoleyl phenylalanine	0.00E+00	8.25E+03	0.00E+00	0.00E+00	0.00E+00	3.02E+04	0.00E+00	1.92E+03	2.04E+04	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	4.24E+03	3.06E+03	8.65E+04	1.36E+05	0.00E+00	1.16E+04	
eicosapentaenoyl valine	1.69E+03	0.00E+00	3.25E+03	0.00E+00	1.52E+03	1.12E+04	3.80E+03	5.28E+03	1.70E+04	5.24E+03	3.13E+03	9.60E+03	6.08E+03	2.09E+03	1.93E+03	2.07E+04	1.36E+04	0.00E+00	8.30E+03	6.90E+03	

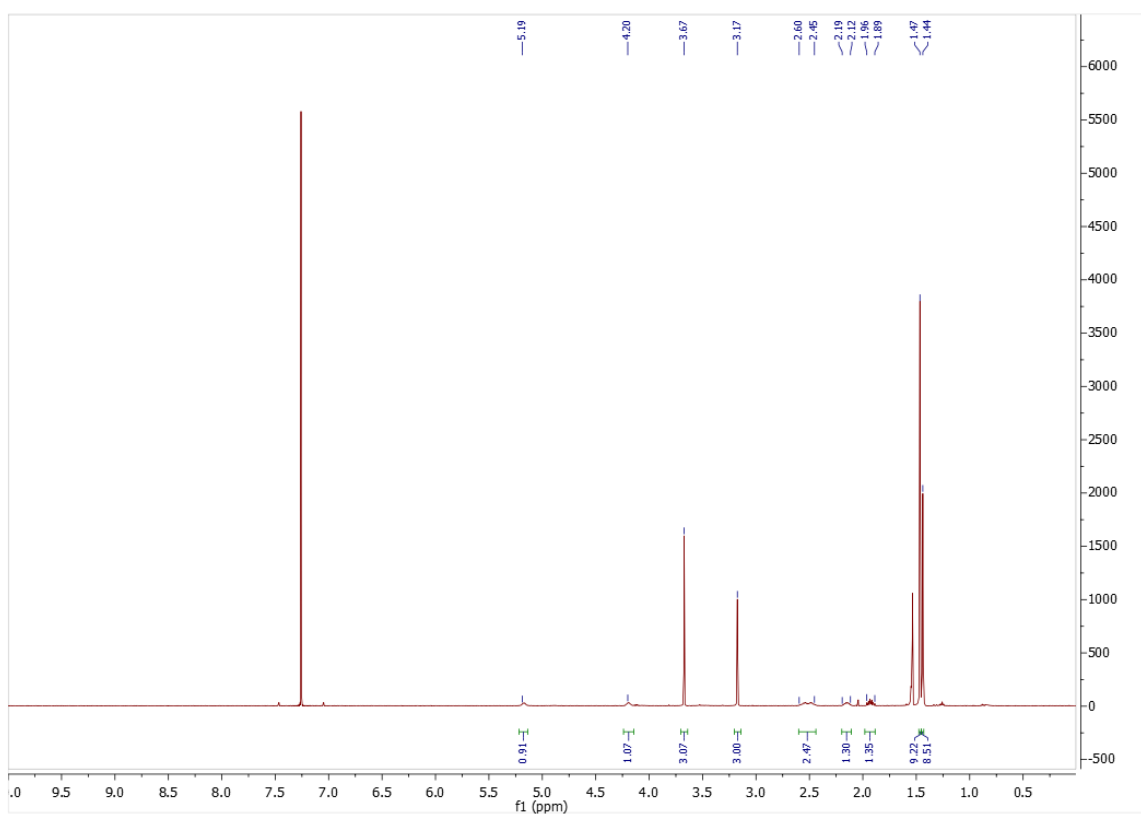
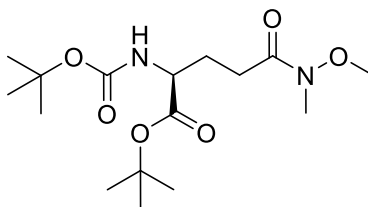
Appendix C: Supporting information for chapter 4

C1: Characterization data for photo-methionine synthesis

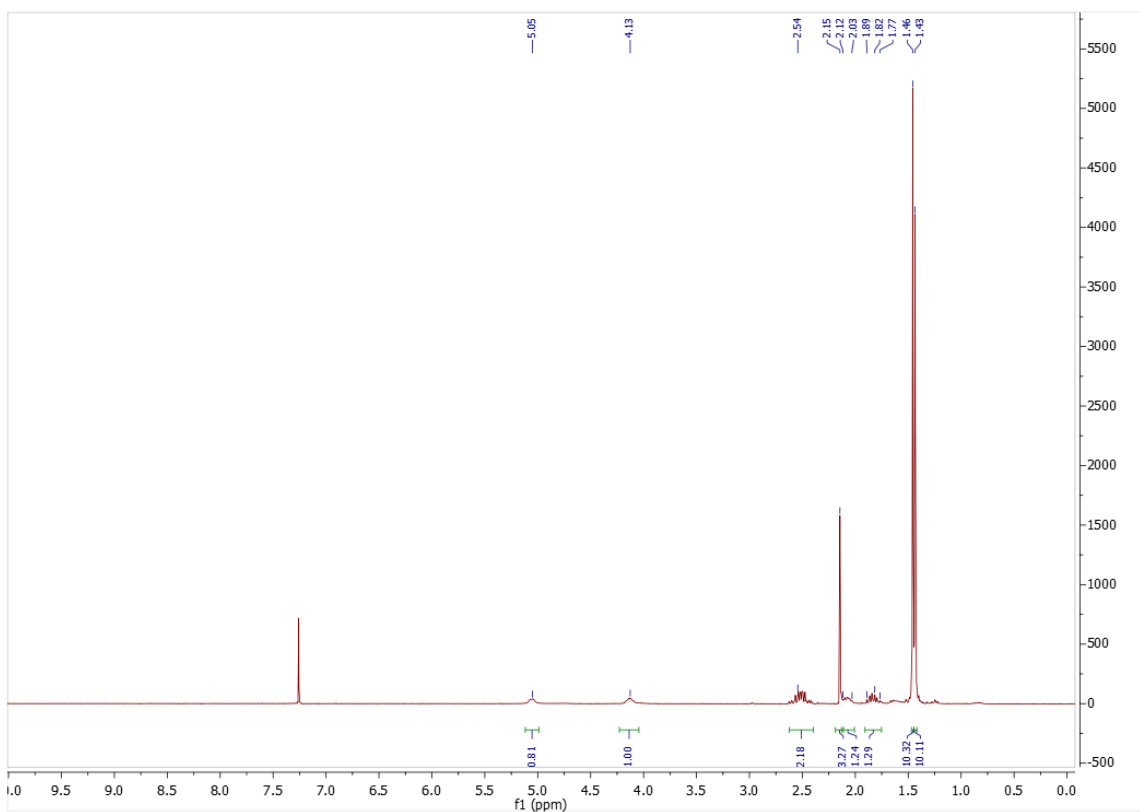
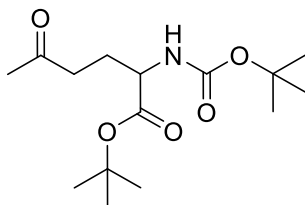
4,4-azohexanoic acid (4): ^1H NMR (300 MHz, CDCl_3): δ 1.01 ppm (s, 3H), 1.38-1.43 (m, 2H), 1.50-1.56 (m, 2H), 2.34 (t, 2H). Solvent residual peak for CDCl_3 at 7.3 ppm.



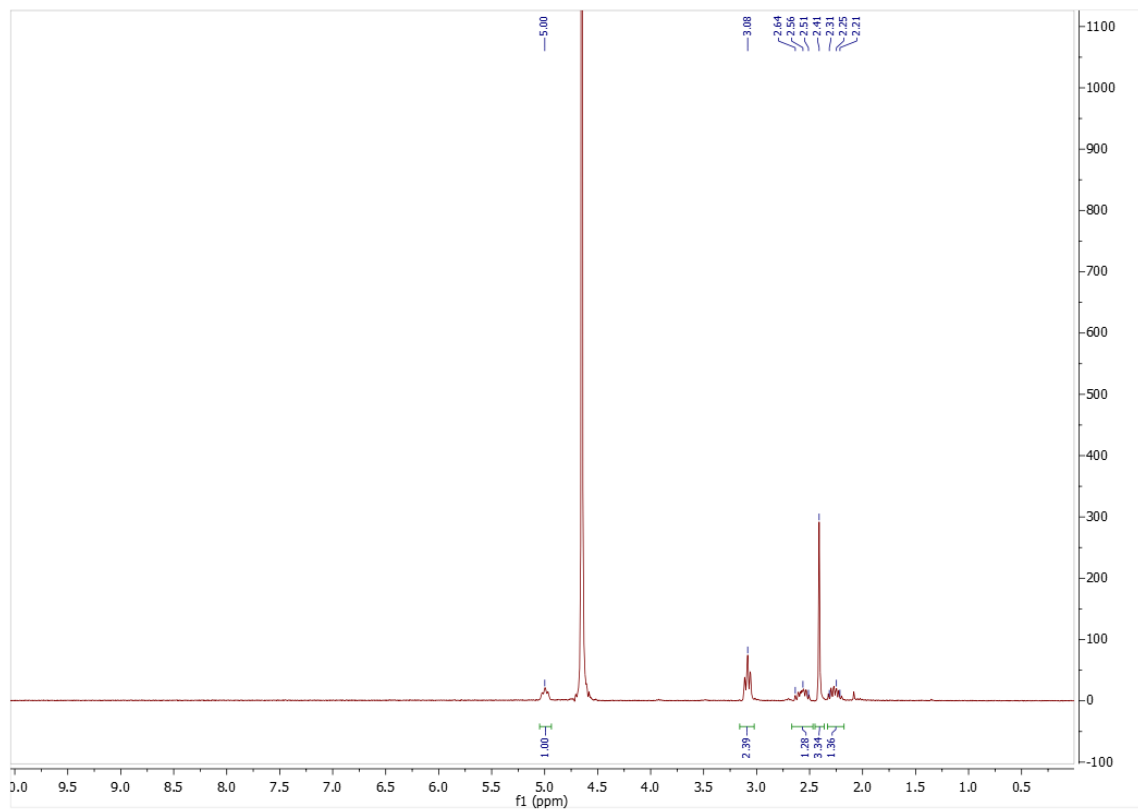
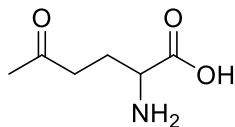
Tert-butyl N²-(tert-butoxycarbonyl)-N⁵-methoxy-N⁵-methylglutamate aka Weinreb amide (7): ¹H NMR (300 MHz, CDCl₃): δ 1.44 ppm (s, 9H), 1.47 (s, 9H), 1.89-1.96 (m, 1H), 2.12-2.19 (m, 1H), 2.45-2.60 (m, 2H), 3.17 (s, 3H), 3.67 (s, 3H), 4.20 (bs, 1H), 5.19 (bs, 1H). Solvent residual peak for CDCl₃ at 7.3 ppm. Residual water peak at 1.6 ppm.



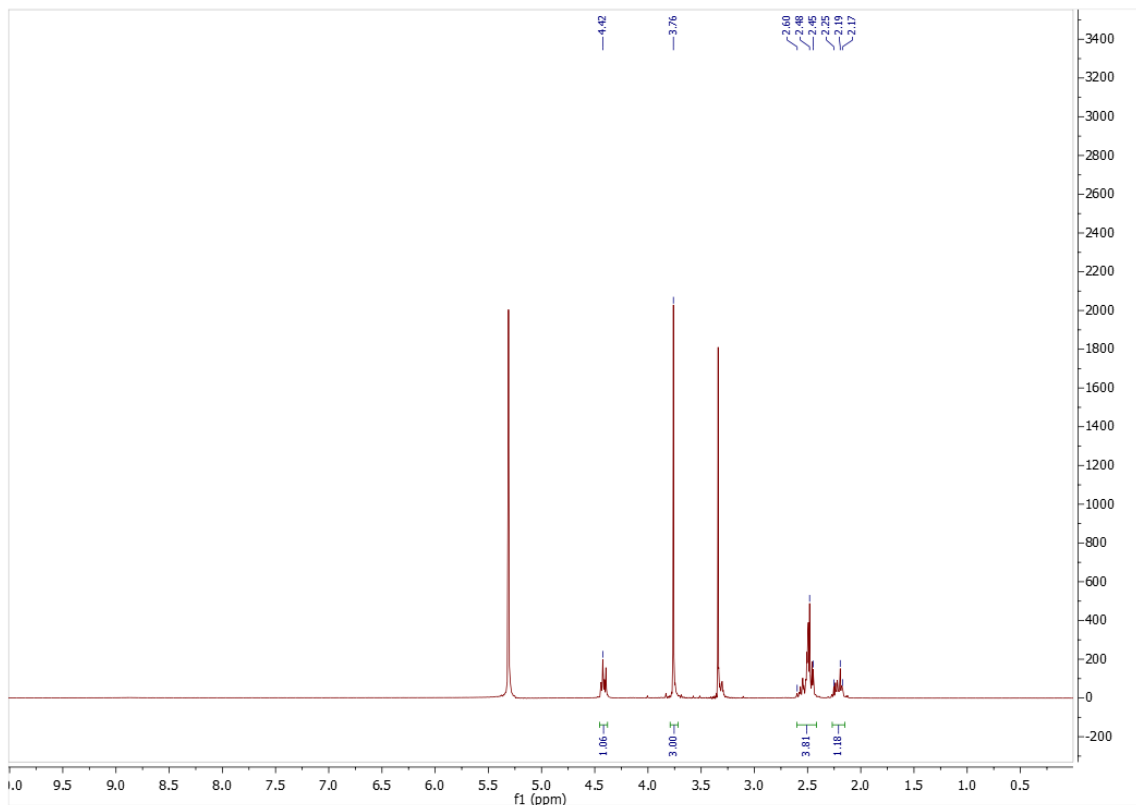
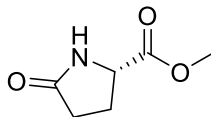
Tert-butyl 2-((tert-butoxycarbonyl)amino)-5-oxohexanoate (8): ^1H NMR (300 MHz, CDCl_3): δ 1.43 ppm (s, 9H), 1.46 (s, 9H), 1.77-1.89 (m, 1H), 2.03-2.12 (m, 1H), 2.15 (s, 3H), 2.54 (m, 2H), 4.13 (bs, 1H), 5.05 (bs, 1H). Solvent residual peak for CDCl_3 at 7.3 ppm.



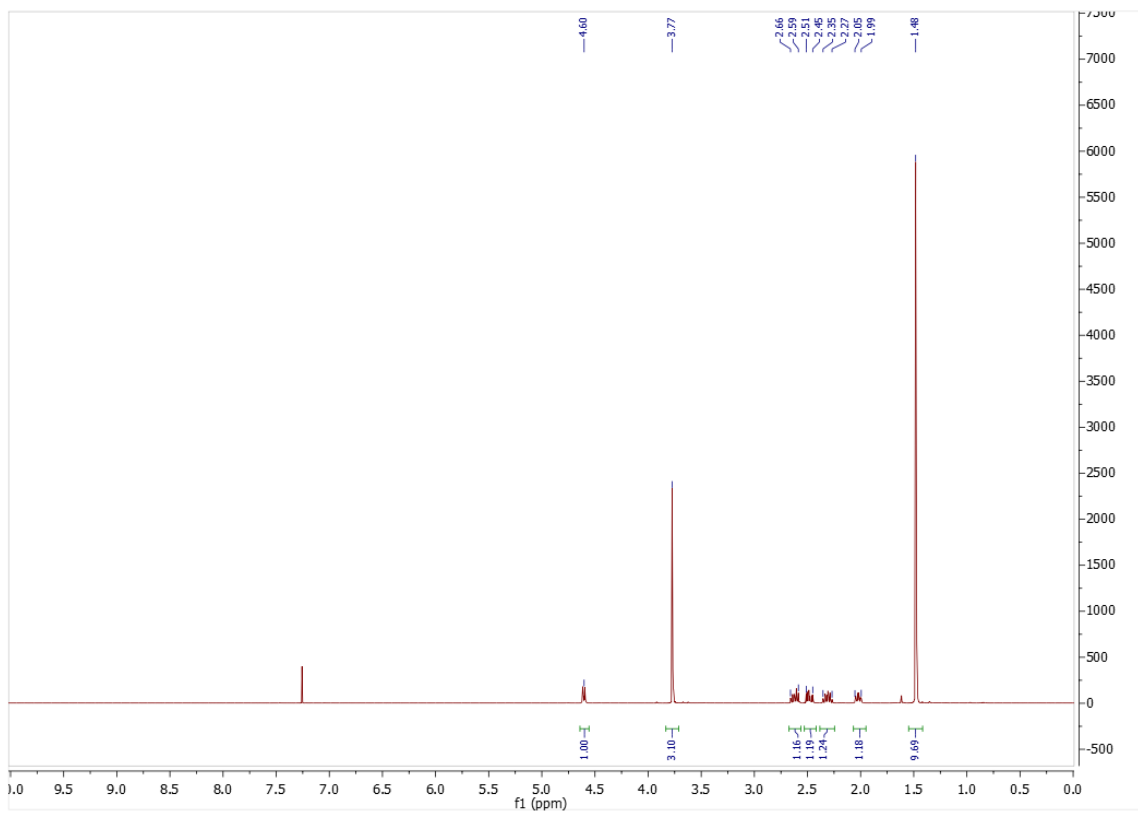
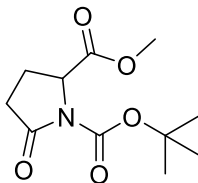
2-amino-5-oxohexanoic acid (10): ^1H NMR (300 MHz, D_2O): δ 2.21-2.31 ppm (m, 1H), 2.41 (s, 3H), 2.51-2.64 (m, 1H), 3.08 (t, 2H), 5.00 (bs, 1H). Solvent residual peak for D_2O at 4.8 ppm.



