

Rockefeller University

Digital Commons @ RU

Student Theses and Dissertations

2024

Discovery and Optimization of a Family of Naturally Encoded Antibiotic Compounds that Evade Antibiotic

Adam Rosenzweig

Follow this and additional works at: https://digitalcommons.rockefeller.edu/student_theses_and_dissertations



Part of the Life Sciences Commons



DISCOVERY AND OPTIMIZATION OF A FAMILY OF NATURALLY ENCODED
ANTIBIOTIC COMPOUNDS THAT EVADE ANTIBIOTIC RESISTANCE

A Thesis Presented to the Faculty of
The Rockefeller University
in Partial Fulfillment of the Requirements for
the degree of Doctor of Philosophy

by
Adam Rosenzweig
April 2024

DISCOVERY AND OPTIMIZATION OF A FAMILY OF NATURALLY ENCODED ANTIBIOTIC COMPOUNDS THAT EVADE ANTIBIOTIC RESISTANCE

Adam Rosenzweig
The Rockefeller University 2024

Antibacterial resistant infections are an ongoing global health emergency. To combat this, novel potent antibiotics with unique modes of action are required. The classic arsenal of antibiotics used in clinical settings are largely natural products discovered by bulk fermentation of bacterial species isolates that are then extracted and purified to yield a bioactive compound. Unfortunately, this discovery pipeline no longer produces novel molecules, thus new discovery methodologies are essential to continue to identify antibiotic clinical candidate molecules. In the modern era, coupling natural product discovery with sequencing technologies has proven to be an efficient and fruitful method of identifying secondary metabolite natural products with unique bioactivity that could not previously be accessed through fermentation-based methods. This novel approach presents a promising reinvigoration to the field of antibiotics discovery.

In our lab, we have developed a discovery method by which sequenced bacterial genomes are analyzed using bioinformatic algorithms to identify biosynthetic gene clusters (BGCs) and to make a structural prediction of the molecular product of a given cluster. This molecular prediction can be built using synthetic chemistry and assayed for its biological activity. The name given to this method and the resulting products that are synthesized is synthetic bioinformatic natural products (synBNP). It is using this method that our lab previously discovered cilagicin, a lipopeptide natural product that shows robust Gram-positive antibiotic activity and evades antibiotic resistance even after prolonged pathogen exposure. This resistance evasion is attributed to a dual polyprenyl phosphate binding mechanism. In this thesis, I present discovery and optimization efforts to expand this promising novel class of antibiotic natural products as well as to develop a singular lead clinical drug candidate that displays optimal bioactivity and *in vivo* efficacy.

In Chapter 2, I present an investigation of bioinformatically screened predicted non-ribosomal polypeptide synthetase encoded structures to identify previously uncharacterized antibiotics that may possess the same molecular targets and resistance evasion ability seen with cilagicin. This structure-based screen yields three BGCs predicted to produce natural analogs of cilagicin. synBNPs for the products of each of these three clusters are synthesized and the resulting molecules are assayed for their biological activity. These compounds, called paenilagicin, bacilagicin, and virgilagicin, are shown to be potent antibiotics against multidrug-resistant Gram-positive pathogens. Paenilagicin and virgilagicin are further shown to engage both of the same polyprenyl phosphate targets as cilagicin, and both also demonstrate the ability to evade antibiotic resistance. Bacilagicin is shown to only bind a single molecular target and is susceptible to antibiotic resistance development. This discovery project expands the members of this family of polyprenyl phosphate binding antibiotics, which allows us to identify a conserved peptide moiety that we suspect may play a role in target engagement.

In Chapter 3, building upon the structural diversity identified among polyprenyl phosphate binding antibiotics, I discuss a structural optimization project in which we sought to design an improved version of the most potent antibiotic in this family, cilagicin. To achieve our goal of a molecule with strong antibiotic activity and low serum protein binding to preserve activity *in vivo*, we conduct two regional analyses of the overall molecular structure. In the first, we

investigate the effect of structurally diverse lipid tail substituents on bioactivity. In the second, we conduct a series of orthogonal scans on the peptide core to explore the impact of different peptide moieties with various chemical properties on bioactivity. Ultimately, we identify an optimized compound, called dodecacilagicin, that maintains high Gram-positive antibiotic activity, shows minimal serum protein binding, and also evades antibiotic resistance.

Overall, the work presented in this thesis represents the application and expansion of the synBNP discovery method as an ever evolving and robust means to discover novel structurally diverse natural products with unique bioactivity that were previously inaccessible by culture-dependent methods.

*I dedicate this thesis to every queer and trans person there is,
has been, and ever will be*

ACKNOWLEDGMENTS

There have been many people who helped get me to the point of completing this degree. Firstly, I must thank my PhD advisor Dr. Sean Brady. Thank you for your years of guidance, mentorship, and patience. Thank you for the freedom to explore and the direction to reorient myself when I would get lost. I also must thank my thesis committee, Dr. Tarun Kapoor, Dr. Kayvan Keshari, and Dr. Katya Vinogradova, your generous guidance as I advanced through my research got me to where I am today, and it has been a pleasure to try and impress you all with my baking when I felt I couldn't impress with my science. To all Brady Lab Members whom I have been able to work alongside during my PhD, thank you for being friends and mentors.

Next, I want to thank my family. Mom and Dad, thank you for believing that I can do anything, and thank you for giving me room to grow. Eric, thank you for being my brother, for hours long phone calls, and for a lifetime of building our brotherhood. Elizabeth, thank you for being my guardian since day one, for being my confidant, for loving me most when I was a PhD Monster, and for always being free to talk.

Delaney, Georgi, and Bella, you are my favorite people on Earth. Thank you for being my best friends for over 10 years. It has been a pleasure to grow up with you and I cannot wait to see how far we will all go in our lives. There are no people I depend on more, and no people I would more quickly move heaven and earth for. I look forward to retiring in our Golden Girls house someday. I love you, I love you, I love you.

Now to my chosen family of amazing gay men who taught me how to have fun, to love life, and to love myself. Made of too many members to count. My Jersey Boys, my Nudies, my Science Gays. The times I have spent with you all; watching Drag Race, at movie nights, laying on the beach, and dancing under a disco ball until sunrise, have been the most meaningful moments of my life. You all took me in at different times, but I learned from you, we laughed until our stomachs hurt, we made a lot of questionable decisions, and it was always worth it. The person I am now would not exist without your constant love, support, and friendship. I look forward to many more years of dancing with you all.

Huge thanks to every person involved with Atoms & Bonds. Jeanne and Imani, thank you for helping me make something I believed in so strongly. To every drag artist who joined me in the lab, thank you for your time, talent, and creativity. To everyone who watched, and to the incredible online queer scientific community I gained through this project, thank you for your support.

To the TPCB class of 2019, thank you for being such a fantastic support network, I can't imagine weathering graduate school during a pandemic with any other group of people. You are some of the kindest, smartest, and most ambitious people I have met. I can't wait to see how you each will change the world with your intellect. A special shout out to Lauren and Gabby, our numerous coffee talks and lunches are often what kept me going through the hardest years of grad school, thank you for being my friends.