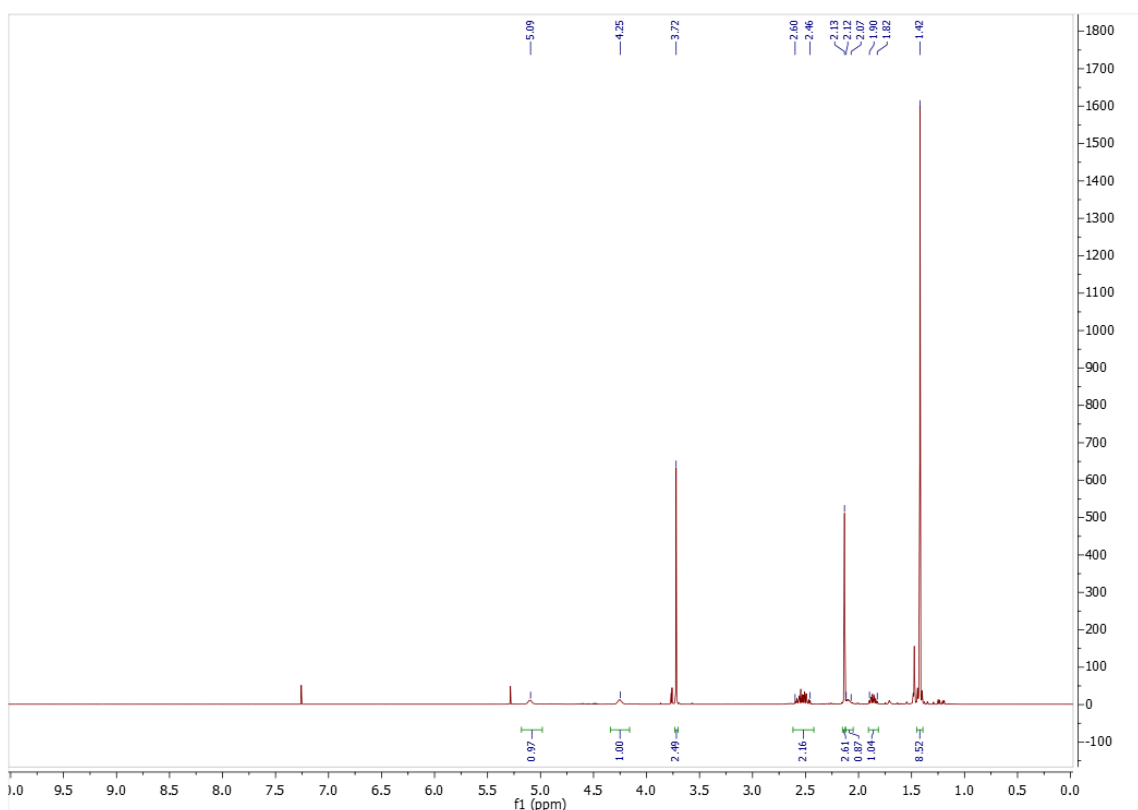
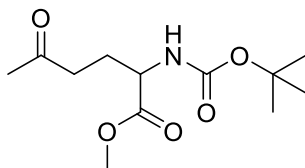
Methyl pyroglutamate (13): ^1H NMR (300 MHz, CD_3OD): δ 2.17-2.25 ppm (m, 1H), 2.45-2.60 (m, 3H), 3.76 (s, 3H), 4.42 (t, 1H). Solvent residual peak for CD_3OD at 3.3 ppm. Residual water at 5.3 ppm.



1-(*tert*-butyl) 2-methyl 5-oxopyrrolidine-1,2-dicarboxylate (14): ^1H NMR (300 MHz, CDCl_3): δ 1.48 ppm (s,9H), 1.99-2.05 (m, 1H), 2.27-2.35 (m, 1H), 3.77 (s, 3H), 4.60 (dd, 1H). Solvent residual peak for CDCl_3 at 7.3 ppm.

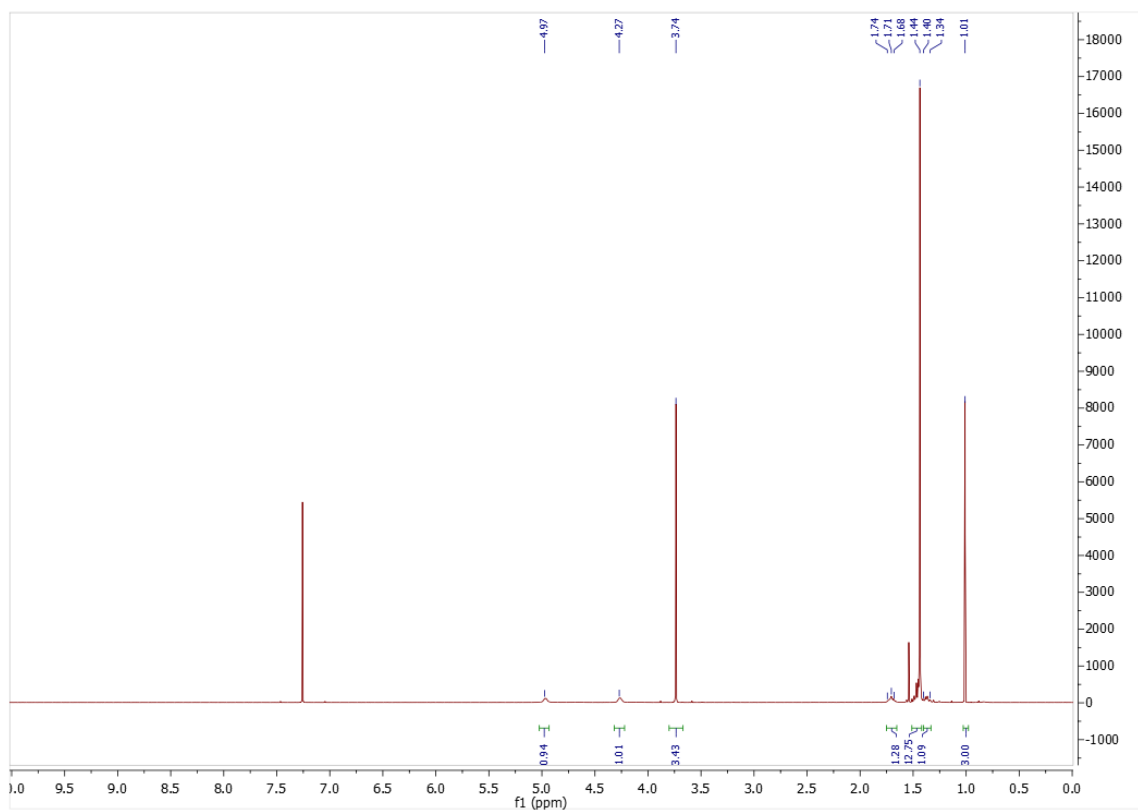
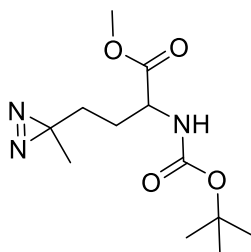


Methyl 2-((*tert*-butoxycarbonyl)amino)-5-oxohexanoate (15): ^1H NMR (300 MHz, CDCl_3): δ 1.42 ppm (s, 9H), 1.82-1.90 (m, 1H), 2.07-2.12 (m, 1H), 2.13 (s, 3H), 2.46-2.60 (m, 2H), 3.72 (s, 3H), 4.25 (bs, 1H), 5.09 (bs, 1H). Solvent residual peak for CDCl_3 at 7.3 ppm. Residual water peak at 1.6 ppm.



Methyl 2-((*tert*-butoxycarbonyl)amino)-4-(3-methyl-3H-diazirin-3-yl)butanoate (16):

^1H NMR (300 MHz, CDCl_3): δ 1.01 ppm (s, 3H), 1.34-1.40 (m, 1H), 1.44 (bs, 12H), 1.68-1.74 (m, 1H), 3.74 (s, 3H), 4.27 (bs, 1H), 4.97 (bs, 1H). Solvent residual peak for CDCl_3 at 7.3 ppm.



2-((*tert*-butoxycarbonyl)amino)-4-(3-methyl-3H-diazirin-3-yl)butanoic acid (17): ^1H NMR (300 MHz, CDCl_3): δ 1.03 ppm (s, 3H), 1.44 (bs, 12H), 1.80 (m, 2H), 4.28 (bs, 1H), 4.95 (bs, 1H). Solvent residual peak for CDCl_3 at 7.3 ppm.

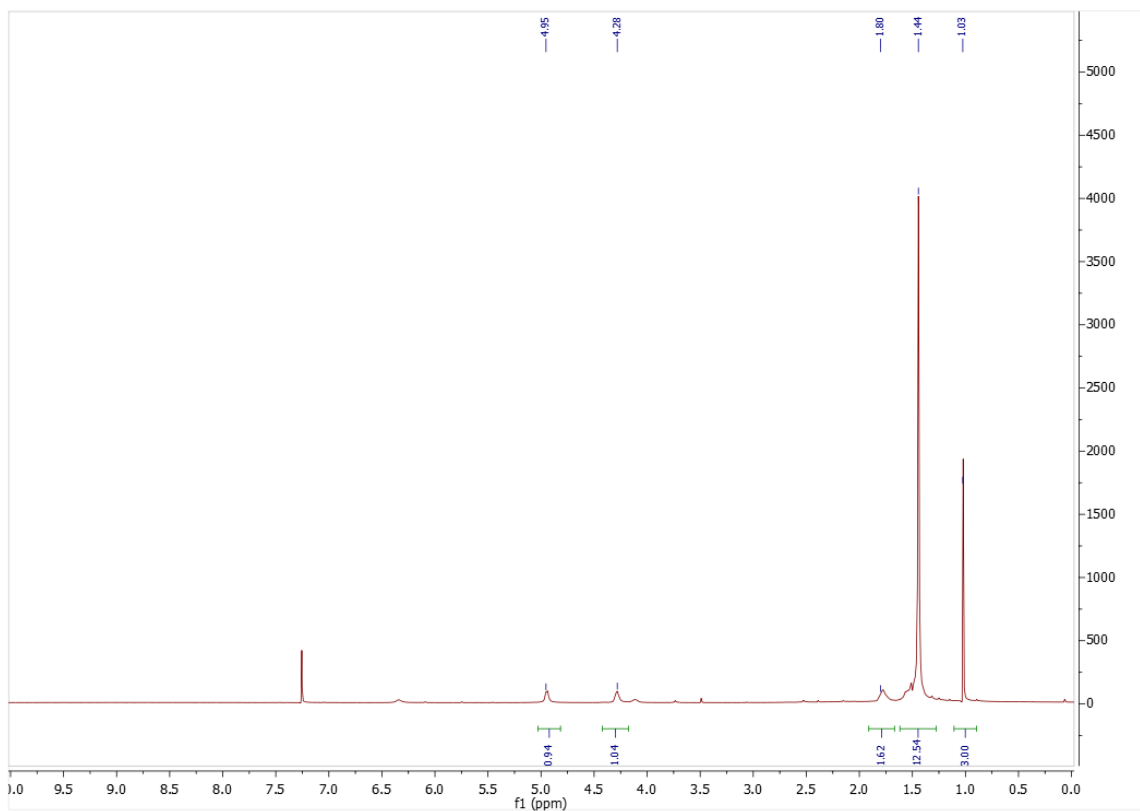
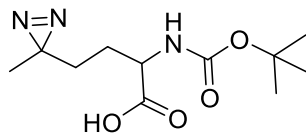
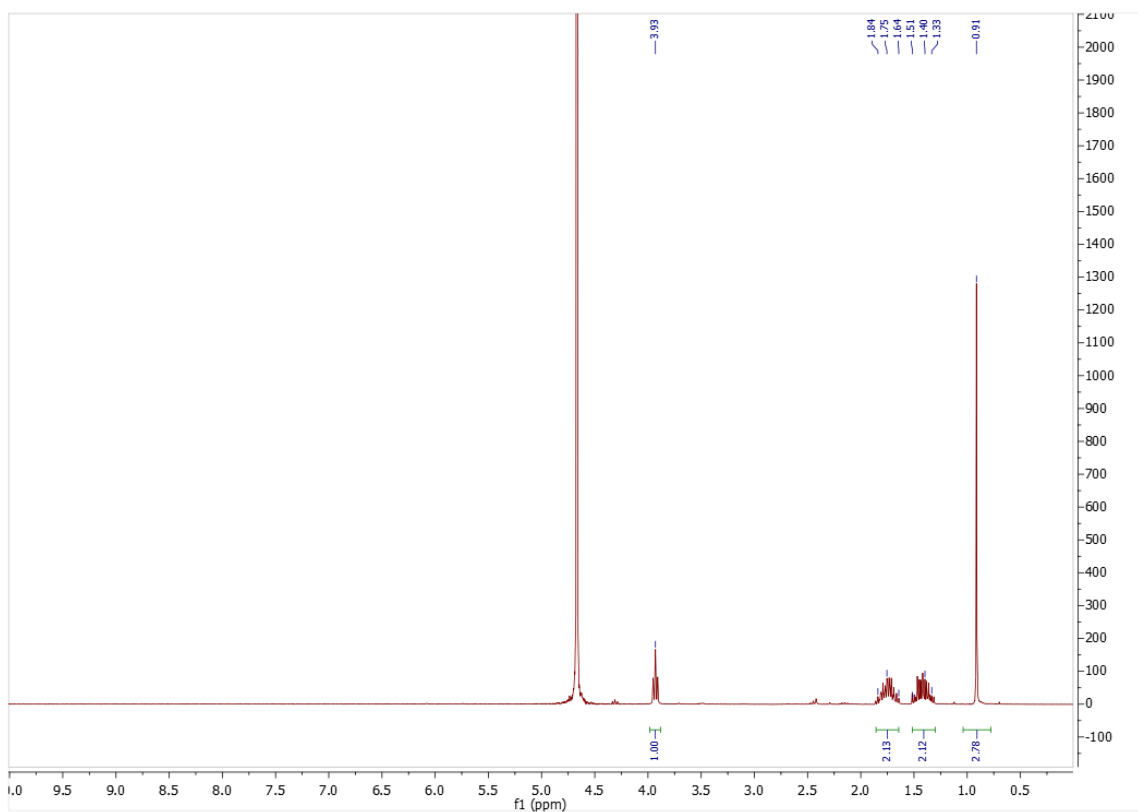
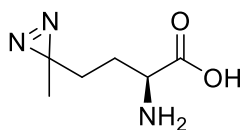


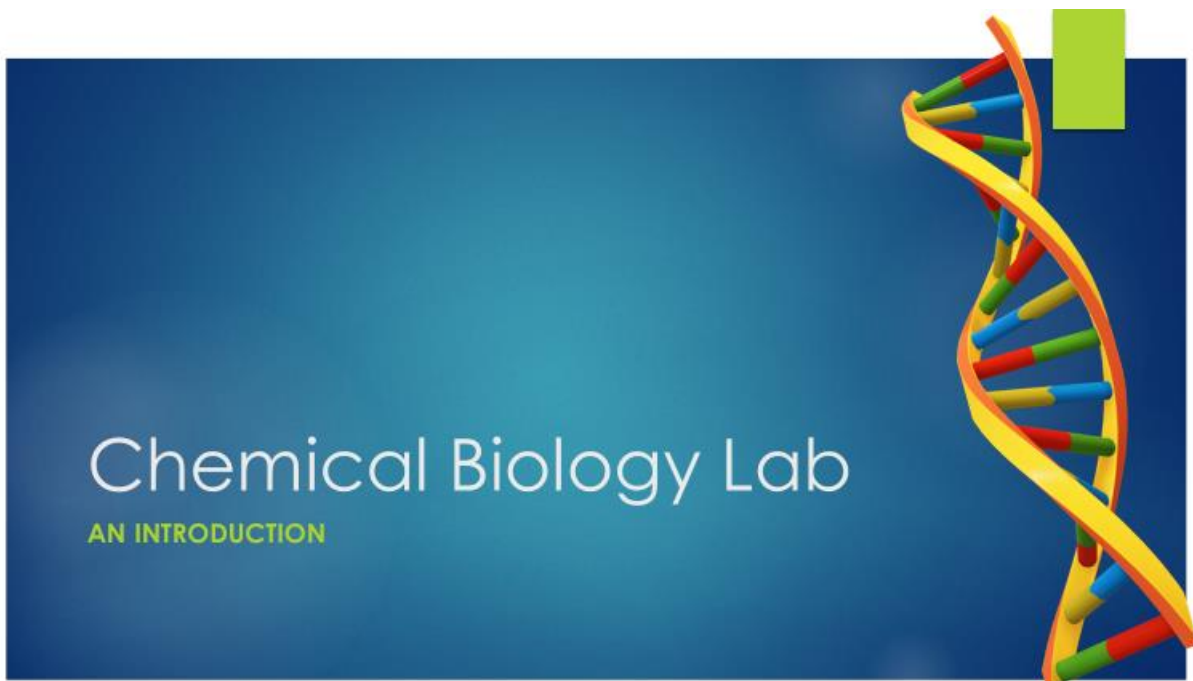
Photo-methionine (1): ^1H NMR (300 MHz, D_2O): δ 0.91 ppm (s, 3H), 1.33-1.51 (m, 2H), 1.64-1.84 (m, 2H), 3.93 (t, 1H). Solvent residual peak for D_2O at 4.7 ppm.



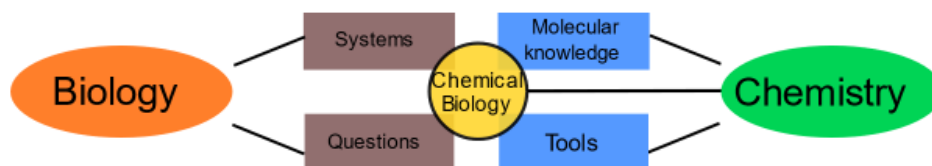
Appendix D: Course content for chapter 5

Week 0: Orientation

Pre-lab material



What is chemical biology?



The application of chemical tools and molecular knowledge to study biological questions and/or biological systems.

What chemists offer to biology

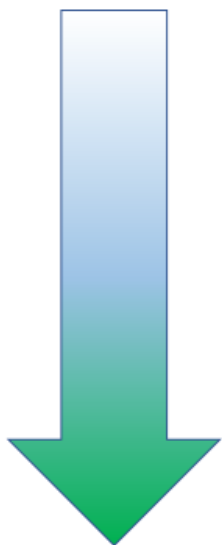
Molecular knowledge

- Catalysis
- Chemical kinetics
- Solution equilibria
- Physical properties of molecules
- Inter- and intra-molecular forces
- Synthetic insights
- Thermodynamics
- Molecule-level measurements

Chemical tools

- Synthesis
- Bioorthogonal reactions
- Separation/purification
- Spectroscopy
- Magnetic resonance
- Mass spectrometry
- Computations
- Microscopy and diffraction

Expectations



- Knowledge
 - Memorizing facts, remembering what you did in lab
- Comprehension
 - Describing techniques
- Application
 - Interpreting results
- Analysis
 - Testing hypotheses, questioning outcomes of experiments
- Synthesis
 - Composing an lab report to support findings
- Evaluation
 - Defending experimental methods

Adapted from Bloom and Krathwohl 1956

Objectives

- You will understand the interface of the two sciences and how chemical biology exemplifies interdisciplinary science.
- You will create a chemical biology concept map for every set of techniques. It should display your understanding of the interface.
- You will be familiar with the basic lab techniques covered, and be able to describe and apply them.
 - PCR, Electrophoresis, MS/MS peptide sequencing, Solid-phase peptide synthesis, etc...
- You will analyze your repertoire of chemical tools to decide which techniques are best for investigating a given biological question.
- You will critically evaluate the outcomes of your experiments and be able to defend your reasoning.
- You will write scientifically in a way that is accurate, descriptive, and effectual.

Week 1: Genes: Primer design and DNA extraction

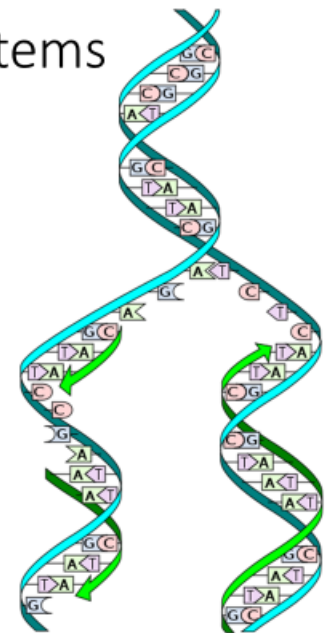
Pre-lab material

An introduction to the concepts of Polymerase Chain Reaction (PCR)



How DNA is copied in biological systems

- DNA helicase separates double-stranded DNA
- Each single strand acts as a template
- DNA polymerase copies the single strands by extending the 3' end of an existing strand
- Polymerase cannot initiate a new strand, only elongate one



What is PCR?

- A way to copy DNA *in vitro* that mimics the natural process
- Uses chemical tools and insights to help the process occur outside of a biological system

The origin of PCR

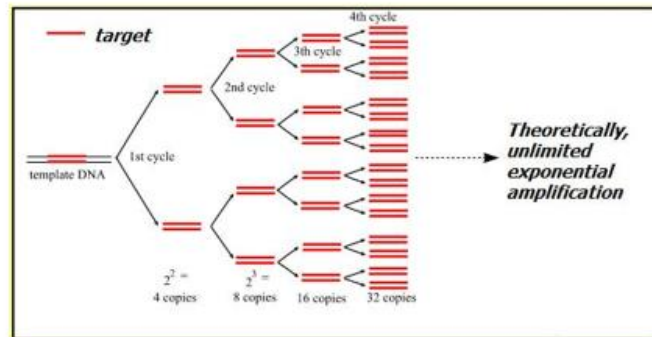
- Dr. Kary Mullis
 - B.S. in Chemistry, P.h.D in Biochemistry
 - Took time off to write fiction and manage a bakery along the way
- In 1983 (working for Cetus Corp.), Mullis was driving in his car late at night when he had a thought:
 - Bracket a section of DNA between primers and copy it using an enzyme to do the synthetic work
- Won Nobel prize (shared) in 1993 for his work on PCR



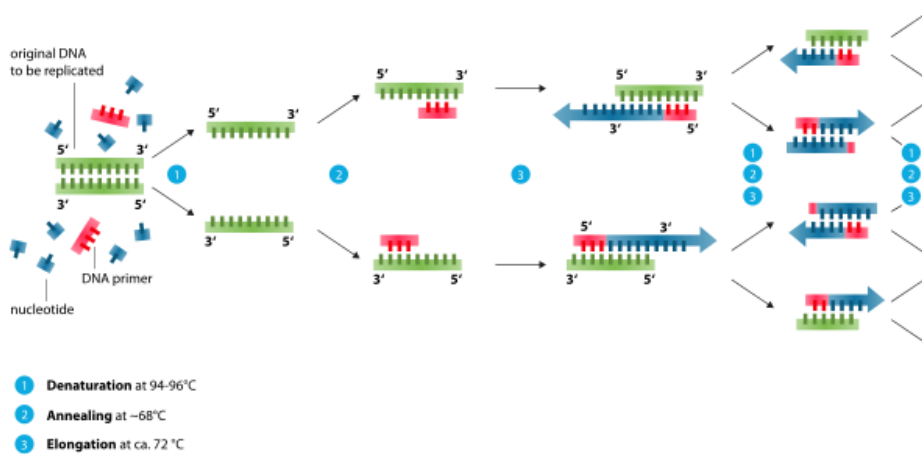
Why copy DNA?

- Amplification

- Quickly produce thousands to millions of copies of a DNA segment
- Low concentrations of DNA can be transformed into a generous and useful concentrations



The 3 Cycles of PCR



By Enzoklop - Own work, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=32003643>

The Necessary Pieces

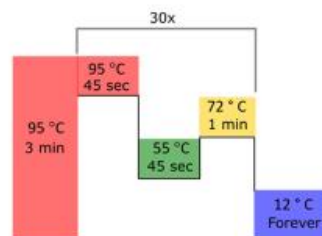
1. A template to copy (DNA target)
2. Markers to select the desired template (primers)
 - Must bind selectively and strongly
3. A catalyst for the polymerization reactions (*Taq* polymerase)
 - Thermostable enzyme from *Thermus aquaticus*. Must be buffered for optimal activity
4. Building blocks to synthesize the copies (nucleotides)

Thermal Cyclers

- Closely regulates temperature
- Will perform automated cycles



{ 95° C (Denature)
55° C (Anneal)
72° C (Elongate) } 30 X



Introduction

Polymerase chain reaction (PCR) is a technique that is used to amplify a region of DNA by rapidly synthesizing thousands to millions of copies. To perform PCR, several basic components are needed: the DNA template to be amplified, deoxynucleotide triphosphate monomers (dNTPs), a DNA polymerase enzyme that performs the polymerization reactions, and DNA “primers” to serve as a starting point for the synthesis. The entire process is carried out in a thermo cycler, a piece of equipment that subjects the reactions to repeated cycles of heating. The procedure can be subdivided into different steps, each with corresponding temperatures. In the first step, the DNA that is intended for amplification must be denatured into a single strand. This is accomplished by heating at 94-98 °C. In the second step, the temperature is dropped to 50-60 °C so the primers can be annealed to the single-stranded DNA template. Next, the temperature is raised to 70-80 °C and the DNA is elongated by the enzyme. The cycles are represented by the figure below.

Primers allow the initiation of the DNA polymerization by providing short sections of double-stranded DNA which can be elongated. They are essential for DNA replication, since the polymerase enzymes are only capable of elongating existing DNA chains.

Objectives

A crucial part of performing a PCR experiment is primer design. Since the primer signals the start position of the DNA synthesis, it controls the specificity of the reaction. In other words, if the primer does not bind exclusively to the intended area of the DNA sequence, then unwanted segments of DNA can be unintentionally amplified. In this lab period, the process of primer design will be explored. Starting from a gene that will serve as the DNA template, primers will be designed and assessed based on several criteria for successful PCR.

Materials

- Laptop/notebook computers
- Primer design handouts

Methods

1. Open web browser and navigate to www.addgene.org.
2. Enter “pUC19” in the “search for plasmids” query and select the top hit. The pUC19 plasmid is isolated from *E. coli* and contains 3 genes, including the gene for resistance to the antibiotic ampicillin, AmpR. The AmpR gene will be the DNA template of this multi-week PCR experiment. The goal is to detect ampicillin resistance in *E. coli* pUC19 DNA. This will be accomplished by PCR amplification of AmpR and electrophoresis to resolve the PCR products.
3. The sequence map should be visible in the top left corner of the pUC19 page. Click on “view all sequences” just below the map image.
4. On the new page, click “analyze sequence” near the map image.
5. On the resulting screen, select the “sequence” tab.
6. Scroll down until you see the AmpR portion of the sequence.
7. Using the select tool (shaped like +), carefully select the entire AmpR gene.
8. Once the entire gene is highlighted, a clipboard window will appear with the text sequence of the AmpR portion of pUC19. Copy the sequence.
9. Open a new window of Microsoft Word and paste the sequence into it.
10. With the sequence highlighted, click on the word count in the lower left corner to bring up the info window. The sequence should be 861 bases in length. This can be confirmed by the character count. If the character count is not 861, repeat steps 7-9.
11. Sequences with fewer than 1000 nucleotides are the easiest to amplify. It is possible to amplify partial genes, but since the entire AmpR sequence is 861 bases, the entire gene can be easily amplified. Remember that the template DNA is not a single strand, but rather a double strand. Next, the complimentary strand will be generated so both strands of the DNA can be visualized.
12. Open a new web browser tab.
13. Navigate to http://molbiol.edu.ru/eng/scripts/01_12.html. You should see a page with the heading “Sequence Utilities.”
14. Copy the AmpR sequence from Microsoft Word and paste it into the sequence box.
15. Check the box that says “display double stranded sequence” and convert.

16. Copy and paste the resulting text in another page of the Microsoft Word document.
17. Both forward and reverse primers are necessary. Recall that the double-stranded DNA will be denatured before annealing and elongation. The forward primer will bind at the 5' end of one strand of the DNA that will be copied, and the reverse primer will bind at the 5' of complimentary strand. There are several initial guidelines for primers that will inform design. Primers should be approximately 18-25 nucleotides in length. Additionally, primers should both begin and end with G/C. This is due to the added advantage of the third hydrogen bond between G/C base pairs. With these guidelines in mind, choose sections of AmpR DNA that could be ideal primers. In Microsoft Word, highlight the potential primer areas in yellow.
18. Copy the portion of the sequence that was chosen as the forward primer. Paste it on a new page of the same Word document and label it "forward primer". Do the same for the reverse primer. Denote the 5' and 3' ends of each primer.
19. Navigate to www.idtdna.com.
20. Under the "tools" dropdown menu, select "OligoAnalyzer Tool"
21. Copy the sequence of the forward primer and paste it into the "sequence box". Click the "analyze" button.
22. The software should generate several parameters that are useful for determining whether or not the primer is suitable. For instance, the GC content should be between 30 and 60 %. Likewise, the T_m (melt temp) should be between 56 and 62 °C. Copy and paste the length, GC content, and T_m into the existing Word document.
23. Repeat steps 20 -22 for the reverse primer.
24. Lastly, it is crucial that primers are highly selective. To ensure that a primer will only bind to the intended sequence, it must be compared against all DNA present. If a primer sequence is repeated elsewhere in the pUC19 plasmid, it could cause amplification of unwanted genes. To confirm selectivity, return to www.addgene.org and navigate to the page for pUC19.
25. Click on "view all sequences".

26. Copy and paste the entire 2686 base pUC19 sequence into a new, blank Word document. Using the “find” function (Ctrl + F or Command + F) search both the forward and reverse primers and ensure they only occur once in pUC19.
27. If primers do not meet the given criteria, consult earlier steps and re-design them.
28. Once primers meet all criteria, save the original document containing the AmpR sequence, double stranded DNA sequence, and selected primers with their properties. Include a short write-up (1 paragraph) describing why the chosen primers are suitable for PCR. Ensure that the document is formatted in an orderly way. Name the file based on the following model: “YYMMDD_LastNamePartner1_LastNamePartner_2_Primer_Design”. Upload the file to Canvas. This will serve as a lab report for week 1 and will be graded.
29. Once you have completed the primer design exercise, see TA to begin DNA extraction.
30. Extract PUC19 DNA using an Invitrogen Plasmid Mini Kit using the supplied instructions (below).

Use a microcentrifuge capable of centrifuging at $>12,000 \times g$. Perform all centrifugation steps at room temperature using a microcentrifuge.

- Optional: Preheat an aliquot of TE Buffer (TE) to 65–70°C for eluting DNA. Heating is optional for eluting 1–30 kb plasmid DNA but is recommended for eluting DNA >30 kb.

- Caution: Buffers contain hazardous reagents. Use caution when handling buffers. Isolate miniprep plasmid DNA

1. Harvest. Centrifuge 1–5 mL of the overnight LB-culture. (Use $1-2 \times 10^9$ E. coli cells for each sample.) Remove all medium.

2. Resuspend. Add 250 μ L Resuspension Buffer (R3) with RNase A to the cell pellet and resuspend the pellet until it is homogeneous.

3. Lyse. Add 250 μ L Lysis Buffer (L7). Mix gently by inverting the capped tube until the mixture is homogeneous. Do not vortex. Incubate the tube at room temperature for 5 minutes.

4. Precipitate. Add 350 μ L Precipitation Buffer (N4). Mix immediately by inverting the tube, or for large pellets, vigorously shaking the tube, until the mixture is homogeneous. Do not vortex. Centrifuge the lysate at $>12,000 \times g$ for 10 minutes.

5. Bind. Load the supernatant from step 4 onto a spin column in a 2-mL wash tube. Centrifuge the column at 12,000 × g for 1 minute. Discard the flow-through and place the column back into the wash tube.
6. Optional Wash. (Recommended for endA+ strains). Add 500 μL Wash Buffer (W10) with ethanol to the column. Incubate the column for 1 minute at room temperature. Centrifuge the column at 12,000 × g for 1 minute. Discard the flowthrough and place column back into the wash tube.
7. Wash and remove ethanol. Add 700 μL Wash Buffer (W9) with ethanol to the column. Centrifuge the column at 12,000 × g for 1 minute. Discard the flowthrough and place the column into the wash tube. Centrifuge the column at 12,000 × g for 1 minute. Discard the wash tube with the flow-through.
8. Elute. Place the Spin Column in a clean 1.5-mL recovery tube. Add 75 μL of preheated TE Buffer (TE) to the center of the column. Incubate the column for 1 minute at room temperature.
9. Recover. Centrifuge the column at 12,000 × g for 2 minutes. The recovery tube contains the purified plasmid DNA. Discard the column. Store plasmid DNA at 4°C (short-term) or store the DNA in aliquots at -20°C (long-term).