To every meltdown, every tear drop, every breakup, rejection, burned bridge, and anxious moment of confusion. You are also a part of this story. Parts I would often like to forget, but these hardships showed me how strong I am. Thank you to my therapist, for helping me navigate these difficult times when they arose, for teaching me how to feel, and for making me a healthier, better Adam.

Finally, thank you to my cat, Cosmo, my best buddy and familiar.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	vi
TABLE OF CONTENTS	vii
LIST OF FIGURES	ix
LIST OF TABLES	x
LIST OF ABBREVIATIONS	xii
CHAPTER 1. INTRODUCTION	1
1.1 Natural Products and Their Applications	1
1.2 Antibiotic Discovery Part 1: The Golden Era	2
1.3 Antibiotic Discovery Part 2: The Modern Era	5
1.3.1 Natural Product Biosynthetic Gene Clusters	5
1.3.2 Sequencing Technologies and New Access to Bacterial BGCs	6
1.4 Culture-Independent Synthetic Methods of Antibiotic Discovery	8
1.5 Discovery of Cilagicin, a Potent synBNP that Evades Resistance Development. 9	
1.5.1 Polyprenyl Phosphates, Key Components of Bacterial Cell Wall Biosynthesis	
.....	11
1.5.2 Binding Multiple Targets Conveys Resistance Evasion to Cilagicin.....	12
CHAPTER 2. STRUCTURE BASED GENOMIC DISCOVERY OF NATURALLY OCCURRING CILAGICIN ANALOG ANTIBIOTICS	13
2.1 Introduction	13
2.2 Results	14
2.2.1 Peptide Sequence Based Genome Library Search for Cilagicin Analog BGCs	
.....	14
2.2.2 SynBNP Prediction and Synthesis of Cilagicin Analog BGCs	15
2.2.3 Bioactive Characterization of Natural Cilagicin Analog synBNPs.....	18
2.2.4 Identification of a Conserved Structural Motif Among Polyprenyl Phosphate Binding Lipopeptides	20
2.2.5 Discussion: Expansion of the Cilagicin Family of Polyprenyl Phosphate Binding Antibiotics	20
CHAPTER 3. DESIGNING AN OPTIMIZED BIOACTIVE AND BIOAVAILABLE CLINICAL DEVELOPMENT CANDIDATE INSPIRED BY CILAGICIN	22
3.1 Introduction	22
3.1 Results	24
3.2.1 Survey of Structurally Diverse Fatty Acid Containing Cilagicin Variants .	24
3.2.2 Orthogonal Scans of Cilagicin Polypeptide Core.....	27
3.2.3 Combining Optimized Features and Final Compound Bioactivity Evaluation	
.....	29
3.2.4 Discussion: Identification of an Optimized Bioavailable Polyprenyl Phosphate Binding Antibiotic.....	31

CHAPTER 4. CONCLUSION AND FUTURE DIRECTIONS	32
4.1 Conclusion	32
4.1 Future Directions.....	33
CHAPTER 5. METHODS.....	35
5.1 Methods for “Structure Based Genomic Discovery of Naturally Occurring Cilagicin Analog Antibiotics”	35
5.1.1 General Experimental Procedures and Materials	35
5.1.2 Identification and bioinformatic analysis of natural cilagicin biosynthetic gene clusters.....	35
5.1.3 Solid Phase Peptide Synthesis	36
5.1.4 Minimum inhibitory concentration (MIC) assay.....	36
5.1.5 Cytotoxicity assay	37
5.1.6 Undecaprenyl phosphate feeding assay.....	37
5.1.7 Evaluating antibiotic resistance by serial passage in liquid broth.....	38
5.2 Methods for “Designing an Optimized Bioactive and Bioavailable Clinical Development Candidate Inspired by Cilagicin”	38
5.2.1 General Experimental Procedures and Materials	38
5.2.2 Solid-Phase Peptide Synthesis.....	38
5.2.3 Determination of Synthetic Peptide Purity	39
5.2.4 Minimum inhibitory concentration (MIC) assay.....	39
5.2.5 Human Serum Binding (MIC) assay	40
5.2.6 Undecaprenyl phosphate feeding assay.....	40
5.2.7 Evaluating resistance development by serial passaging in liquid broth	41
CHAPTER 6. APPENDIX	42
6.1 Appendix Figures for Chapter 2.....	42
6.1 Appendix Figures for Chapter 3.....	47
REFERENCES	61

LIST OF FIGURES

Figure 1.1. The diversity of bacterially derived natural products	2
Figure 1.2. Fermentation based antibiotic discovery.	3
Figure 1.3. Timeline of antibiotic resistance.	4
Figure 1.4. Example of modular secondary metabolite biosynthesis.	6
Figure 1.5. Genetically driven methods for silent BGC activation.....	7
Figure 1.6. synBNP culture-independent antibiotic discovery	9
Figure 1.7. Cilagycin BGC bioinformatic analysis and peptide sequence prediction	10
Figure 1.8. synBNP cilagycin	10
Figure 1.9. Bacterial carrier lipid metabolic cycle.....	11
Figure 2.1. Cilagycin, a dual polyprenyl phosphate binding antibiotic.	14
Figure 2.2. Bioinformatic search of predicted NRPS structures for antibiotics that evade resistance.....	16
Figure 2.3. Polyprenyl phosphate binding synBNP BGCs and structure predictions	17
Figure 2.4. Synthetic scheme for synBNP lipodepsipeptides	18
Figure 2.5. Suppression of antibiotic activity by polyprenyl phosphates and development of resistance in serially passaged cultures.....	19
Figure 3.1. Lipid substituent library synthesized to reduced cilagycin serum binding	22
Figure 3.2. Lipid substituent scans of cilagycin	25
Figure 3.3. Orthogonal polypeptide scans of cilagycin	27
Figure 3.4. Activity summary of combined feature cilagycin variants	30
Figure 3.5. Mode of action and resistance development studies for dodecacilagycin.....	31
Appendix Figure 6.1. NMR spectra for paenilagycin.....	42
Appendix Figure 6.2. NMR spectra for bacilagycin.....	43
Appendix Figure 6.3. NMR spectra for virgilagycin.....	44
Appendix Figure 6.4. HRMS analysis of paenilagycin	45
Appendix Figure 6.5. HRMS analysis of bacilagycin	45
Appendix Figure 6.6. HRMS analysis of virgilagycin	46
Appendix Figure 6.7: High-resolution mass spectrometry data for natural cilagycin analogs.....	46
Appendix Figure 6.8: All species experimental culture conditions	47
Appendix Figure 6.9. NMR spectra for dodecacilagycin	48
Appendix Figure 6.10. HRMS analysis of dodecacilagycin.....	49
Appendix Figure 6.11. Structures and number assignments of synthesized cilagycin variants	50-61
Appendix Figure 6.12. High-resolution mass spectrometry data for all synthetic cilagycin variants	62-64

LIST OF TABLES

Table 2.1. NRPS structure prediction database search results of BGCs ranked by percent identity match to cilagicin linear polypeptide sequence	15
Table 2.2. Biological activity (MICs) of natural cilagicin analogs, reported in $\mu\text{g}/\text{mL}$	18
Table 3.1. Antibiotic activity of cilagicin and cilagicin-BP in a panel of clinically relevant pathogens	23
Table 3.2. Spectrum of activity of analogs with different acyl substituents in the absence (-) and presence (+) of 10% serum	26