Week 2: Genes: Amplification with PCR

Pre-lab material

Review slides from week 1

Required reading

Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase

Randall K Saiki et. al.

Science

January 1988

DOI: 10.1126/science.239.4839.487

Introduction

Polymerase chain reaction (PCR) is a technique that is used to amplify a region of DNA by rapidly synthesizing thousands to millions of copies. To perform PCR, several basic components are needed: the DNA template to be amplified, deoxynucleotide triphosphate monomers (dNTPs), a DNA polymerase enzyme that performs the polymerization reactions, and DNA “primers” to serve as a starting point for the synthesis. The entire process is carried out in a thermal cycler, a piece of equipment that subjects the reactions to repeated cycles of heating. The procedure can be subdivided into different steps, each with corresponding temperatures. In the first step, the DNA that is intended for amplification must be denatured into a single strand. This is accomplished by heating at 94-98 °C. In the second step, the temperature is dropped to 50-60 °C so the primers can be annealed to the single-stranded DNA template. Next, the temperature is raised to 70-80 °C and the DNA is elongated by the enzyme.

The most common polymerase employed in PCR is isolated from the thermophilic bacterium *Thermus aquaticus*, which lives in hot springs and hydrothermal vents. Often abbreviated Taq Pol or simply Taq, this enzyme is unique among polymerases as it can withstand the high temperatures required for DNA denaturation in the initial step of a PCR cycle. To function properly, Taq must be buffered. It has optimal activity at a pH of 9. Once the primers are annealed to the denatured DNA, Taq will extend the primers until the end of the elongation step. Then, the process of denaturation, annealing, and elongation will be repeated sequentially until all cycles are complete. This should yield millions of exact copies of the target sequence. This makes PCR an extremely powerful tool, as endogenous levels of DNA can be very difficult to detect using common laboratory techniques.

Objectives

In this experiment, the gene for ampicillin resistance in the pUC19 plasmid isolated from *Escherichia coli* (*E. coli*) will be amplified using primers like the ones that were designed in the previous exercise. Each pair of students will be responsible for running two PCRs: a pUC19 isolate and a control sample that does not contain the target AmpR gene. While

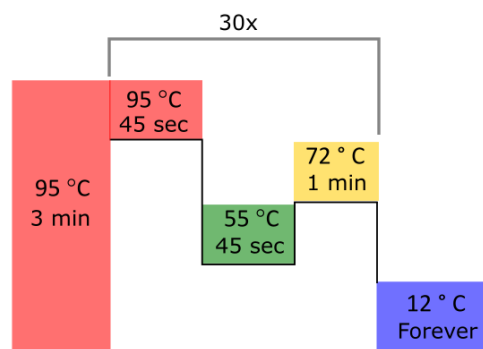
the reactions are being performed in the thermal cycler, the concepts of PCR will be solidified by completing a series of short-answer questions.

Materials

- pUC19 plasmid isolate
- Solution of forward primer (20 μ M)
- Solution of reverse primer (20 μ M)
- Illustra™ Ready-to-Go™ PCR beads
- Milli-Q water
- Thermal cycler

Methods

1. On the thermal cycler, program the following conditions (the bolded steps are the repeated cycles, which should be performed 30 times):
 - a) Hold at 95 °C for 3 minutes
 - b) Hold at 95 °C for 45 seconds
 - c) Hold at 55 °C for 45 seconds
 - d) Hold at 72 °C for 1 minute
 - e) Final hold at 10-12 °C indefinitely



2. To each Illustra™ Ready-to-Go™ PCR bead in a 500 μ L Eppendorf tube, add 22 μ L of Milli-Q water.

3. To each tube, add 1 μL forward primer and 1 μL of reverse primer.
4. Add 1 μL Puc19 sample to one tube, and 1 μL negative control to the other. Label the tubes appropriately with your name and the contents of the tube.
5. Place to tubes in the thermal cycler and start the reaction cycles.
6. While reactions are occurring, answer the questions below.
7. After the questions have been completed, proceed to the gel loading station and practice loading gels.
8. Once amplification is finished, remove vial and return it to the refrigerator.

Questions

1. The Illustra™ Ready-to-Go™ PCR beads are pre-packaged for fast and convenient PCR. In order to perform a successful PCR, the beads must contain 3 things. What are the three things? Explain why each one is necessary.

2. Assume that there is a sufficient amount of primer in each reaction vial such that each cycle produces a perfect copy of every available DNA strand. How many single strands of DNA can be produced from one fragment of double-stranded DNA after all 30 cycles? Show your math.

3. The initial cycle of copies of the DNA template are different from all subsequent copies. How and why do they vary? HINT: Where is the polymerization reaction halted?

4. Predict the results of a PCR experiment in which the following temperatures are too low. Discuss how the products would be altered.
 - a. Denaturation temperature

 - b. Annealing temperature

 - c. Extension temperature

Week 3: Genes: Separation and Detection with Gel Electrophoresis

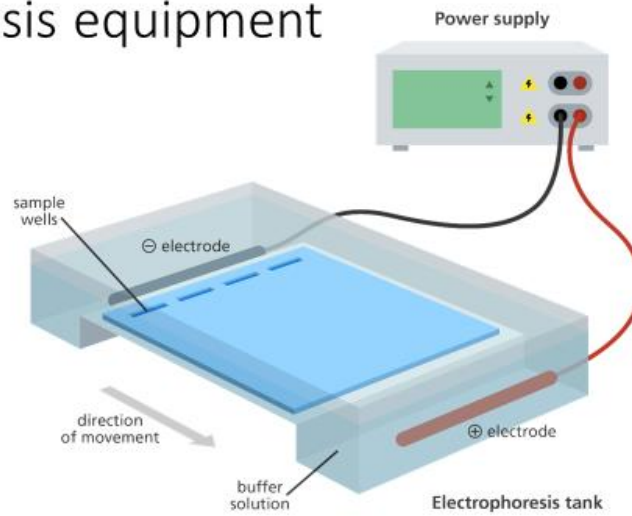
Pre-lab material



What is gel electrophoresis?

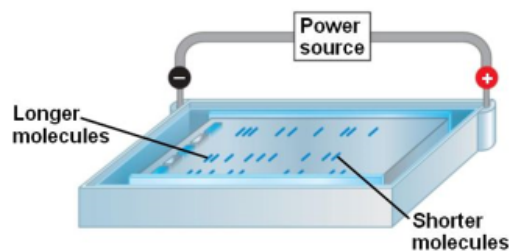
- Technique for separation and analysis.
 - Most commonly used for DNA, RNA, and proteins.
 - Performs separation based on size and charge
 - DNA is loaded into gel, electric field is applied, DNA fragments move through gel with the electric field.
- Agarose (polysaccharide extracted from seaweed) is most commonly used for DNA/RNA separations

The electrophoresis equipment



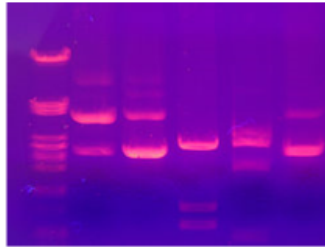
How does the separation happen?

- DNA is negatively charged (due to the phosphate groups)
- When electric field is applied, DNA will migrate towards the (+) electrode
- The agarose matrix will slow down the larger fragments



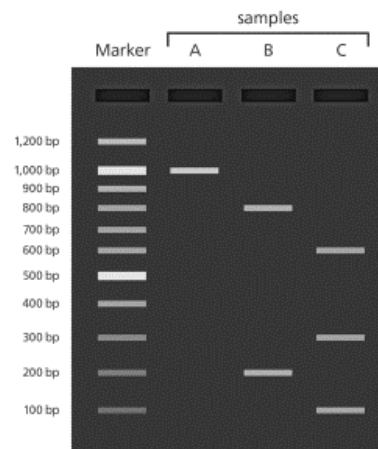
Visualizing the bands

- Many dyes available, but ethidium bromide (EtBr) is common for DNA
 - **CAUTION:** EtBr is a mutagen. Use extreme care.
- EtBr intercalates between DNA and absorbs UV light
 - Must use short wave UV light to see the bands



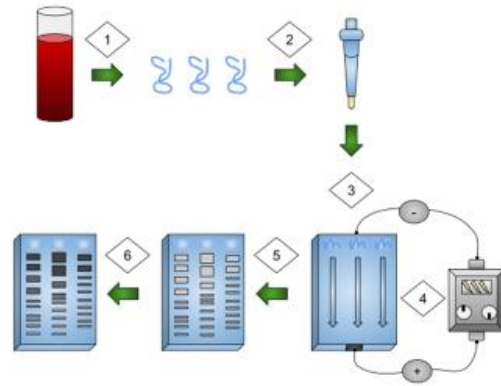
Interpreting the results

- Use size markers as standards to determine size of DNA fragments.
 - Mixture of markers called “ladder”
- Ladder will be loaded into the gel alongside the samples and used to determine the size of the DNA in the samples.



The workflow

1. Extract DNA (Done)
2. Amplify DNA with PCR (Done)
3. Load DNA in gel wells
4. Apply electric current
5. Allow "bands" to separate over time
6. Visualize DNA with dye



Required reading

Gel Electrophoresis

Reiner Westermeier

Wiley Encyclopedia of Life Sciences

2005

DOI: 10.1038/npg.els.0005335

Introduction

PCR, an incredibly beneficial tool for molecular biology, is widely used for amplification of DNA. However, it does not yield useful data unless it is paired with another technique. A common method of detecting PCR products is gel electrophoresis—a technique which separates biopolymers in a gel matrix by their charge and size. Electrophoresis can be employed for the separation of proteins, as well as DNA and RNA. For nucleic acid polymers (DNA and RNA), the gel matrix is usually comprised of agarose, a polysaccharide polymer extracted from algae. In DNA electrophoresis, the sample is loaded into a small well in the agarose gel, which is immersed in a buffered electrolyte solution. An electric field is applied to the chamber via an external power supply. As the negatively-charged DNA migrates towards the positive terminal of the chamber, the fragments are slowed by their movement through the pores in the agarose, a phenomenon called “sieving”. Since the smallest fragments of DNA are least hindered on their path through the gel, they move the fastest—and therefore the furthest in a given amount of time. For the data to be meaningful, the migration of unknown DNA must be compared to a reference mixture of known DNA fragments known as a “ladder”. A visualization reagent (dye) must be used, as the DNA is not detectable without some kind of dye or combination of dye and UV light. These dyes can either be incorporated into the gel matrix itself or introduced during a staining procedure after electrophoresis. Together, the ladder and visualization reagent allow the size of a DNA fragment to be estimated with a surprising level of precision.

Objectives

The PCR product from the previous lab exercise will be detected using agarose gel electrophoresis. Ethidium bromide (EtBr) will be used as an in-gel dye that must be visualized under ultraviolet light. It works by inserting between planar bases of DNA (intercalating) and, upon exposure to UV light, fluorescing with an orange color. A 1 kilobase DNA ladder will be used as a reference for determining the size of the amplified DNA fragment. Since the size of the AmpR gene is known, a successful PCR can be confirmed by the electrophoresis result.

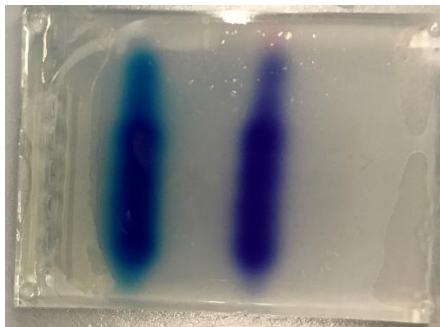
Materials

- Agarose
- TAE buffer
- 1% Ethidium bromide (EtBr) solution
- Electrophoresis cells
- Power supply
- Quick-Load Purple 1 kb DNA ladder
- Loading dye
- PCR products from week 2
- UV gel imaging system

Methods

1. Pour the gel (done as a group under close TA guidance):
 - a) Measure 30 mL of TAE buffer using a graduated cylinder and add it to a 250 mL erlenmeyer flask.
 - b) Using the analytical balance, weigh out 300 mg of agarose on weigh paper.
 - c) Add the agarose to the flask of water and microwave until clear, swirling occasionally. Total microwave time should not exceed 1 minute.
 - d) Add 1 uL of a 1% ethidium bromide solution to the warm agarose solution. CAUTION: ethidium bromide is mutagenic and should be handled with extreme caution.
 - e) Insert the comb into the electrophoresis well.
 - f) Pour the mixture into the well and let solidify (approximately 30 minutes).
 - g) Once solidified, fill the chamber with TAE buffer until the top of the gel is completely covered.
 - h) Remove the comb.
2. Retrieve the PCR products from the previous experiment.

3. One group at a time (following instructions given by the TA), proceed to the gel loading station.
4. Load 5 uL of 1 kb DNA ladder into lane 1.
 - a) Using careful pipetting techniques, fill the pipet tip with 5 uL of ladder solution.
 - b) Insert the tip into the well. Be careful not to disrupt the gel or poke the bottom of the chamber.
 - c) Slowly dispense the solution into the well. Unlike normal pipette use, do not depress the plunger to the second stop! This could cause air to force the ladder solution out of the well.
5. Obtain a small square of parafilm.
6. Dispense a 5 uL drop of the loading dye onto the parafilm.
7. Dispense 5 uL of the PCR product onto the parafilm and mix it with the drop of loading dye.
8. Load 5 uL of the mixture of DNA/loading dye into the appropriate lane.
9. Repeat steps 5-8 for the negative control sample.
10. Connect the electrodes of the electrophoresis chamber to the power supply using the included cables.
11. Run the gel at 100 V for 45 minutes to 1 hour, or until the purple loading dye migrates roughly 60% across the gel (see picture below).

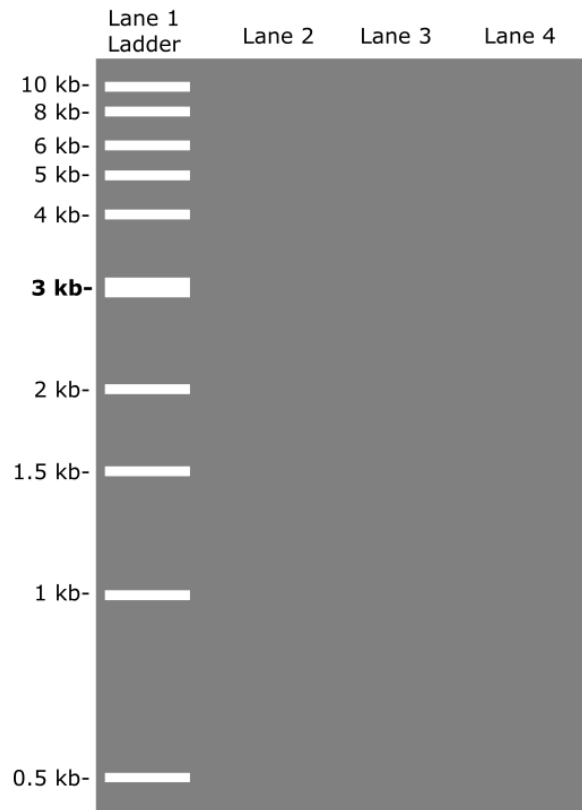


12. While the gel is running, answer the questions at the end of the procedure.
13. Once the gel has finished running, carefully remove it from the chamber while wearing gloves.

14. Use the UV lamp to visualize the gel.
15. Take a picture of the gel under the UV lamp, such that the bands are clearly visible in the image. Take care to avoid contamination by disposing of gloves before touching cell phones.
16. Complete the questions on the next page.

Questions

1. The Quick-Load Purple 1 kb DNA ladder is shown below in lane 1 with all the bands labelled. (Notice that the 3 kb band is intentionally more intense than the other bands. This will aid in the correct annotation of the ladder bands in the experimental gel.) In lane two, draw the expected band(s) for the PCR product(s) of the AmpR gene amplification from pUC19. In lane 3, draw the expected band(s) for the PCR product(s) of the negative control sample.



2. Imagine that you performed this PCR experiment and discovered, to your dismay, that your primers were not specific enough. Looking back at the sequence of the entire plasmid, you notice that there is a sequence identical to your forward primer and another identical to your reverse primer with 520 bases in between. In lane 4 of the image above, draw the expected band(s) for the product(s) of the poorly-designed PCR.

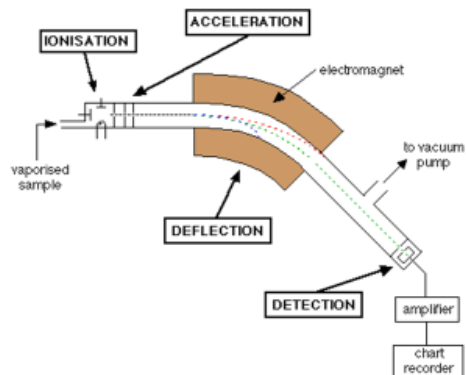
Week 4: Polypeptides: Sequencing with mass spectrometry

Pre-lab material

Peptide Sequencing with MS/MS

What is Mass Spectrometry?

- Analytical tool that detects individual molecules by their mass-to-charge ratio (m/z)
- Works on the basic principle of using electric and/or magnetic fields to move charged particles
- Composed of 3 basic parts:
 - Something to give the molecules charge (ion source)
 - Something to separate the ions by m/z (mass analyzer)
 - Something to record the results (detector)



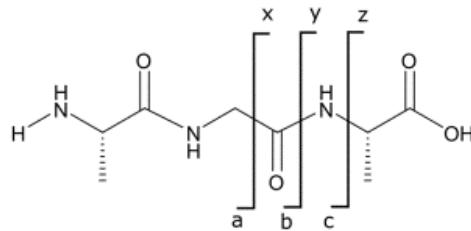
What is MS/MS?

- The best way to understand what something is made of? Smash it and look at the pieces. Do this using 2 mass analyzers.

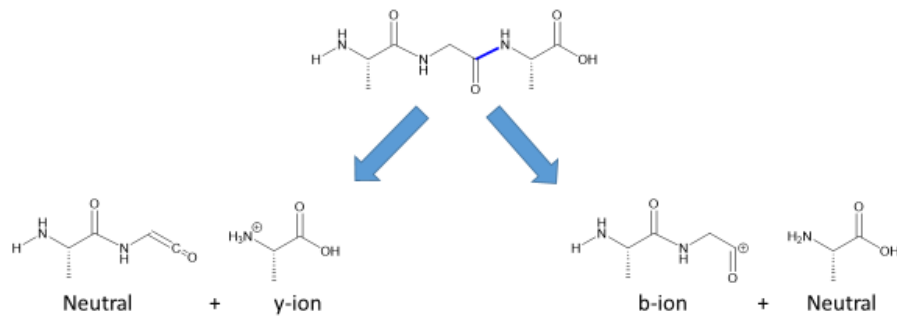


How does a peptide fragment?

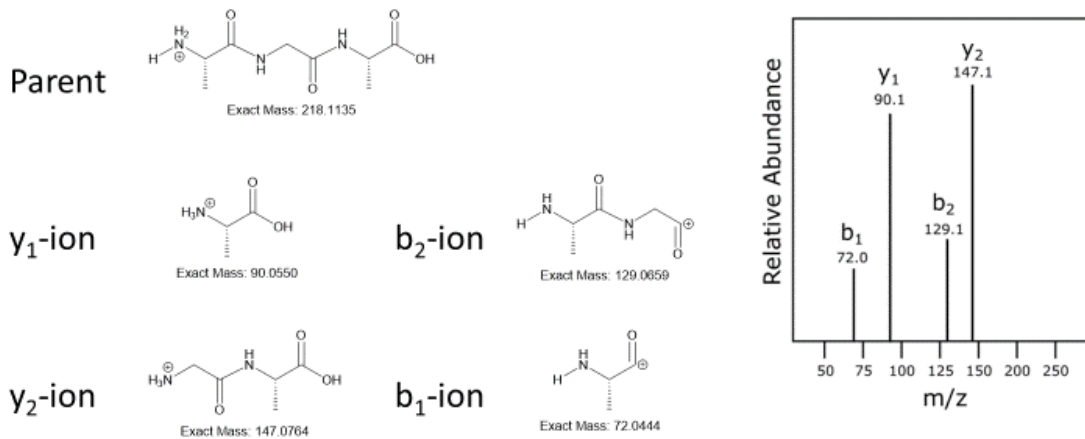
- Along the backbone (3 options)
- b-type and y-type ions are the most common in collisional dissociation (CID/CAD). CID/CAD is the most common fragmentation mechanism.



What about charge?



Putting the pieces together



Required reading

Mass spectrometry of peptides and proteins

Vicki H. Wysocki et. al.

Elsevier Methods

August 2004

DOI: 10.1016/j.ymeth.2004.08.013

Introduction

Tandem mass spectrometry (MS/MS) can be used to elucidate the sequence of proteins and peptides, due to the predictable fragmentation of their amide backbone. Large proteins are commonly digested into peptides before analysis, so that each peptide can be fully sequenced (termed “bottom-up” proteomics). Most complex proteins require search algorithms and reference databases for efficient sequencing. However, in simpler cases, MS/MS spectra of peptides can be manually annotated for effective sequence elucidation. This is accomplished by observing the difference between prominent peaks in MS/MS spectra of peptides and correlating those differences to the masses of individual amino acid residues. Additionally, MS/MS sequencing can illuminate post-translational modifications (PTMs) of known peptides by highlighting residues with a mass shift.

Goals

Initially, the expected b-type and y-type ions of a known peptide will be predicted. Once this is accomplished, an unknown peptide will be identified from an MS/MS spectrum using the manual annotation method. Lastly, a mass spectrum of a known sequence containing a PTM on one or more residues will be analyzed. Both the identity and the location of the modification(s) will be determined from the MS/MS data.

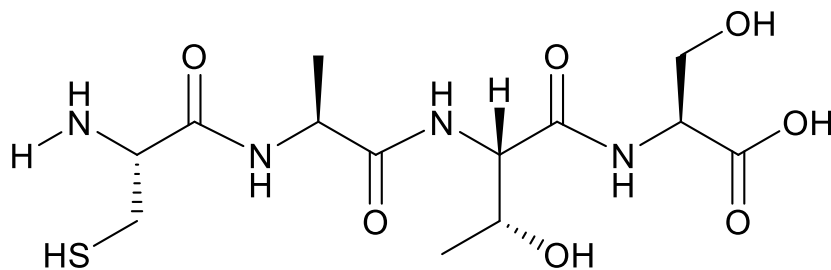
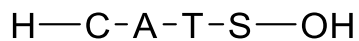
Materials

- Spectra packets
- Reference packets containing lists of residue masses, as well as masses of PTMs and possible sites of modification
- List of strategies for determining sequence from MS/MS spectra
- Calculators
- Pens/pencils

Strategies for peptide sequencing with MS/MS

1. In general, the most commonly used mass spectrometers for peptidomics create b-type and y-type ions. Y-type ions are usually more abundant than b-type.
2. Estimate the number of residues by dividing the parent mass by the average mass of an amino acid residue (110 Da). This is an approximate calculation, and will be subject to some error based on the peptide in question. For example, if the parent mass of the peptide is 1023.5 Da, then there will be $1023.5/110$ or 9.3 residues (8-10 residues, to be safe).
3. Identify the y-type ion with the largest m/z and subtract from the parent mass. This will give the mass of the N-terminal residue. (Do not forget to account for the proton).
4. Identify subsequent y-type ions and determine the m/z difference between each successive peak.
5. Using the chart of masses corresponding to amino acid residues, assign an amino acid to each difference in m/z between y-ions.
6. Find the b-type ion with the smallest m/z. This should give you the mass of the N-terminal residue.
7. Continue through the spectrum and find the difference in m/z between each subsequent b-type ion.
8. Using the chart of masses corresponding to amino acid residues, assign an amino acid to each difference in m/z between b-ions.

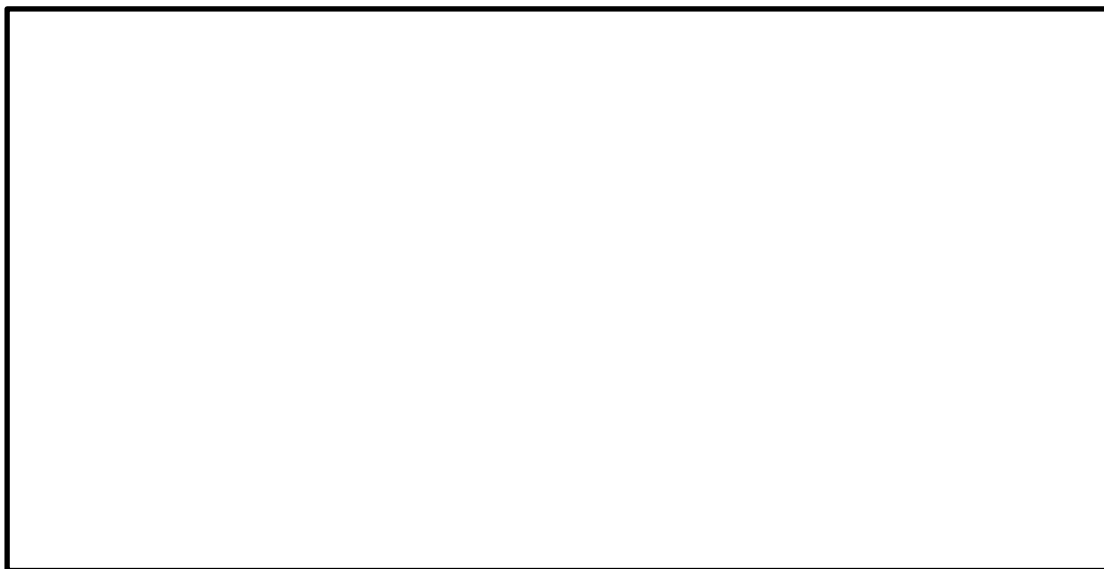
Problem 1: Consider the following peptide. On the sequence, annotate and label all possible b-type and y-type ions. In the spaces below, write the sequences of the b-type and y-type ions and calculate their masses.



y-ions:

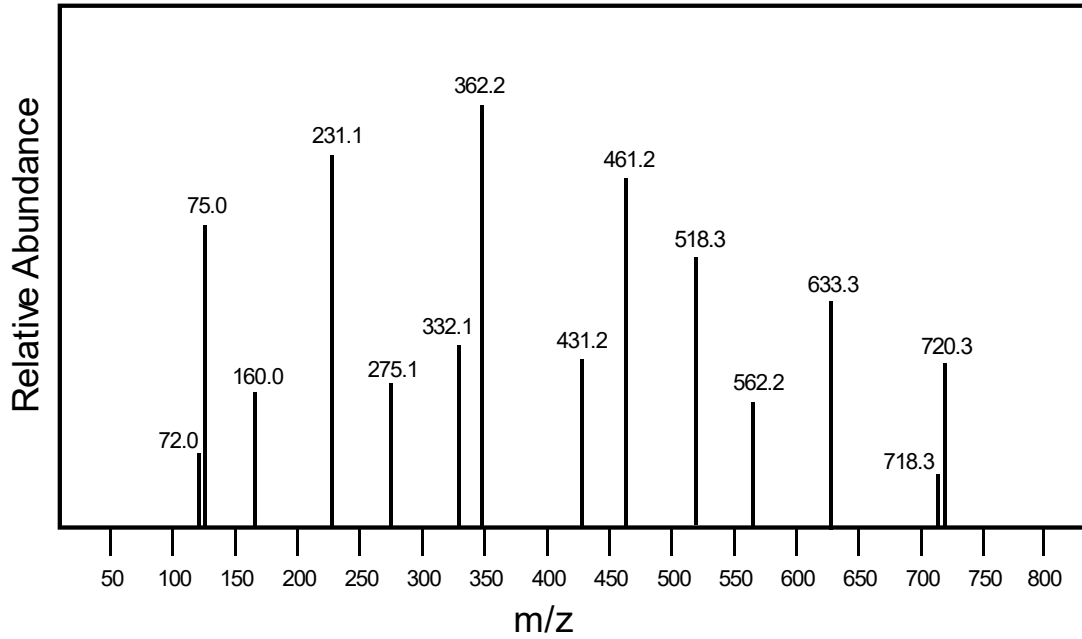
b-ions:

Problem 2: Based on these b-type and y-type ions from problem 1, draw the expected MS/MS spectrum of the peptide. Make sure to include all axis labels. Remember to consider the intensity of the ions.



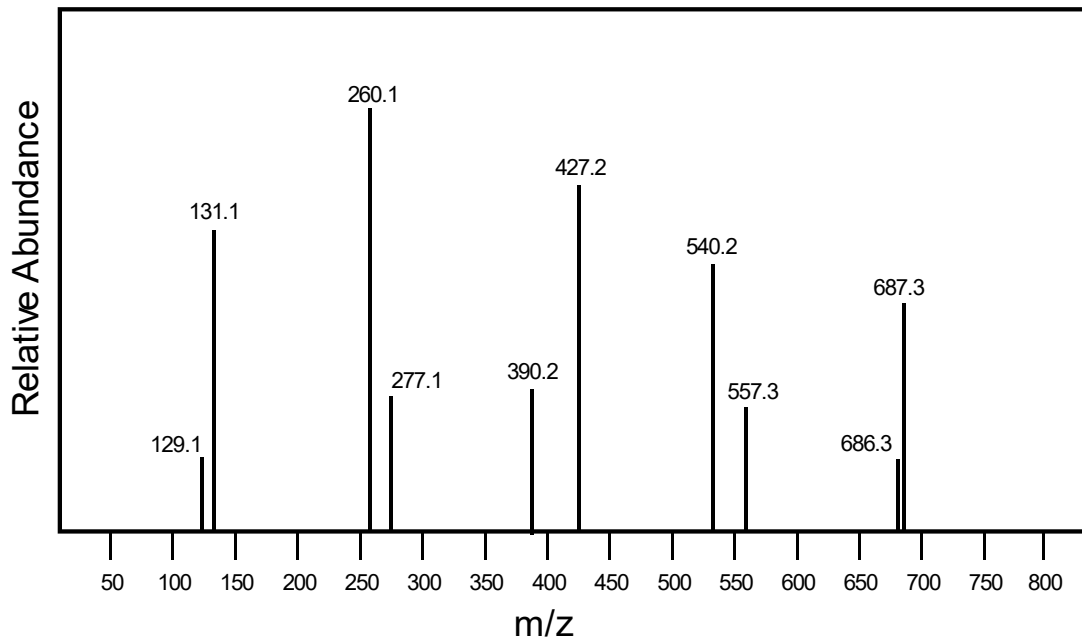
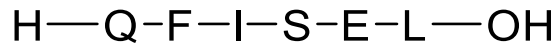
What are the mass and structure of the parent (molecular) ion for the peptide?

Problem 3: Below is the MS/MS spectrum of an unknown peptide with an $[M+H]$ molecular ion of 792.37 Da. Annotate the spectra and determine the sequence. Show all work. Clearly label the b-type and y-type ions.



Sequence: _____

Problem 4: Consider the following peptide sequence. It has undergone a post-translational modification, the location and identity of which are unknown. Using the experimental MS/MS spectra below, determine the location and identity of the post-translational modification.