LIST OF ABBREVIATIONS

NP:	natural product
FDA:	Food and Drug Administration
DNA:	deoxyribonucleic acid
RNA:	ribonucleic acid
BGC:	biosynthetic gene cluster
A:	adenylation
C:	condensation
T:	thiolation
E:	epimerization
TE:	thioesterase
JGI:	Joint Genome Institute
NCBI:	National Center for Biotechnology Information
eDNA:	environmental DNA
NRPS:	non-ribosomal polypeptide synthetase
PKS:	polyketide synthases
ACP:	acyl-carrier protein
synBNP:	synthetic bioinformatic natural product
SPPS:	solid phase peptide synthesis
C _{starter} :	condensation starter
C55:P :	undecaprenyl phosphate
C55:PP :	undecaprenyl pyrophosphate
CDA:	calcium-dependent antibiotics
CAL:	CoA-ligase
PP:	phosphopantetheine-binding
KS:	ketosynthase
AT:	acyl transferase
HPLC:	high-performance liquid chromatography
prepHPLC:	preparative high-performance liquid chromatography
UV:	ultraviolet
HRMS:	high-resolution mass spectrometry
NMR:	nuclear magnetic resonance
MHz:	megahertz
MIC:	minimum inhibitory concentration
IC ₅₀ :	half maximal inhibitory concentration
LB:	Lysogeny broth
BP:	biphenyl
C10:	decanoic acid
C12:	dodecanoic acid
C14:	myristic acid
C11(unsat.):	10-undecenoic acid
MS/MS:	tandem mass spectrometry
UPLC:	ultra-performance liquid chromatography
QTOF:	quadrupole time-of-flight
DCM:	dichloromethane

DMF:	N,N-dimethylformamide
Fmoc:	fluorenylmethoxycarbonyl
AA:	amino acid
HATU:	hexafluorophosphate azabenzotriazole tetramethyl uranium
DIPEA:	N,N-diisopropylethylamine
DIC:	N,N'-diisopropylcarbodiimide
DMAP:	4-dimethylaminopyridine
HFIP:	hexafluoroisopropanol
PyAOP:	(7-azabenzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate
TFA:	trifluoroacetic acid
TIPS:	triisopropylsilane
DMSO:	dimethylsulfoxide
ATCC:	american type culture collection
USA:	united states of america
H ₂ :	hydrogen gas
CO ₂ :	carbon dioxide
N ₂ :	nitrogen gas
H ₂ O:	water
MTT:	3-(4,5-dimethyl-2-thiazolyl)-2,5-diphenyl-2H-tetrazolium bromide
HEK:	human embryonic kidney
DMEM:	Dulbecco's modified eagle medium
Pen/Strep:	penicillin and streptomycin
DPBS:	Dulbecco's phosphate-buffered saline
SDS:	sodium dodecyl sulfate
OD:	optical density
rpm:	rotations per minute
CaCl ₂ :	calcium chloride
Å:	angstrom
mm:	millimeter
µm:	micrometer
nm:	nanometer
M:	molar
mg:	milligram
mL:	milliliter
µL:	microliter
°C:	degrees Celsius
n:	number of replicates
equiv:	equivalents
v/v:	volume per volume
ppm:	parts per million
min:	minute
Da:	dalton
cps:	counts per second

CHAPTER 1. INTRODUCTION

1.1 Natural Products and Their Applications

Nature is the ultimate organic chemist. Evolution of living organisms over millions of years has resulted in wildly diverse metabolic pathways across all species of life. These pathways produce libraries of different chemical compounds with highly varied structures and functions¹. These compounds are referred to as Natural Products (NPs) and are produced and utilized by living organisms as part of their natural biology. They are also called metabolites, as they are products of metabolism. Natural Products can be separated into two main groups of metabolites; primary and secondary².

Primary metabolites are those compounds which are essential for the normal function and survival of an organism. Examples of these compounds include nucleotides, lipids, amino acids, and molecules found in energy metabolism pathways such as the Krebs Cycle³. Secondary metabolites are products of metabolism that do not have a direct role in standard organismal function and survival but do convey some advantage to the species that produces them⁴. Different species of plants, fungi, and bacteria produce diverse secondary metabolites to improve their fitness for their given environment. The roles that these molecules play are numerous and include but are not limited to chelators, inter-organismal communication, chemical weapons against competitors, and sex hormones⁵.

As well as serving diverse biological functions, the chemical and structural diversity of natural products must also be appreciated. Often based on the scaffolds of products of primary metabolism, secondary metabolites include complex molecules based on polypeptides, polyketides, lipids, nucleic acid derivatives, shikimic acid derivatives, and aminoglycosides⁶. Due to the chemical complexity of natural products, and their diverse biological functions, natural products have been a long-standing source of inspiration to the field of pharmacology and drug discovery⁷. Even in pre-industrial times, humans used herbs and soil as sources of medicine and healing. Now, in modern times, we have been able to isolate compounds from their source organism and take advantage of their specialized biological function⁸. This has led to massive strides in the field of human health, increasing modern quality of life and human longevity. Beyond simple product isolation, scientists are able to perform chemistry on these products to develop better and more potent varieties of naturally inspired therapeutic molecules⁹. In fact, approximately 65% of FDA approved therapeutics are natural products or natural product derived¹⁰.

Some of the most influential natural products that have been discovered in the last century have come from bacterial sources. Bacterial natural products have been rich sources of antibiotics, anticancer, and antiviral compounds (Figure 1.1). Sources of bacteria that produce such valuable compounds vary widely and span the entire earth. Over the years bacteria found in soil, in marine environments, and in the intestinal microbiome of other living organisms have yielded compounds that have contributed to the modern revolution of human health¹¹.

Perhaps the most influential of these compounds are antibiotics. Since the discovery of the first antibiotic, salvarsan, in 1910, the discovery and utilization of antibiotics in medicine has tremendously reduced deaths due to infection, which was once a leading

cause of human mortality, and has increased the average human lifespan by over 20 years¹². The molecules are tremendously important for the advancement of human health in modern society. To maintain this advancement, the field of antibiotic discovery has been striving to unveil novel and potent therapeutics for over a century.

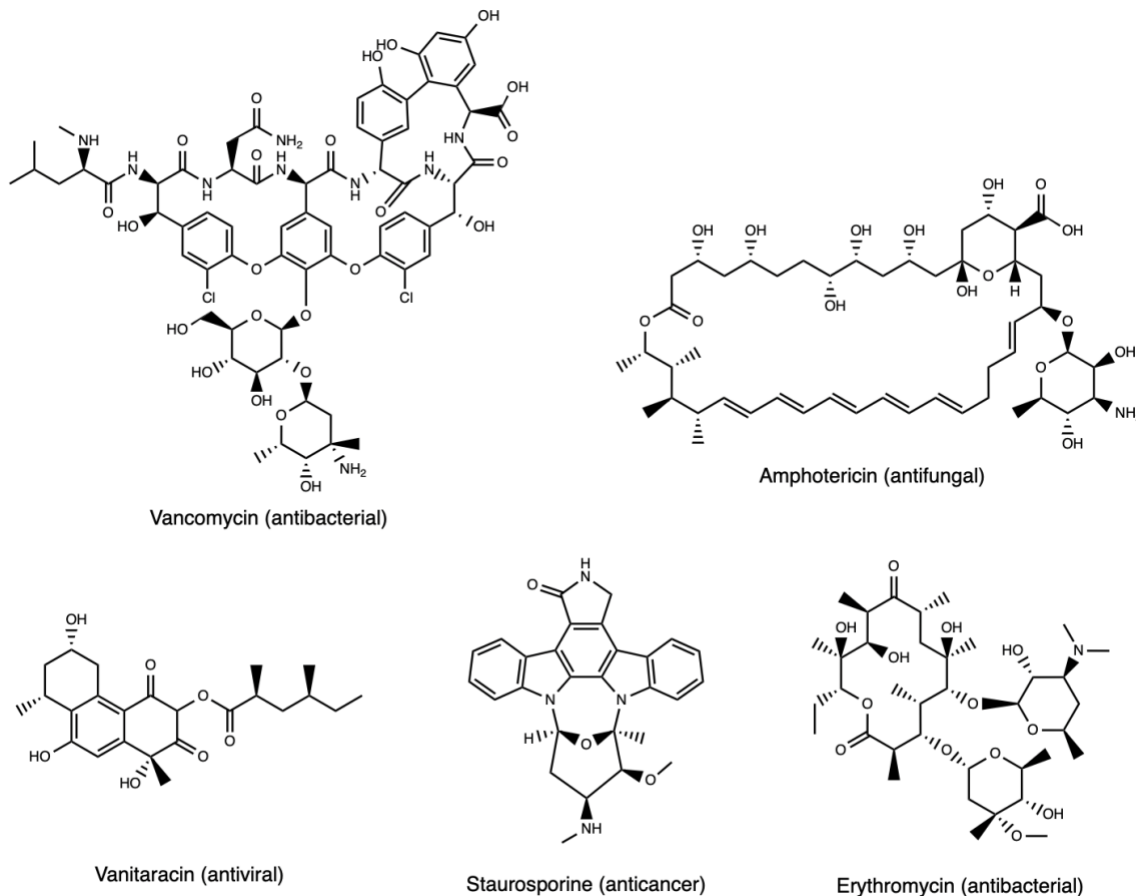


Figure 1.1. The diversity of bacterially derived natural products. Examples of structurally varied secondary metabolites originally discovered in different bacterial species. Therapeutic function of each compound in parentheses.

1.2 Antibiotic Discovery Part 1: The Golden Era

The discovery of penicillin by Alexander Fleming in 1929 marked the beginning of the Golden Age of Discovery for antibiotics which lasted from 1940 to 1962¹³. This discovery begot the first widely used clinical antibiotic therapy, which forever changed the face of the field of infectious diseases. More important than the discovery was the methodology behind penicillin's discovery and isolation. Scientists began performing bulk fermentation of bacterial isolates, followed by fractionation of the bacterial culture extract followed by phenotypic screen of the fractions against a test pathogenic bacterium¹⁴ (Figure 1.2). The bacteria and other microbes used to fuel this boom of fermentation-based discovery were largely derived from soil samples recovered from all over the Earth. This phenotype-based screening method yielded thousands of newly identified biologically active compounds, dozens of which went on to become the arsenal

of antibiotics that made up standard of care therapeutics for bacterial infections the world over. In fact, many of these compounds are still in clinical use today.

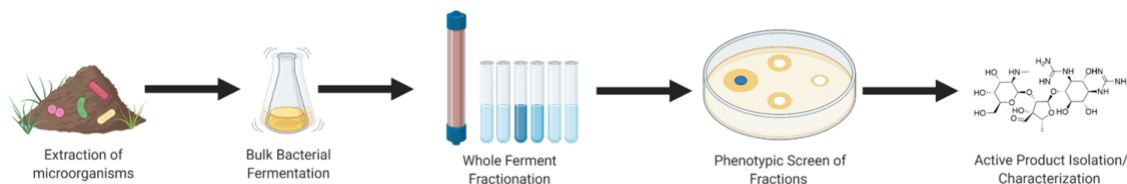
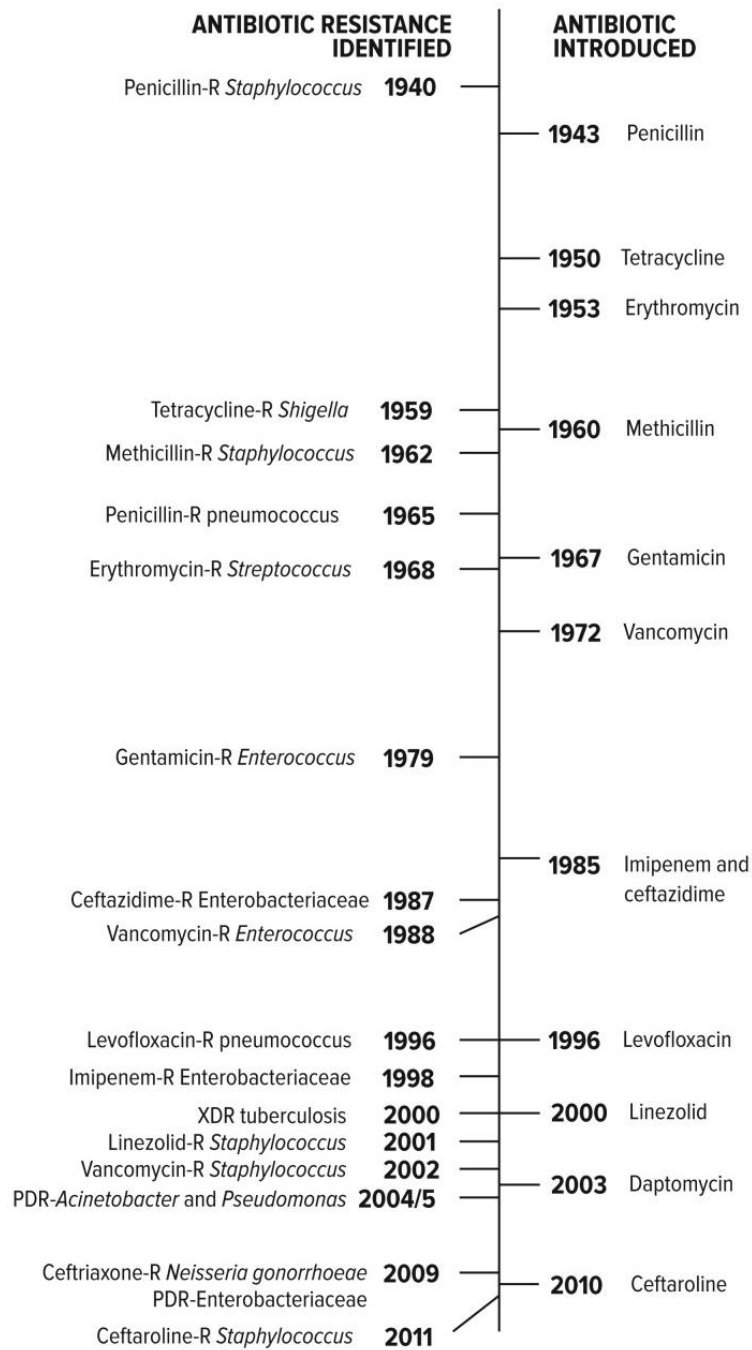


Figure 1.2. Fermentation based antibiotic discovery. The discovery pipeline for traditional antibiotic drug discovery. A bacterial species isolated from a soil sample is grown in bulk culture then fractionated to separate the bacterial metabolites. Fractions are assayed against a test pathogen in a phenotypic screen. The active fraction is further purified to isolate the bioactive compound.