Modification: _____

Location of Modification: _____

Tables

Amino Acid Masses*		
1-letter code	3-letter	Mass of residue
A	Ala	71.0
R	Arg	156.1
N	Asn	114.0
D	Asp	115.0
C	Cys	103.0
E	Glu	129.0
Q	Gln	128.1
G	Gly	57.0
H	His	137.1
I	Ile	113.1
L	Leu	113.1
K	Lys	128.1
M	Met	131.0
F	Phe	147.1
P	Pro	97.1
S	Ser	87.0
T	Thr	101.0
W	Trp	186.1
Y	Tyr	163.1
V	Val	99.1

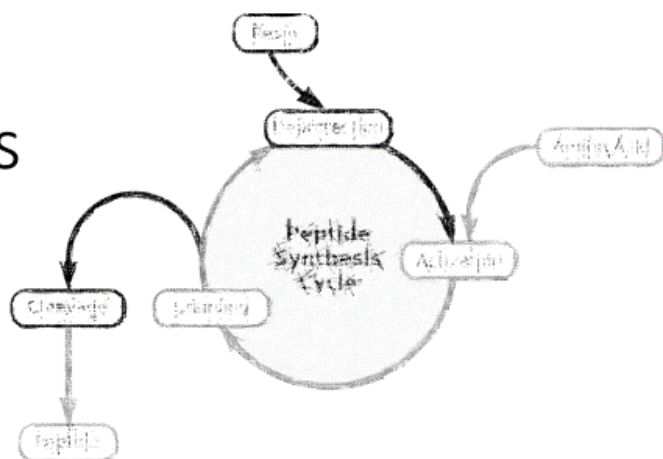
***Note that these are the masses of the internal residues, not the terminal ones.**

Common Protein Modifications		
Modification	Affected Residues	Mass Shift
Phosphorylation	Ser (common), Thr, Tyr (rarely)	+ 80 Da
Sulfation	Tyr	+ 80 Da
Methylation	Lys	+ 14 Da
Hydroxylation	Lys, Pro	+ 16 Da
Acetylation	Primary amines (Lys)	+ 42 Da

Week 5 and 6: Polypeptides: Synthesis of fluorogenic tripeptides

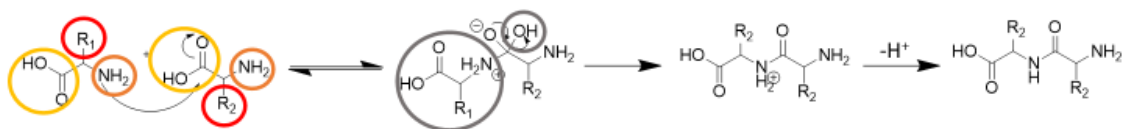
Pre-lab material

Peptide Synthesis



Repetitive Chemistry

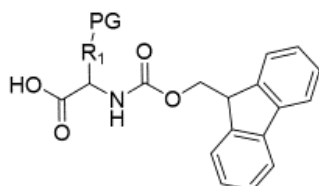
- Successive amide couplings to grow the length of a peptide



- **PROBLEM:** Unwanted reactions
- **PROBLEM:** Bad leaving group = equilibrium with starting material = slow & low yielding reaction

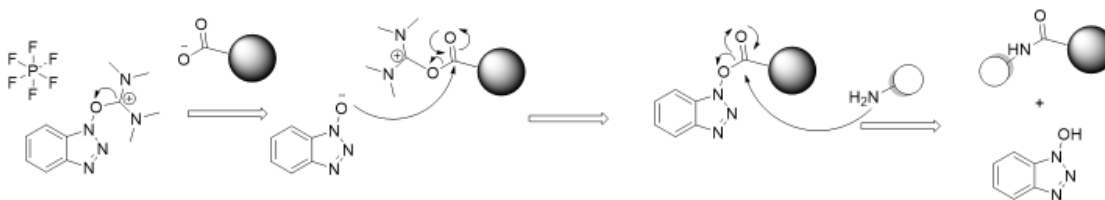
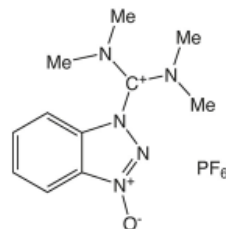
Protecting Groups: The solution to unwanted reactions

- N-terminus protection with fluorenylmethoxycarbonyl (Fmoc) group
 - Allows for specific coupling by blocking the extra N-terminus
 - Deprotection with base (piperidine, 20 min)
- Side-chain protection with various groups
 - Depends on the functional group within the side chain
 - Must have orthogonal chemistry to Fmoc group (deprotect with acid)



Coupling Reagents: The solution to slow reactions

- 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU)
- See also: HATU
- Allows for complete coupling in ~30 min



Problem #3: Technical cost

- Solution-phase chemistry requires cleanup after every step.
 - Perform extraction (~10-15 minutes, use of solvents)
 - Evaporate solvent (~10-15 minutes, requires rotary evaporator)
 - Purify by flash chromatography (~2+ hours, ~1+ L of solvents, ~100-500 g silica)
- This makes the overall synthesis of a peptide much less cost and time effective.
- The answer?

Solid-Phase Peptide Synthesis (SPPS)

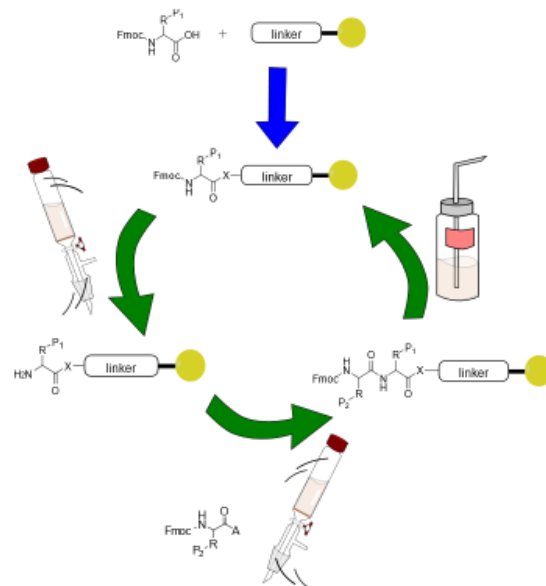
- Pioneered by Robert B. Merrifield (who won the Nobel prize for it)
- The basic concept? Anchor the peptide to polymer beads and build from C-terminus to N-terminus
- This allows the peptide to be purified by a quick rinse over a filter after each step.
 - Much faster cleanup (3 minutes per reaction)
 - Much more cost effective cleanup (5-10 mL of solvents per reaction)
- 3 min, \$ vs 3 hours, \$\$\$

SPPS: How it works

1. Anchor first amino acid to resin via linker (Time varies)
2. Deprotect N-terminus of the first amino acid (20 min)
3. Rinse (3 min)
4. Couple with second amino acid (30 min)
5. Rinse (3 min)
6. Deprotect N-terminus of anchored peptide (20 min)
7. Rinse (3 min)
8. Couple with next amino acid (30 min)
9. Rinse (3 min)

·
·
·

Cleave complete peptide from resin while simultaneously deprotecting side chains

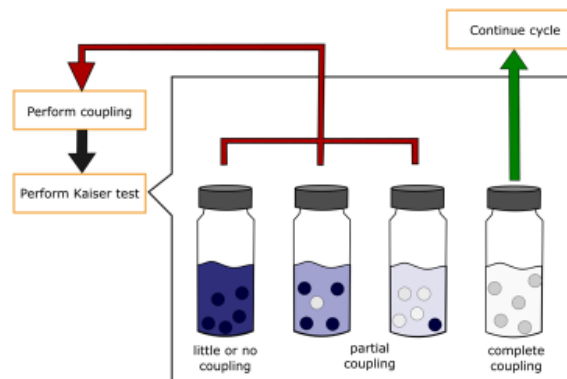


Monitoring the synthesis: Kaiser Test

- Ninhydrin based qualitative staining.
- Blue = free primary amines

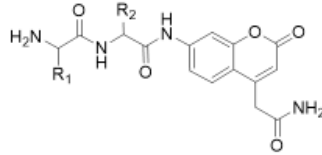
• Test

1. 3-4 drops solution A
2. 3-4 drops solution B
3. 3-4 drops solution C
4. Heat for 5 min at 90 – 110 °C

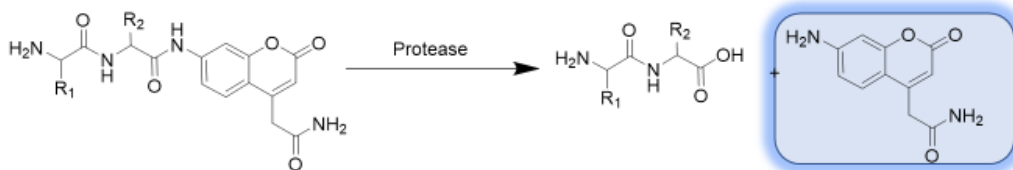


Our synthetic targets

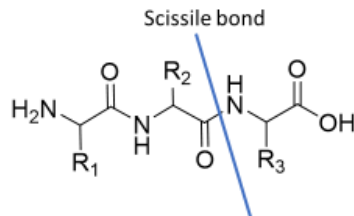
- Modified peptides w/ C-terminal 7-amino-4-carbamoyl methylcoumarin (ACC)



- Why? ACC will fluoresce when cleaved from peptide. Thus, it can be used to determine the specificity of enzymes that cleave peptide bonds (peptidases/proteases).



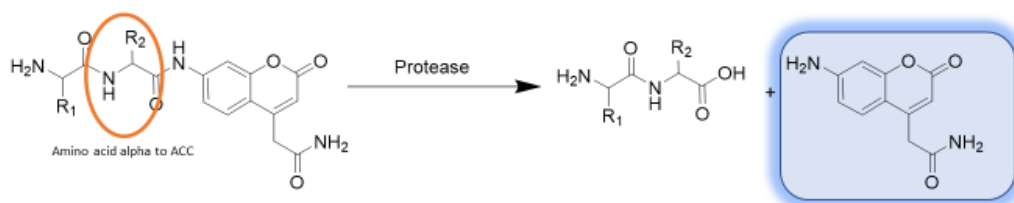
The proteases



- Trypsin
 - Cleaves the amide bond at the c-terminal side of arginine and lysine residues (R₂ = R, K).
- α Chymotrypsin
 - Cleaves the amide bond at c-terminal end of tyrosine, tryptophan, and phenylalanine residues (R₂ = Y, W, F).

The goal:

- Distinguish between trypsin, chymotrypsin, and the absence of either using the fluorescence of peptide-ACC substrates
 - Use synthetic peptides which have either lysine (sensitive to trypsin), tyrosine (sensitive to chymotrypsin), or alanine (inert) in the position alpha to the ACC moiety
 - Cleavage of the scissile bond by the protease will liberate the ACC and yield fluorescence



Required reading

Methods and Protocols of Modern Solid Phase Peptide Synthesis

Muriel Amblard et. al.

Molecular Biotechnology

July 2006

DOI: <https://doi.org/10.1385/MB:33:3:239>

Introduction

Modified peptides containing coumarin moieties have previously been employed as a means of probing protease specificity through simple fluorescence assays. A common approach is to test protease activity on small peptides bearing c-terminal 7-amino-4-carbamoyl methylcoumarin (ACC) moieties. Cleavage of the appropriate peptide bond releases a free ACC which can be detected through basic excitation /emission fluorescence. By testing two peptidases against several peptide-ACC substrates, the specificity of proteolytic cleavage can be ascertained. In order to facilitate this type of assay, modified peptides must be synthesized. The most common reliable means of accomplishing this is solid-phase peptide synthesis (SPPS), a technique for which the Nobel prize was awarded to Robert Bruce Merrifield in 1984. This approach begins by anchoring the C-terminal amino acid to a polymer resin and elongating through successive amidation

Goals

Tripeptide analogs with C-terminal ACC moieties will be synthesized with SPPS using an ACC-modified resin. These will be designed to strategically serve as substrates for the proteases trypsin and chymotrypsin by incorporating the appropriate amino acid in position next to the scissile bond (the P1 position). In short, peptides which have either lysine (sensitive to trypsin), tyrosine (sensitive to chymotrypsin), or alanine (inert) in the P1 position relative to the ACC moiety will be made using SPPS. After successful synthesis, these peptides will be used to determine protease activity via ACC fluorescence using a microplate fluorimeter. Fluorescence will be recorded after addition of enzymes trypsin, chymotrypsin, or neither (control) to the synthesized peptide substrates. The change in fluorescence of each well will be monitored and the data will be plotted and analyzed in Microsoft Excel.

Materials

- ACC Resin
- Peptide synthesis vessels

- HBTU (coupling reagent)
- Diisopropylethylamine (DIPEA)
- Dimethylformamide (DMF)
- Dichloromethane (DCM)
- Teflon stir rods
- Spatulas
- Fmoc-Ala-OH
- Fmoc-Val-OH
- Fmoc-Gly-OH
- Fmoc-Leu-OH
- Kaiser test solution A [5% ninhydrin in ethanol (w/v)]
- Kaiser test solution B [80% phenol in ethanol (w/v)]
- Kaiser test solution C [KCN in pyridine (2 mL 1 mM KCN in 98 mL pyridine)]
- Micropipettes
- ½ Dram to 1 Dram glass vials
- Bath for heating vials (Sand in evaporating dish, or metal beads in a heating mantle)

Schedule

Week 1: Consider the possible proteases, design the two substrates that each group will synthesize, and complete the first week of synthesis.

Week 2: Complete the second week of synthesis of their peptides.

Week 3: Complete fluorescence assay as described in Protease Specificity Lab Hand-out

Design of tripeptide-ACC substrates

- Each student will choose ACC coupled amino acid to begin.
 - The options are Ala-ACC, Lys-ACC, and Tyr-ACC.

- Then, each group chooses two more “inert” residues to couple, either alanine, glycine, leucine, or valine.

- Examples:

H-Ala-Val-**Ala-ACC-NH₂**

H-Leu-Gly-**Lys-ACC-NH₂**

H-Val-Leu-**Tyr-ACC-NH₂**

- Record your choices and make note of how many grams to use and how these choices affect your synthesis timeline.

Draw the chemical structure of your tripeptide-ACC final product and answer the following questions:

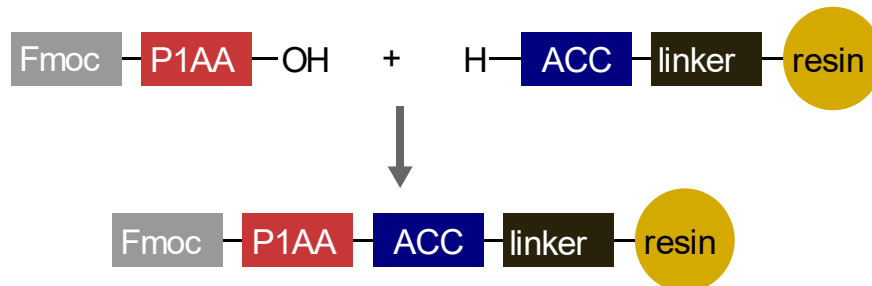
- 1) Which protease, if either, will cleave your tripeptide?
- 2) At what bond will the protease enzyme cut?

Proteases used in lab and their specificities

Protease	Cleavage Site
Trypsin	C-Terminal side of Lysine and Arginine
Chymotrypsin	C-Terminal side of Tyrosine, Tryptophan, and Phenylalanine

Experiment Synthesis Flowchart

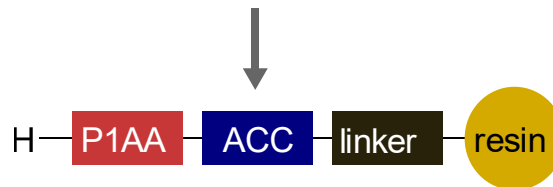
Stockroom Preparation: Coupled First Fmoc Protected Amino Acid to Resin. P1AA = Ala, Lys, or Tyr



Experiments in Lab:

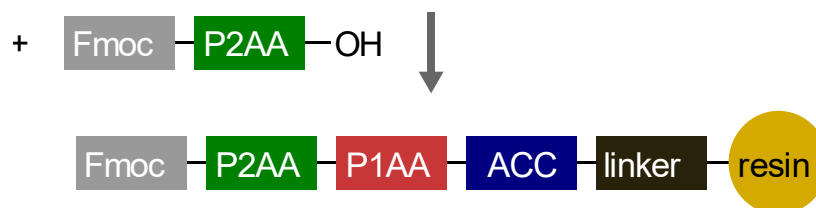
Week 1:

- 1) Swell Resin
- 2) Deprotect



- 3) Couple Second Fmoc Protected Amino Acid to Resin

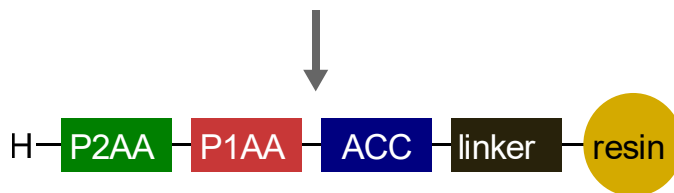
P2AA = Val, Ala, Gly, or Met



- 4) Run Kaiser Test. If bad results, go back to step 3 and redo coupling. If good results, continue.
- 5) Dry Resin

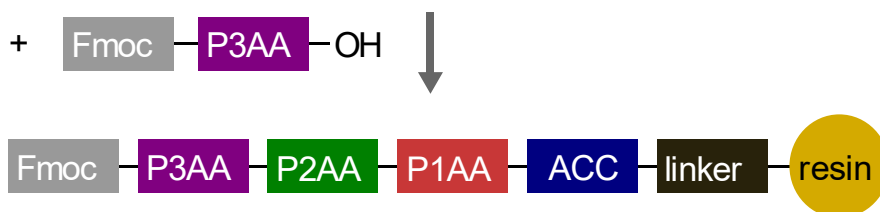
Week 2:

- 1) Swell Resin
- 2) Deprotect

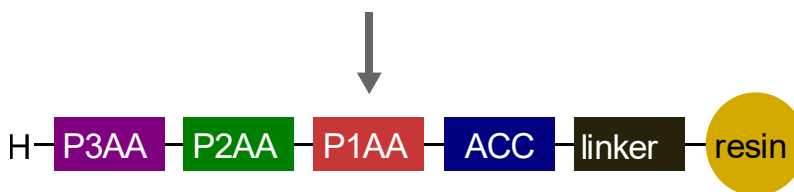


- 3) Couple Third Fmoc Protected Amino Acid to Resin

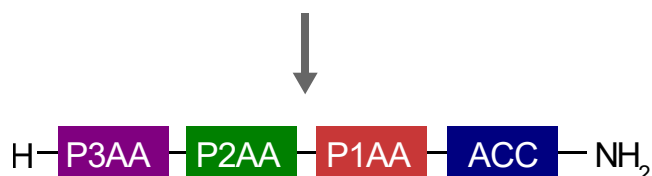
P3AA = Val, Ala, Gly, or Met



- 4) Run Kaiser Test. If bad results, go back to step 3 and redo coupling.
If good results, continue.
- 5) Final Fmoc Deprotection



- 6) Cleavage of Amino Acid protecting groups and linker/resin.