Laboratory culture-based phenotypic discovery was the dominant method that allowed for the rapid expansion of the antibiotic catalog from the 1940s to the 1970s, marking this period as the Golden Era of antibiotic discovery. This bountiful period was marked by the identification of the different drug classes of antibiotics¹⁵. However, moving forward into the next 30 years of discovery, far fewer new classes were found, and the approaches used had to become more sophisticated¹⁶. Firstly, the sources of bacteria and other microbes expanded. Natural products scientists began culturing microbes found in other biomes, particularly those found in samples from marine and aquatic environments¹⁷. Adding additional sources to the pool of microbial species allowed even more compounds to be discovered, with bioactivity beyond just antibiotics. Additionally, sources of microbes improved, as did the discovery methods employed by researchers. Target-based assays were developed to evaluate potential hits against known biological targets that had been identified from the first wave of antibiotic discovery¹⁸. The incorporation of chemical techniques to modify known molecular scaffolds also gained prevalence in the 1990s as a *de novo* discovery approach¹⁴. Combinatorial chemistry allowed for the discovery of novel antibiotic compounds that were inspired by natural compounds¹⁹. These molecules could be produced via completely synthetic or biochemical means. These post-Golden Era years were marked by the discovery of new members of existing families of antibiotics, though discovery of new families of antibiotics with novel mechanisms of action came to a near standstill.

While the prevalence of antibiotics was improving the quality of human life on earth, it came with a critical problem. The emergence of pathogens that were resistant to the therapeutics being used was happening faster than new antibiotics were being discovered²⁰ (Figure 1.3). This of course presented a significant problem, as the field of antibiotic discovery now found itself in an arms race against potentially fatal infectious pathogens, where novel molecules with unique modes of action need to be identified, before the current arsenal of antibiotics are rendered completely ineffective.



PDR = pan-drug-resistant; R = resistant; XDR = extensively drug-resistant

Figure 1.3. Timeline of antibiotic resistance. Chronological order of the introduction of antibiotic compounds into common use and the first identified emergence of clinical pathogens resistant to the antibiotic from 1940-2015. (figure adapted from Ventola, C. 2015)

1.3 Antibiotic Discovery Part 2: The Modern Era

As we entered the new millennium, the problem of slowing rates of antibiotic discovery and increasing rates of antibiotic resistant bacteria only grew. It seemed that the traditional fermentation-based discovery methods had yielded all the results that they could, so new methodologies were required to be able to find novel antibiotic secondary metabolites.

It turns out that the limited application of traditional discovery methods is due to the ways in which the bacterial genes responsible for the production of secondary metabolites are expressed²¹. Across all living species, gene expression, and metabolism is optimized to function in the most energy efficient way possible to best preserve an organism's energy and material resources. Therefore, in the case of antibiotic secondary metabolite production, it is a waste of energy and carbon for a bacterium to be constantly producing these compounds if there are no competing bacteria, or other threat, in its immediate environment. Thus, the library of antibiotics that had been discovered in the Golden Era represented those antibiotics that happened to be highly expressed, from bacterial species which happened to be able to be cultured under laboratory conditions, which ultimately represents a mere fraction of the bacterial species on Earth²².

Enter the age of sequencing technologies. Beginning with the first full sequence of a bacterial genome in 1995, the field's understanding of genetics and the acquisition of sequencing data unlocked a whole new era for antibiotic and secondary metabolite discovery²³.

1.3.1 Natural Product Biosynthetic Gene Clusters

To fully appreciate the impact that sequencing technologies has had on the field of natural product discovery, we need to understand the ways in which secondary metabolite natural products are biosynthesized. The central dogma of biology dictates that genetic information stored in DNA is translated into shorter single stranded copies of RNA and this RNA is translated by ribosomes into proteins. However, where metabolism is concerned, this process can be taken a step further. When the proteins translated from RNA are molecular machines that can perform chemical reactions, i.e. enzymes, these proteins can make molecular products called metabolites. So, following the chain of biochemical information encoded by this expanded central dogma, we can understand that the enzymatic machinery that build secondary metabolites, and to a greater extent the resulting metabolites themselves, are all encoded in the DNA of the producing bacteria.

Evolution has led to bacterial DNA being organized into clusters that group genes responsible for the production of a single metabolite into a linear arrangement with limited gaps²⁴. This is believed to have been adapted by Nature not only as an efficient storage of genetic information but also for the process of horizontal gene transfer, in which bacteria transmit DNA between one another in a non-reproductive process that expands genetic diversity and fitness amongst a bacterial species population²⁵.

Biosynthetic gene clusters (BGCs) are the cornerstone of secondary metabolite production and this organized assembly line is seen across all different classes of secondary metabolites²⁶. The linearity of a BGC is preserved from the DNA level to the

enzyme level and gives rise to a very efficient modular biochemical production pipeline for making secondary metabolites (Figure 1.4).

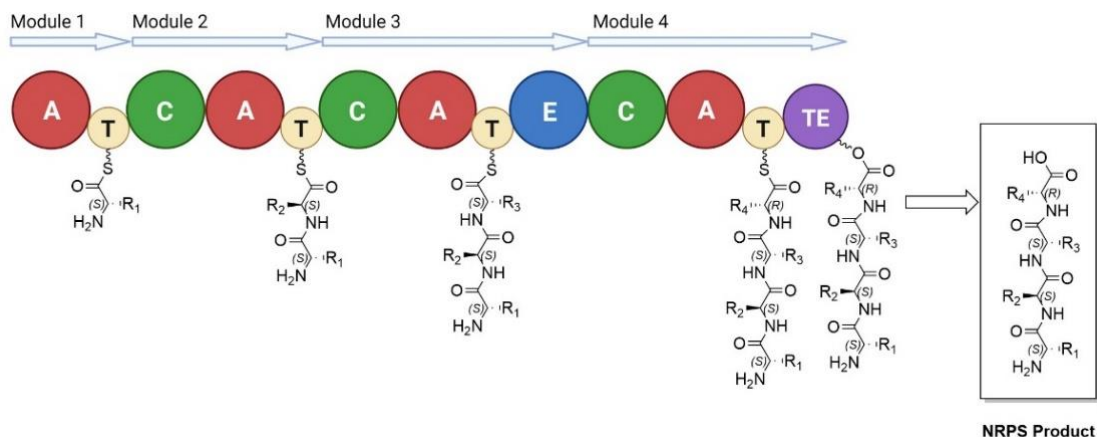


Figure 1.4. Example of modular secondary metabolite biosynthesis. Illustration of modular polypeptide assembly by a generic non-ribosomal polypeptide synthetase (NRPS). (A= Adenylation Domain, T= Thiolation Domain, C= Condensation Domain, E= Epimerization Domain, TE= Thioesterase domain.)

1.3.2 Sequencing Technologies and New Access to Bacterial BGCs

As researchers began sequencing increasing numbers of bacterial species, scientists could begin manipulating BGCs and accessing previously inaccessible natural products²⁷. Currently on publicly available genome repository databases such as GenBank and the Joint Genome Institute (JGI) there are nearly 500,000 sequenced microbial genomes recorded^{28, 29}. As previously stated, the antibiotics that were identified during the Golden Era were mostly the “low hanging fruit” antibiotics. They were products of bacterial species that could be cultured in the laboratory, and products of gene clusters that were either highly expressed or constitutively active. However, these products represent a scant amount of the biochemical diversity that is encoded in bacterial genomes. So, the challenge becomes, how do we get these BGCs that are “off” to be turned “on” so we can access their products.

These “off” BGCs, called silent gene clusters, or cryptic gene clusters can be accessed through a variety of methods that include chemical and genetic manipulation³⁰ (Figure 1.5). All of these methods are dependent on an understanding of the genetic composition of a target BGC. Therefore, a BGC must be identified in a bacterial genome and manipulated by various strategies, some of which include; installing a highly active promoter to the beginning of a BGC or deleting a repressor gene (classical genetics), excising a BGC from a host species and implanting it into a bacterial host with a much higher rate of gene expression (heterologous expression), exposing a bacterial culture to chemical stimuli that alter bacterial metabolism and may stimulate BGC activation as a survival mechanism (chemical genetics), and co-culturing two or more bacterial species to simulate the competitive native environment of the natural world there-by potentially stimulating the bacteria to biosynthesize products of otherwise inactive BGCs (culturing modalities).

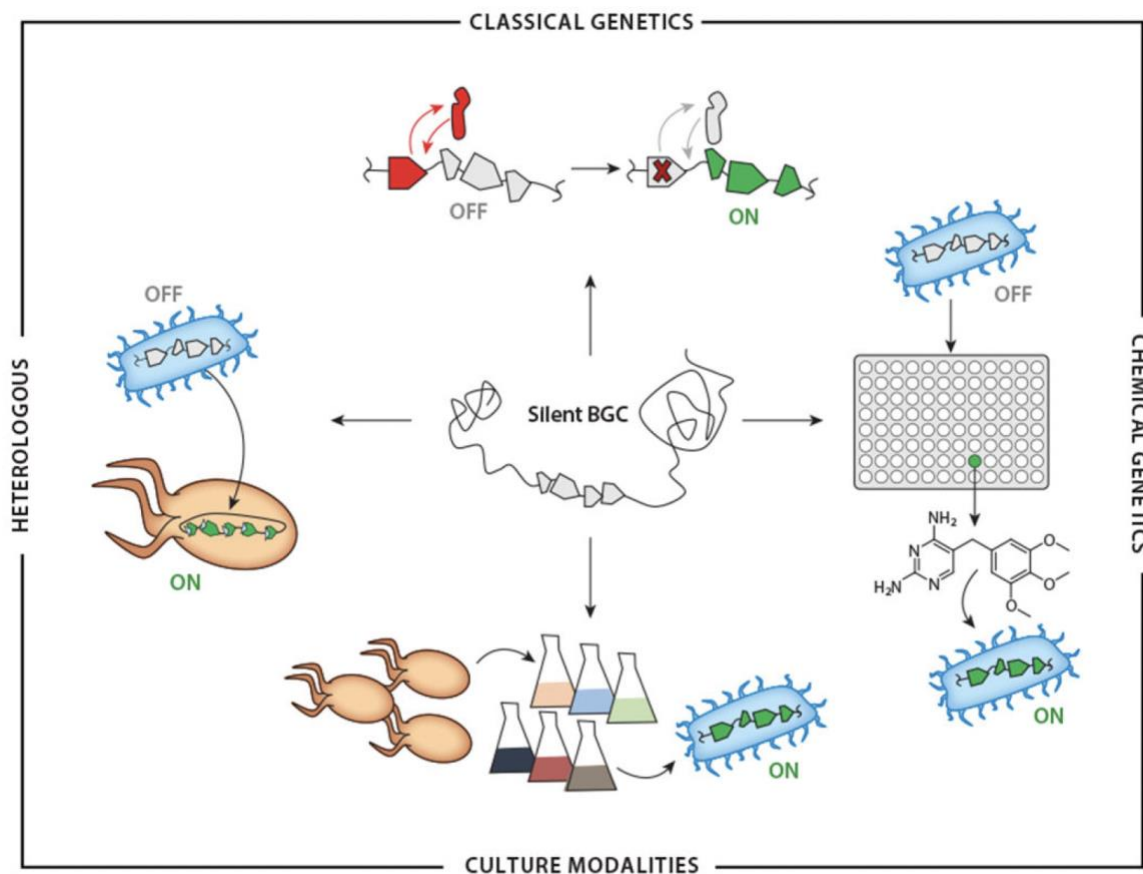


Figure 1.5. Genetically driven methods for silent BGC activation. Various culture-dependent genetically informed approaches to activate silent and cryptic bacterial BGCs (figure adapted from Covington, B. 2021).

While clever approaches for accessing silent gene clusters have yielded novel antibiotic natural products in recent years, these approaches still face one critical hurdle. These methods are still dependent on the ability for an isolated bacterial species to be cultured under laboratory conditions. The field of metagenomics circumvents this by transferring environmental DNA (eDNA) from an unknown host in a mass extraction of a soil or fecal sample into a library of highly culturable bacteria such as *Escherichia coli* followed by thorough sequencing and BGC recovery, before transforming a BGC into a host bacterium for heterologous expression^{31, 32}. Genomic sequencing has found that the majority of BGCs in culturable organisms are silent under laboratory conditions, and so the simple transfer of a metagenomic BGC to a heterologous host does not guarantee expression. Fortunately, numerous strategies (such as taxa matched hosts, hosts engineered for improved expression, increased copy number, promoter refactoring, transcription factor engineering) to activate production of silent BGCs have been developed which are directly applicable to metagenomic BGCs³³.

A revelation in the antibiotic discovery field has come from the emergence of methods that dispose of the need for culturing approaches all together. These culture independent methods are a rapidly developing force in antibiotic discovery and is the chief discovery method behind the work presented in this thesis.

1.4 Culture-Independent Synthetic Methods of Antibiotic Discovery

Until the last decade, all existing methods for decoding BGCs relied on biological processes to convert genetic instructions into natural products. However, as previously described, there is a clear link between the gene sequences in bacteria and the molecular products that they produce. So, the Brady Lab has pioneered a natural product discovery methodology employing bioinformatic algorithms and total chemical synthesis in which BGCs are identified from sequenced genomic data and further analyzed to give a prediction of the molecule that is encoded by the BGC³⁴.