Peptide synthesis steps

Swelling Resin (begin with 150 mg of resin on week 1):

1. Add DCM to the peptide vessel until the resin is barely covered
2. Let the resin sit for 10-15 minutes so it swells to maximal surface area

Deprotection:

1. Fill peptide vessel $\frac{1}{4}$ full with a 20% piperidine/DMF solution
2. Mix carefully to ensure that there are no clumps of beads (to facilitate access to all bead surfaces)
3. Shake/stir for 1 minute, and drain the solution
4. Fill the peptide vessel $\frac{1}{4}$ full with a 20% piperidine/DMF solution
5. Mix carefully to ensure that there are no clumps of beads (to facilitate access to all bead surfaces)
6. Shake intermittently over 15 minutes
7. Drain the solution
8. Rinse 3x with DMF

Coupling:

1. Weigh out the appropriate amount of amino acid and place it in a test tube
2. Weigh out 101 mg of HBTU and place it in the same test tube
3. Add DMF to the test tube and mix until all contents are dissolved
4. Add the activated amino acid to the peptide vessel containing the resin and shake to ensure adequate dispersion of beads
5. Add 94 μ L of DIPEA
6. Shake intermittently over 30 minutes
7. Drain the solution
8. Rinse 3X with DMF

Kaiser test:

1. Remove a few clean beads from the peptide vessel and place in a small vial. Add 3-4 drops of each Kaiser test solution (A, B, and C)
2. Heat at 90 – 110 °C for 5 minutes
3. With adequate lighting, determine if any of the beads maintain a blue color. Kaiser test is a Ninhydrin based qualitative staining. Blue = free primary amines = coupling unsuccessful.

Cleavage and final side-chain deprotection:

1. Dry the resin
2. Add to round-bottom flask equipped with a stir bar
3. Add 5-10 mL of cleavage solution
4. Stir for 90 minutes

Week 7: Polypeptides: Probing protease specificity

Pre-lab material

Review slides from weeks 5 & 6.

Required reading

Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries

Jennifer L. Harris, et. al.

Proceedings of the National Academy of Sciences (PNAS)

July 2000

DOI: <https://doi.org/10.1073/pnas.140132697>

Introduction

Enzymes are powerful biological catalysts that are responsible for the chemical reactions of metabolism, among other functions. Enzymes are folded proteins with specific shapes that are able to bind other molecules and alter them in precise ways. A subset of enzymes called proteases are responsible for digesting proteins into smaller peptides. They accomplish this by cleaving the amide bond between amino acid residues in the protein. Many proteases are specific, and preferentially cleave bonds adjacent to certain amino acids based on the ways they interact with the binding “pocket” of the enzyme. This function will be assessed through the current experiment.

Goal

The Tripeptide-ACC molecules made during previous lab sessions will be employed in this experiment. The cleavage of the peptide-ACC bond liberates the ACC molecule, which fluoresces. When excited at 365 nm, it fluoresces with an emission wavelength of 445 nm. Microplate fluorimeters can constantly excite samples at a given wavelength, and monitor emission over time to give quantitative data. Using this instrument and the peptide-ACC substrates, it is possible to differentiate proteases based on their cleavage sites. The synthetic peptides have either lysine (sensitive to trypsin), tyrosine (sensitive to chymotrypsin), or alanine (inert) in the P1 position relative to the ACC moiety. Fluorescence will be recorded after addition of enzymes trypsin, chymotrypsin, and the absence of either using the fluorescence of peptide-ACC substrates. These experiments should be run at saturating substrate concentrations to give maximum rates, allowing the results to be observed within a short minute time span.

Materials

- Microplate fluorimeter
- 96-well plates
- Synthetic tripeptide-ACC substrates from previous experiment
- DMSO
- Buffer solution

- Buffered solution of trypsin
- Buffered solution of Chymotrypsin
- ACC standard
- Micropipettes
- Microsoft Excel

Methods

A. Calculations

Begin by creating stock solutions of the synthetic tripeptide-ACC substrates in DMSO. Answer the following questions to help guide you through the dilution calculations. Have TA check calculations before creating the solution.

- a. Write the abbreviated name and chemical structure of the tripeptide-ACC molecules your group designed and synthesized.
- b. What is the molecular weight of your tripeptide-ACC compound? Tip: an online peptide mass calculator can accelerate the process.
- c. Based on the mass of tripeptide-ACC you synthesized, how many moles are present?
- d. Using this amount of peptide, what total volume of solution is required to yield 10 mM tripeptide-ACC in DMSO?

Using these concentrations as the stock solutions, calculate the volume needed to aliquot into each well. Hint: Use the dilution equation $M_1V_1=M_2V_2$

- e. How many μL of the 10 mM stock solution and how many μL of buffer/protease mixture should you add to each well to yield 500 μM tripeptide-ACC in each 100 μL well?

μL of 10 μM stock: _____

μL of buffer solution: _____

B. Fluorescence assay

- Fill each well with a total volume of 100 μL .
- To each well, add 95 μL buffer/protease mixture.
- To each well, add 5 μL of the 10mM tripeptide-ACC in DMSO stock solution, such that the total concentration of substrate is 500 μM , as calculated previously.
- In addition to the wells containing both enzyme and substrate, prepare wells for each substrate without enzyme, so that any uncatalyzed hydrolysis of the tripeptides can be monitored. Also prepare wells of only ACC standard, only DMSO solution, only enzyme, and only tripeptide-ACC to be used as controls.
- Fluorescence (excitation, 360 nm; emission, 445 nm) will be measured at 25° C for 21 readings (approximately 1 minute each).

C. Data analysis

Using Microsoft Excel, analyze data for all Tripeptide-ACC variants and enzyme conditions, (not just the two peptides designed by your group).

- Once the excel data has been distributed, compare the cells and well assignments on the spreadsheet to the layout of the well plate used in lab.

- b. Since samples were run in triplicate, average the triplicates to get the mean value for each time point. It is important to note that there are two versions of each Tripeptide-ACC type.
- c. Create fluorescence (y) vs time (x) scatterplots from the fluorescence data for each protease. Graph the fluorescence results from all 3 Tripeptide-ACC versions (P1-Ala-ACC, P1-Lys-ACC, P1-Tyr-ACC) treated with the same protease and the blank on the same graph.
- d. Include legends to identify the data, graph titles, and axis titles with units. Since the values plotted are mean readings, include custom error bars for the min (difference between the mean and smallest reading in the set) and max (difference between the mean and the largest value in the set) for each time point.

Week 8: Connecting carbohydrate structure to reactivity

Pre-lab material

Qualitative Analysis of Carbohydrates

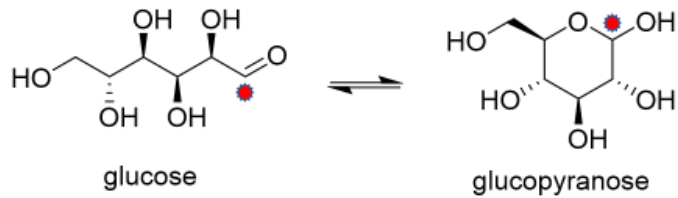


Carbohydrates

- Generally contain H:O ratio of 2:1
- Play important roles in cell structure, energy storage, immune system, fertilization, blood clotting, and composition of *ribonucleic acids* (among others).
- Includes sugars, starches, cellulose, glycogen, etc...

Monosaccharides

- Are polyols with a carbonyl group at anomeric carbon
- Have a linear structure and cyclic structure that are in equilibrium.



Ketoses vs Aldoses

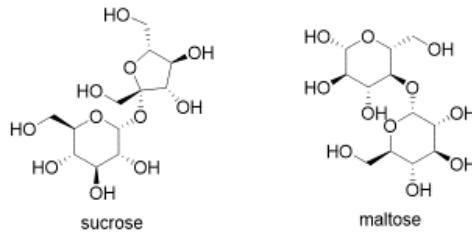
- Ketoses have a ketone in the linear form and a hemiketal in the cyclic form
- Aldoses have an aldehyde in the linear form and a hemiacetal in the cyclic form



Disaccharides and polysaccharides

- Dimers of monosaccharides

- Can be two of the same monomers (maltose) or two different ones (sucrose)

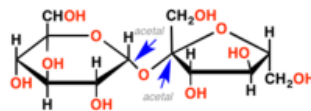


- Polysaccharides are larger oligomers of monosaccharides

Reducing sugars

- Any sugar that is capable of acting as a reducing agent.
- Must have an accessible ketone or aldehyde to act as a reducing agent.
 - Thus, all monosaccharides are reducing sugars.
 - Disaccharides and polysaccharides can be reducing sugars, but it depends on connectivity

Sucrose is a non-reducing sugar



The structure lacks any hemiacetal functional groups and is therefore "locked" in its cyclic form

Objectives

- Use qualitative chemical tests to characterize carbohydrates.
 - Main descriptors:
 - Monosaccharide vs. disaccharides vs. Polysaccharide
 - Ketose vs aldose
 - Reducing vs Nonreducing sugar
- Collect observations using known carbohydrates, and use the information you gather to help identify the two unknown solutions.

Required reading

A Reagent for the Detection of Reducing Sugars

Stanley R. Benedict

Journal of Biological Chemistry

January 1909

PMID: 11953443

Optional references

The Composition of Reducing Sugars in Solution

Stephen Angyal

Advances in Carbohydrate Chemistry and Biochemistry

1984

DOI: [https://doi.org/10.1016/S0065-2318\(08\)60122-5](https://doi.org/10.1016/S0065-2318(08)60122-5)

Introduction

Carbohydrates assume many central biochemical functions including structural support of DNA and RNA, composition of coenzymes, and storage of energy. Monosaccharides, the single sugar units, can be polymerized to create larger saccharides ranging from disaccharides to high molecular weight polysaccharides. Chemical tests can be used to broadly categorize and differentiate carbohydrates based on structural differences. These tests are often colorimetric—qualitative tests based on a change in color. For example, Fehling's test is a classic colorimetric test which produces a red-brown cuprous oxide precipitate in the presence of aldehydes. Thus, it can be used to quickly differentiate reducing sugars from nonreducing sugars.

Objectives

The purpose of this experiment is to observe the outcomes of various chemical tests on carbohydrate substrates and employ those tests to determine the identity of two unknown carbohydrate samples. Initially, each of the known saccharide compounds will be tested with each qualitative test, and observations should be recorded appropriately. Observations should be related to the structure of the saccharide tested. Once a familiarity with the tests is established, two unknowns will be tested and identified as specifically as the data allows.

Materials

- Saccharide knowns: 1 M glucose, 1 M fructose, 1 M maltose, 1 M sucrose, 1 M ribose, and soluble starch
- Molisch's reagent (10% α -naphthol in ethanol)
- Concentrated sulfuric acid
- Fehling's solution A (aqueous CuSO_4)
- Fehling's solution B (aqueous solution of potassium sodium tartrate)
- Barfoed's reagent (cupric acetate in acetic acid)
- Seliwanoff's reagent (resorcinol in $\text{HCl}/\text{H}_2\text{O}$)
- Bial's reagent (Orcinol, HCl , ferric chloride)
- Concentrated HCl
- Litmus paper
- 1 M NaOH

- Iodine solution
- Hot water bath (150 mL beaker half full of water on hotplate with 2 boiling chips)
- Unknown carbohydrates
- Small test tubes and rack
- Pasteur pipettes
- Rubber pipet bulbs
- Graduated cylinder

Methods

1. Obtain approximately 10-15 mL of each known saccharide from the dispensing area.
2. Label clean, dry test tubes and place them in the test tube rack.
3. Perform each of the following tests on each of the unknowns, as indicated in the data table. NOTE: exceptions are marked with an "X", and do not need to be performed
4. It is crucial to perform a control for each test. This should be done by using water instead of the saccharide solutions for each test. Use this control as a comparison for each colorimetric test.
5. After testing all the known compounds, acquire 2 unknown saccharide solutions to identify.
6. Perform all necessary tests on the unknown compounds. It is not necessary to perform every test, but make sure to collect enough data to have ample evidence for the identification of each unknown.
7. Answer all discussion questions.

Molisch's Test:

A sensitive test for all carbohydrates (and some compounds functionalized with carbohydrates), Molisch's test is a tandem acid-catalyzed dehydration followed by condensation of the resulting aldehyde with naphthol to produce a red to purple color.

NOTE: The appearance of a brown color is due to charring. In the case of charring, repeat the test with more dilute solutions of saccharide knowns/unknowns.

1. Add 1 mL DI water and 5 drops of known sugar solution to each respective test tubes.
2. Add 1 drop of Molisch's reagent.
3. Carefully pour 1 mL of concentrated sulfuric acid (CAUTION: corrosive) down the side of the tube. It should form a layer at the bottom of the tube.
4. Observe the color at the interface between the layers, and record results.

Fehling's Test:

Together, Fehling's solutions A and B react with reducing sugars to produce a red-brown precipitate of copper(I) oxide. The sugars are oxidized in the process of reducing Cu^{2+} to Cu^+ , acting as a reducing agent like their names suggest.

1. Add approximately 1 mL of solution B to 1 mL of solution A in a small test tube.
2. To this mixture, add 1 mL DI water and 1 mL of each sugar solution.
3. Observe the reaction for a few minutes, then record the results in the data table.

Barfoed's Test:

Barfoed's reagent is balanced so that it can only be reduced by monosaccharides. Some disaccharides may also be reduced, but the reaction takes much longer. The reaction reduces copper (II) acetate to copper(I) oxide. Since the product is the same as Fehling's test, a positive reaction should produce a brick red precipitate.

1. Add 2 mL of Barfoed's reagent, 1 mL DI water, and 1 mL sugar solution to each test tube.
2. Heat in a water bath (100° C) for 5 minutes and let stand.
3. Observe the color and record.

Seliwanoff's test:

Seliwanoff's test is a simple way of distinguishing aldoses from ketoses. It relies on the fact that ketoses more rapidly dehydrate than aldoses when heated. Ketoses should quickly produce a red color, while aldoses may slowly form a lighter pink hue.

1. Add 1 mL of saccharide solution to 3 mL Seliwanoff's reagent.
2. Heat in a water bath (100° C).
3. Ketoses will produce the red color within 1 minute. If the red color does not appear, continue heating for another 1-2 minutes, or until it turns pink.

4. Make observations and record them in the data table.

Bial's Test:

Pentoses can be identified with Bial's test, which relies on the dehydration of pentoses to dehydrate to furfural and then react with orcinol. A positive test will result in a blue color, and potentially a precipitate.

1. Add 2 or 3 drops of saccharide solution to 5 mL Bial's reagent.
2. Heat in a water bath (100° C) for approximately 1 minute.
3. Record observations.

Inversion of Sucrose:

Sucrose, a disaccharide, can be hydrolyzed into a mixture of its monosaccharide constituents. This hydrolysis can be accomplished with only water and heat but is accelerated by acid. Once inverted, the mixture of monosaccharides can be subjected to other colorimetric tests.

1. Obtain clean test tubes and add 1 mL DI water and 2 mL sugar solution to each.
2. Add 3 drops of concentrated HCl.
3. Heat in a water bath (100° C) for 10 minutes.
4. Cool solution and add 1 M NaOH until neutral. Monitor the pH with litmus paper.
5. Perform Barfoed's test on the resulting solutions.
6. Record results.

Iodine Test:

Iodine can indicate the presence of starch (polysaccharides) by producing a blue-black color.

1. Add 1-2 drops of iodine solution to 2-3 mL of saccharide solution.
2. Observe any color changes and record findings.

Observations

Perform these →	Glucose	Fructose	Maltose	Sucrose	Ribose	Starch
Molisch's Test						
Fehling's Test						
Barfoed's Test						
Seliwanoff's Test						✗
Bial's Test			✗	✗		✗
Inversion of Sucrose	✗	✗			✗	✗
Iodine Test	✗		✗		✗	

Unknowns

Unknown:		
Molisch's Test		
Fehling's Test		
Barfoed's Test		
Seliwanoff's Test		
Bial's Test		
Inversion of Sucrose		
Iodine Test		

Questions

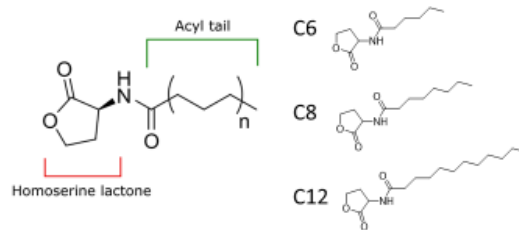
1. What are the identities of your two unknowns? Make sure to explain your reasoning and to avoid making assumptions. Be as specific as possible based on the data.
2. Draw the chemical structure of maltose and use it to explain the result that maltose gave for Fehling's test.
3. Could sucrose be differentiated from lactose based on these colorimetric tests alone? Why or why not?
4. In the required reading, you learned about Benedict's test for reducing sugars. How is Benedict's test different from Fehling's test and what are the similarities?