The ability to decipher natural product structural predictions from a DNA sequence relies on a few key pieces of information. The first, is the modular structure of BGCs. The previously mentioned linear nature of BGCs allows for prediction of which chemical reactions will take place in which order as a product is synthesized down the assembly line of enzyme megasynthases. To this end, some classes of secondary metabolites are more modularly organized than others. The two classes that are most linear in their genetic sequences and step-wise in their enzyme biochemistry are polyketide synthases (PKS) and non-ribosomal polypeptide synthetases (NRPS). The second key piece of information is that prior sequencing analysis of known BGCs allows for the prediction of what enzymatic domains of a BGC are encoded by which regions of DNA. Domains of a BGC, as illustrated in Figure 1.4, are the individual enzymes that each perform a single biochemical function, but work collaboratively to create a molecular product^{35, 36}. Anticipating what domains appear in what order allows for the prediction of which chemical transformations are occurring as a product is being made. The final piece of information required to make structural predictions from sequenced DNA is predicting which carbon units are incorporated into the final product at each module of a BGC. In PKS BGCs the carbon units incorporated are malonyl extender units selected by an acyl-carrier protein (ACP) domain, and in NRPS BGCs the carbon units incorporated are amino acids selected by adenylation (A) domains. A revolutionary discovery by Stachelhaus et.al at the turn of the millennium unveiled the specificity-conferring code of NRPS A-domains by which the amino acid residues at ten key positions in the substrate binding pocket of the A-domain can be interpreted to predict the amino acid that will be selected for adenylation by the A-domain and incorporation into a BGC product³⁷. From this discovery, bioinformatic algorithmic tools have been developed that can very accurately analyze bacterial genomes and both identify BGCs and predict the molecular product of the clusters³⁸⁻⁴¹.

With a well-informed structural prediction of a molecular product, total synthesis can be used to produce the desired product. It is this combination of bioinformatic prediction and total synthesis that the Brady Lab has used to discover a host of molecules in recent years called synthetic bioinformatic natural products (synBNP)⁴². This of course works best for molecules that are synthetically tractable, therefore most BGCs pursued by synBNP discovery have been NRPS gene clusters as this class of secondary metabolites can be accurately bioinformatically predicted and can be quickly synthesized via solid phase peptide synthesis (SPPS), then assayed for bioactivity (Figure 1.6). The synBNP method is entirely culture-independent liberating discovery efforts carried out by this method from the restriction of requiring a bacterial species to be culturable in the

laboratory. Additionally, because the basis of this technique is genomic or metagenomic sequencing data, silent and cryptic BGCs that would not be readily expressed in culture can be easily identified and accessed. The molecules discovered by the synBNP method are not guaranteed to be the exact molecule that would be synthesized *in vivo* by the native BGC in its producing host bacterium, but the accuracy of the structural prediction is expected to be a true enough replica that it will maintain the same bioactivity and mechanism of action as the native natural product. Thus, these synBNP molecules can serve as synthetic lead compounds for further optimization and development before becoming candidates for clinical evaluation. SynBNPs have proved to be a rewarding source of small molecules with novel or rarely observed modes of action and potent *in vivo* activity as both antibiotic and anti-cancer agents⁴³⁻⁴⁵.

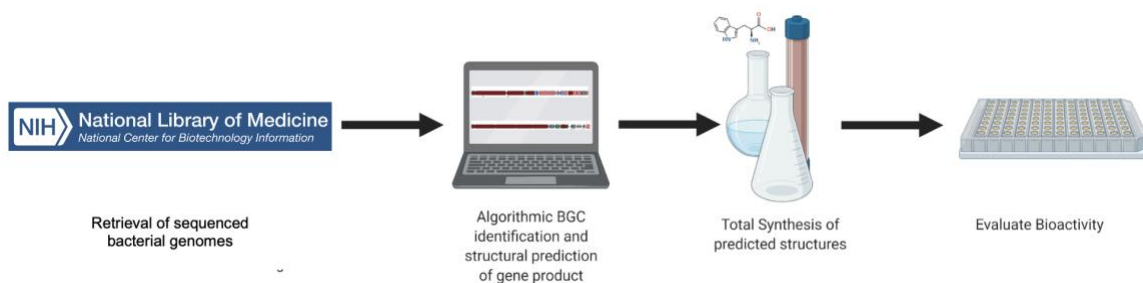


Figure 1.6. synBNP culture-independent antibiotic discovery. The discovery pipeline for synthetic bioinformatic natural product drug discovery. Whole sequenced bacterial genomes are retrieved from publicly available databases. Bioinformatic analysis identifies BGCs and predicts the structure of the genetically encoded molecular product. The prediction is synthesized using chemistry then assayed for bioactivity.

1.5 Discovery of Cilagicin, a Potent synBNP that Evades Resistance Development

One of the most promising antibiotic compounds identified using the synBNP method was recently discovered in a genome mining expedition that yielded a molecule called cilagicin⁴⁶. The BGC that yielded cilagicin was identified by a genome search in which a phylogenetic tree was created comparing the condensation-starter (C_{starter}) domains of bioinformatically predicted BGCs. The rationale behind this approach was that lipopeptides are historically natural products with antibiotic activity⁴⁷. A key feature of lipopeptide NRPS BGCs is the presence of a C_{starter} domain which catalyzes the amide bond formation of a specific fatty acid to the first amino acid substrate on the nascent polypeptide⁴⁸. By identifying C_{starter} domains across sequenced genomes sourced from genome database repositories, the nucleotide sequences of the C_{starter} domain gene could be aligned to construct a phylogenetic tree. This tree revealed clustering of C_{starter} domains from BGCs with known products that share a mechanism of action. It also revealed C_{starter} domains from BGCs that did not have a known product and are less similar in sequence to known lipopeptide antibiotics that may represent undiscovered lipopeptide antibiotics with unique mechanisms of action.

This phylogenetic tree of C_{starter} domain identified one BGC of particular interest that had not been previously characterized and stood alone on the tree but was fairly closely associated with C_{starter} domains of known BGCs that produce antibiotic molecules. This BGC of interest came from the genome of *Paenibacillus mucilaginosus* K02 (Accession

no. NC_017672.3). Following identification and confirmation that this BGC was not associated with any known natural products, this cluster was analyzed for bioinformatic prediction under the synBNP method, and predictions were made for the peptide sequence of the final molecular product of the BGC (Figure 1.7).

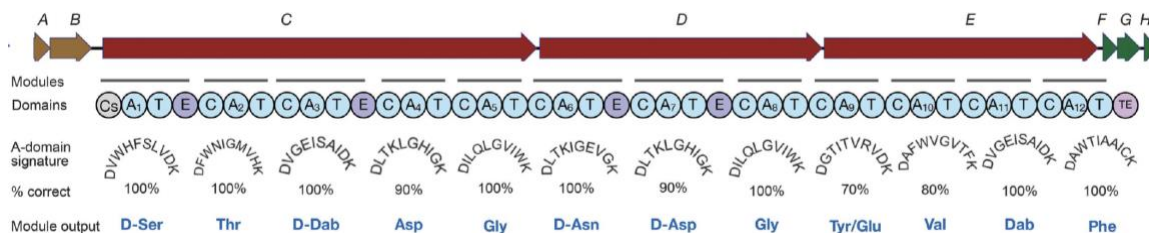


Figure 1.7. Ciligacin BGC bioinformatic analysis and peptide sequence prediction. [From top to bottom] Annotated BGC for *Paenibacillus mucilaginosus* K02 BGC responsible for producing ciligacin. Biosynthetic domains predicted from gene sequence. A-Domain signatures from each BGC module. Percent match to known A-Domain signatures. Predicted amino acid output from each module of the BGC incorporated in final peptide product [Dab = diaminobutyric acid]. (figure adapted from Wang, Z. 2022).