Week 9 and 10a: Chemical signaling: Synthesis of AHLs

Pre-lab material

Synthesis of AHLs: chemical signaling week 1

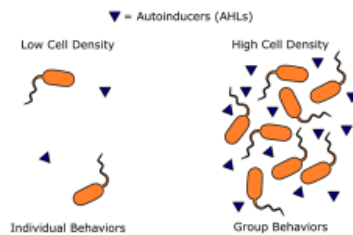
N-acyl-L-homoserine lactones



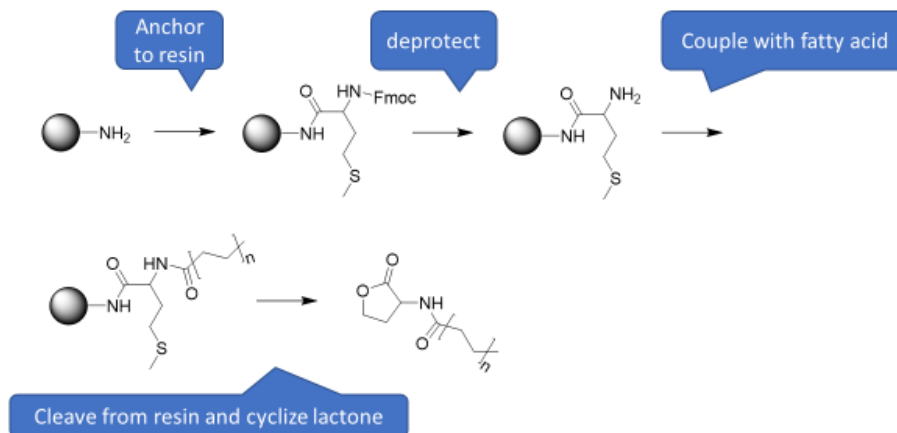
- Produced by many species of bacteria as signaling molecules.
- Signal gene expression that can effect various pathways and phenotypes
 - Dye production, biofilm formation, bioluminescence, etc...

Quorum sensing

- Based on a population density threshold
- Once a certain extracellular concentration of autoinducer is achieved, the threshold is reached and the signaling pathways are activated
- These phenotypes are generally expressed this way because they are advantageous for larger populations



The synthesis of AHLs



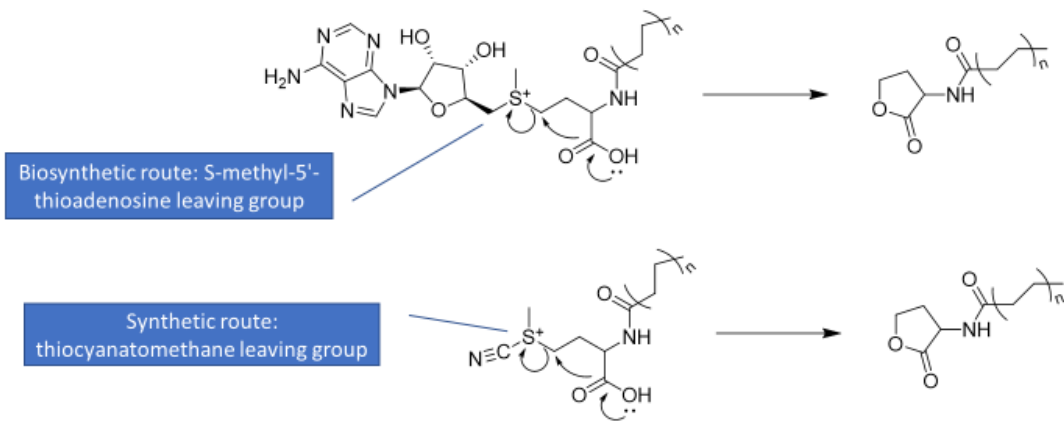
Solid-phase synthesis

- Minimized workup and purification after each step
- Robust, well-developed chemistry
- Shorter, quicker reactions by use of peptide coupling reagents (activating groups)
- More environmentally friendly than traditional synthesis in solution

Microwave-assisted synthesis

- Microwave assistance produces higher yields over shorter reaction times (sometimes hours >>> minutes)
- Microwave synthesis allows rapid optimization
- Microwave heating is more energy efficient on top of being faster

Biomimetic synthesis



Required reading

Structural identification of autoinducer of Photobacterium fischeri luciferase

A. Eberhard et. al.

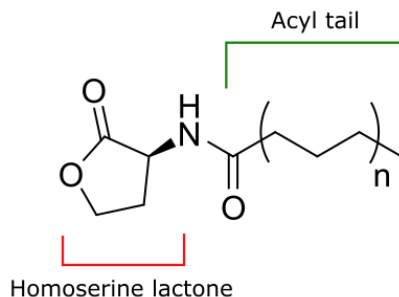
Biochemistry

April 1981

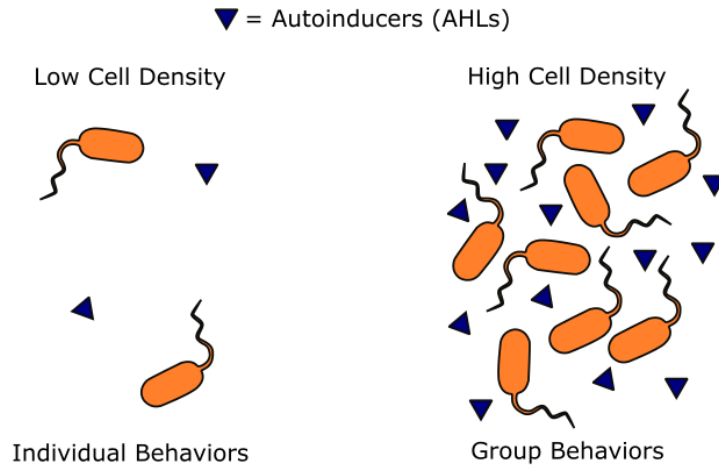
DOI: 10.1021/bi00512a013

Introduction

N-acyl-L-homoserine lactones (AHLs) are a class of signaling molecules involved in various signal transduction pathways across many species of bacterial microorganisms.



They are responsible for signaling many changes in gene expression that can impact various phenotypes including biofilm formation, bioluminescence, and production of pigments. The AHL signaling pathways are an example of quorum sensing, a stimuli/response system that depends on population density. As the bacteria grow, they produce AHL signaling molecules (also called autoinducers) and release them into the environment. Once microbial communities achieve the population density threshold, which is determined by the extracellular concentration of the autoinducers, the pathways are activated and the signaling cascade produces the desired phenotype. Generally, these phenotypes are expressed through quorum sensing because they are advantageous for the larger population. The illustration below demonstrates quorum sensing.



Synthetic AHLs can be used to artificially modulate these phenotypes, and to help study the signaling pathways. The synthesis of the homoserine lactone portion of the AHL can be achieved through the proteolytic action of cyanogen bromide (CNBr) on the methionine residue of a peptide. The acyl tail of the AHLs can be added with traditional amide-forming chemistry identical to those used in solid-phase peptide synthesis.

Objectives

The C8 AHL will be synthesized, and later used to modulate the expression of luminescence in bacterial cultures. The synthesis protocol is unique for two reasons: it features a solid support and uses a microwave synthesizer. The use of a solid support allows the purification steps between reactions to be minimized. Additionally, the use of a microwave synthesizer drastically cuts reaction times and may allow milder conditions. Both techniques also serve to alleviate the environmental impact of the synthesis.

Materials

- Aminomethyl polystyrene resin
- Fmoc-L-methionine
- Chloroform (CHCl₃)
- Diisopropyl carboimide (DIC)
- Hydroxybenzotriazole (HOBT)
- Methylene chloride/dichloromethane (DCM)

- Dimethylformamide (DMF)
- Octanoic acid
- 1.5 M cyanogen bromide (CNBr) in chloroform (cleavage solution A)
- 1% trifluoroacetic acid (TFA) in water (cleavage solution B)
- 25% piperidine in DMF
- Diisopropylethylamine (DIPEA)
- Microwave (MW) synthesis vials
- Filter funnels and filter flasks
- Ethanol (EtOH)
- Magnesium sulfate
- Deuterated chloroform (CDCl₃) for NMRs
- NMR Tube
- Benchtop NMR
- Spatulas
- Micropipette

Methods

1. Weigh 100 mg of aminomethyl resin and place it in a reaction vial.
2. Add a few drops of CHCl₃ (such that the resin is just covered by solvent) and allow the resin to swell for 10 minutes.
3. To a separate round-bottom flask, add approximately 3 mL DMF along with 129 mg Fmoc-L-methionine, 54 mg HOBT, and 73 μ L DIC. Allow activation by stirring for 10 minutes at room temperature.
4. Add the activated solution of Fmoc-L-methionine to the swelled resin.
5. Equip the reaction vial with a stir bar and cap it.
6. Irradiate the reaction according to the following parameters: 10 min at 50 °C.
7. Filter the resin and wash with approximately 5 mL each of DMF, water, EtOH, and DCM and dried *in vacuo*.
8. To deprotect, place the dried resin into a clean MW flask in 1-1.5 mL of 25% piperidine in DMF and irradiate at 150 °C for 6 minutes.

9. Filter the resin and wash with 3 times with 5 ml of DMF.
10. To a clean microwave vial with stir bar, add the dried resin (approx. 100 mg) and swell with a minimal amount of CHCl_3 (approximately 0.5 mL) for 5 min.
11. In a separate vial, prepare a mixture of 0.344 mmol fatty acid (49.6 mg or **54.5 μL** octanoic acid), and 0.5 mmol DIC (78 μL) in 2 mL DMF.
12. Activate by stirring at room temperature for 5 minutes.
13. Add the activated solution of fatty acid to the swelled resin and irradiate (hold for 10 minutes at 50 °C).
14. Filter the resin and wash 2x with 20 mL each of DMF, Water, EtOH, and DCM.
15. Dry the resin *in vacuo* (stop here at the end of week 9).
16. To cleave the AHL from the resin, 3 mL of cleavage solution (5:2 1.5M CNBr to 1% TFA) is added to the resin in a clean MW vial.
17. Irradiate (30 min at 60 °C).
18. Filter the resin and collect the filtrate.
19. Add 5 ml CHCl_3 and wash the organic layer 3x with water 10 mL each). Dry over magnesium sulfate.
20. Concentrate *in vacuo* to yield the AHL as a white powder.
21. Once the product is dry, dissolve a small amount in CDCl_3 and transfer to NMR tube.
22. Record proton NMR spectra using the benchtop 40 MHz NMR.

Week 10b: Chemical signaling: Modulation of luminescence using synthetic quorum sensing autoinducers

Pre-lab material

Required reading

Shedding light on bioluminescence regulation in Vibrio fischeri

Tim Miyashiro et. al.

Molecular Microbiology

June 2012

DOI:10.1111/j.1365-2958.2012.08065.x.

Modeling Analysis of Signal Sensitivity and Specificity by Vibrio fischeri LuxR Variants

Deanna M. Colton et. al.

PLOS One

May 2015

DOI:10.1371/journal.pone.0126474

Introduction

N-acyl-L-homoserine lactones (AHLs) are a class of signaling molecules involved in various signal transduction pathways across many species of bacterial microorganisms. In wild type bioluminescent Gram-negative bacterium *Vibrio fischeri*, sufficiently high bacteria density triggers the expression of a set of genes called the *lux* operon¹⁸⁹. These genes induce bioluminescence when active. In wild type, 3-oxohexanoyl-homoserine lactone (3OC₆-AHL) and C8-AHLs are the primary quorum sensing signals for *V. fischeri*, allowing them to sense cell density and trigger bioluminescence. From an enzymatic point of view, bioluminescence is brought on by the enzyme luciferase, which catalyzes the oxidation of a luciferin, causing it to produce this visible glow. During this lab, the C8-AHL synthesized in the previous week will be used to modulate luminescence in *V. fischeri* DC22, a mutant strain of which lacks the ability to produce its own AHL but responds to exogenous C8-AHL. If laboratory synthesis of C8-AHL was successful, addition to the DC22 mutant *V. fischeri* should result in bioluminescence.

Objectives

The C8 AHL synthesized during week 1 will be used to modulate the expression of bioluminescence in *V. fischeri* bacterial cultures. Multiple concentrations of AHL will be run in order to produce a dose-response curve, and all samples will be run in duplicates. Overnight cultures of DC22 will be incubated in wells of a 96-well plate with varying concentrations of AHLs, and luminescence will be monitored over time.

Materials

- *V. fischeri* cultures
- YTSS Media (per liter: 4g tryptone, 2.5g yeast extract, 15g Sea Salts)
- Octanoyl-L-homoserine lactones (C8-AHLs) synthesized by the previous procedure
- 96 Well Plate with clear bottom
- Petri dish
- Spatulas
- Micropipette

- Microsoft Excel

Methods

1. NOTE: 1:100 *V. fischeri* cultures diluted into sterile YTSS media will be prepared by the stockroom.
2. Each AHL treatment will be run in duplicate. The treatments are 0 μM , 2.5 μM , 5 μM , and 10 μM AHL in *V. fischeri* culture (final concentrations).
 - 90 μL of a solution containing 1:100 of *V. fischeri* culture : YTSS media, and
 - 10 μL of AHL solutions

A. Calculations

The two 0 μM AHL wells will be loaded with 100 μL of 1:100 of *V. fischeri* culture : YTSS media solution. Each of the AHL containing sample wells will be loaded with 90 μL of 1:100 *V. fischeri* : YTSS media solution and 10 μL of AHL solution.

Using the following questions as a guide, make a 10 mM stock solution of C8-AHL, then dilute aliquots of the stock solution with DMSO to concentrations of 100 μM , 50 μM , and 25 μM . Note: The molecular weight of C8-AHL is 227.30 g/mol.

- a) Based on the mass yielded by the AHL synthesis, how many millimoles are present?

- b) Using the synthetic C8-AHL, what total volume is required to yield a 10 mM solution in DMSO?

Calculate the dilutions required to give 10 μM , 5 μM , and 2.5 μM solutions of C8-AHL.

Hint: Use the dilution equation $M_1V_1=M_2V_2$

- c) What volumes of the 10 mM stock solution and of DMSO should be used to make 1 mL of 10 μM AHL?
- d) How many μL of the 100 μM solution should be used to make 1mL 50 μM solution?
- e) How many μL of the 50 μM solution should be used to make 1mL of 25 μM solution?
- f) Calculate the volume of each AHL solution required for achieving the desired final concentrations in each well.

B. Assay

1. Obtain approval from the TA before proceeding with these dilution steps, as incorrect calculations will lead to faulty experimental results.
2. Based on the calculations above, prepare 10 mM a stock solution of C8-AHL. Swirl the DMSO around the flask to ensure all AHL has been dissolved.
3. Based on the above calculations, prepare the three solutions of C8-AHL (100 μM , 50 μM and 25 μM) using serial dilutions.
 - First prepare the 100 μM solution
 - Dilute and aliquot of the 100 μM dilute solution to prepare the 50 μM solution
 - Lastly, using an aliquot of the 50 μM solution, prepare the 25 μM solution
4. To each control well, add 100 μL of 1:100 of *V. fischeri* culture : YTSS media solution.
5. To each well containing AHL sample, add 90 μL of prepared 1:100 *fischeri* : YTSS solution. Use the same tip for all wells.
6. To each well containing AHL sample, add 10 μL of the desired dilute AHL sample to its designated well. Change tips between dilution types, and pre-wet the tip. Mix contents well by pipetting up and down.

7. Place plate in Biotech Synergy HTX microplate reader.
8. Incubate at 24 °C while shaking 200 rpm.
9. Measure luminescence every 30 min for 5 h
10. Record optical density (OD) at 595 nm/600 nm for each time point to provide a set of metadata for normalizing the luminescence values.

C. Data Analysis

1. Once the excel data has been distributed, compare the cells and well assignments on the spreadsheet to the layout of the well plate used in lab.
2. Each group will only analyze the data from their own wells.
3. Normalize to cell density by dividing peak luminescence by OD600
4. Create a scatterplot from the normalized luminescence data. Show increase in luminescence on the y-axis vs. time (min) on the x-axis. Graph the fluorescence results from all 4 treatments (0 μ M, 10 μ M, 5 μ M, and 2.5 μ M) as well as the controls on the same graph.
5. Make sure to include a legend to identify the data; graph title; axis titles and units; and since we are averaging measurements, include standard error bars for the min and max of each set of replicates.

VITA

Caleb Michael Gibson attended Carson-Newman University for his undergraduate studies, participating in chemistry research under Dr. Paul Martino. He received a bachelor's degree in chemistry and philosophy with honors in 2014. Continuing to graduate school at the University of Tennessee, he joined the lab of Dr. Shawn Campagna. During graduate school, Caleb worked as a research assistant, teaching assistant, and instructor. He uses chemical tools to explore systems biology while maintaining a passion for chemical education and academia. In an alignment of these interests, he developed and authored curriculum for an undergraduate chemical biology laboratory.