After determining the prediction of the linear peptide sequence for the molecular product of this BGC. Total synthesis was employed to synthesize the predicted product. When making molecular predictions, some features are more difficult to predict than others. In lipopeptide BGCs, predicting the specific fatty acid appended to the N-terminus of the final product is a challenge, so in this synBNP, myristic acid was used as it is among the most common simple straight-chain saturated lipids observed on lipopeptide natural products. Additionally, peptide macrocyclization between the C-terminus and a nucleophilic amino acid residue in the linear peptide is a common feature observed in NRPS natural products, but the specific ring size and site of cyclization is difficult to predict. To address this, several variations of the predicted product were synthesized to assess different linear and cyclic possibilities. These synthetic products were then assayed for antibiotic bioactivity and a single product emerged as the bioactive variant. This molecule was accepted as the synBNP, called ciligacin, and was carried forward for further bioactive characterization (Figure 1.8).

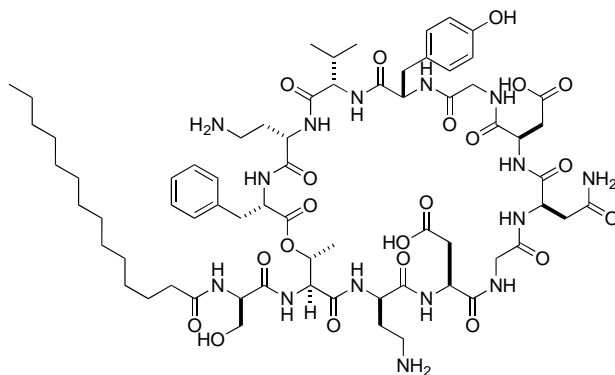


Figure 1.8. synBNP ciligacin. Synthesized predicted structure of antibiotic ciligacin.

The bioactivity profile of ciligacin when assayed against a variety of pathogenic bacteria revealed that ciligacin is a potent antibiotic against Gram-positive bacteria, while largely ineffective against Gram-negative bacteria, and non-cytotoxic against human cells. Notably, ciligacin was effective against a variety of multidrug resistance pathogen isolates. Seeing as ciligacin was ineffective against the outer membrane of Gram-negative bacteria, and that ciligacin is a lipopeptide with a two positively charged residues, it was presumed that ciligacin did not likely exert its antibiotic activity by entering the cell. Instead, it was thought that the mechanism of action may involve binding the cell wall of the pathogen target, as is a common property of positively charged lipopeptides⁴⁹. Molecules like ciligacin often function by rupturing the cell membrane, as seen with polymyxin⁵⁰. Cell membrane depolarization assays with ciligacin indicated that this was not the mechanism of action by which ciligacin was acting as an antibiotic. The next rationalization for a mechanism was that ciligacin was binding a component of the bacterial cell wall and interrupting normal biosynthetic pathways, as seen with vancomycin and daptomycin^{51, 52}. Assays testing binding between ciligacin and known bacterial cell wall metabolites revealed that ciligacin strongly bound two similar but distinct cell wall components, the polyprenyl phosphates, undecaprenyl phosphate (C55:P) and undecaprenyl pyrophosphate (C55:PP).

1.5.1 Polyprenyl Phosphates, Key Components of Bacterial Cell Wall Biosynthesis

C55:P and C55:PP are phospholipids found in the cell membrane of bacteria composed of iterating isoprene units⁵³. They serve an essential function to the construction of the peptidoglycan cell wall that envelops Gram-positive bacteria by chaperoning individual peptidoglycan components from inside the cell to the extracellular space⁵⁴. The phosphorylation state of the carrier lipids changes as they traverse the cell membrane (Figure 1.9)

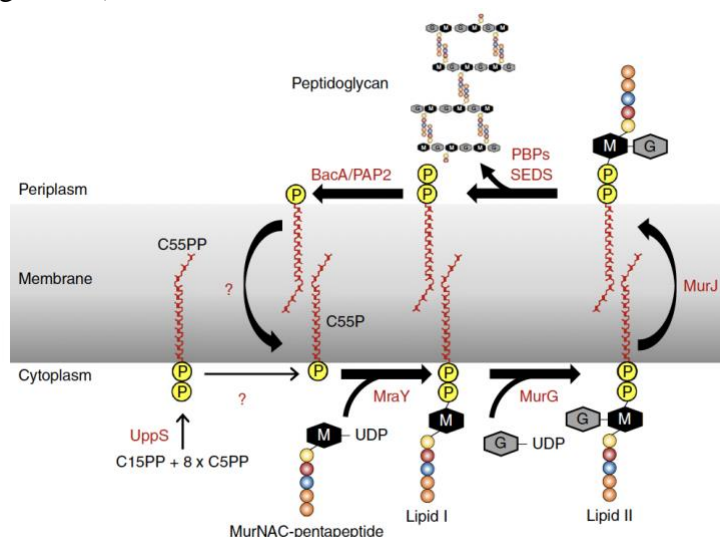


Figure 1.9. Bacterial carrier lipid metabolic cycle. C55:P and C55:PP chaperone peptidoglycan units across the bacterial cell membrane where the units are polymerized into the bacterial cell wall. The polyprenyl phosphates are recycled continually during cell wall biosynthesis. (figure adapted from Ghachi, M. 2018).

Polyprenyl phosphates are known targets of antibiotic molecules, such as bacitracin and friulimicin^{55, 56}. Molecular binding of these carrier lipids disrupts cell wall biosynthesis preventing further bacterial replication and eventual cell death. Polyprenyl phosphate production is tightly regulated by bacterial metabolism, so there is a very small pool of these lipids in a bacterial cell wall, $\sim 10^5$ molecules per cell⁵⁷. Thus, sequestration of free carrier lipids by small molecule binding is a very effective mechanism of action for an antibiotic.

1.5.2 Binding Multiple Targets Conveys Resistance Evasion to Cilagicin

Cilagicin is a unique molecule because it exhibits a rare mechanism of action as it is a dual binder of two components of cell wall biosynthesis, undecaprenyl phosphate (C55:P) and undecaprenyl pyrophosphate (C55:PP). However, its most impressive quality observed during its bioactivity characterization was that after 25 days of serial culture passage while exposed to sub-lethal quantities of cilagicin, test pathogens did not develop resistance to cilagicin⁴⁶. Antibiotic resistance development was observed with control compounds bacitracin and amphomycin, which are single molecule binding antibiotics with affinity to C55:PP and C55:P respectively. This observed resistance evasion is thought to be due to cilagicin binding multiple targets of the same biosynthetic pathway. For antibiotic resistance to develop, functional mutations must occur that either allow bacteria to decompose an antibiotic, or that alter the protein or molecular target of an antibiotic⁵⁸. When two components of the same molecular pathway are sequestered by a single molecule, it would require an immense amount of metabolic restructuring and accumulation of a vast number of specific mutations to evade the antibiotics effects.

In an era where antibiotic resistance is one of the greatest threats to public health, discovering potent antibiotics that evade resistance is the gold standard of antibiotic drug discovery⁵⁹. Therefore, cilagicin is an incredibly unique molecule that shows tremendous promise as a lead compound for clinical drug development. Additionally, further discovery efforts for cilagicin-like molecules would be very advantageous for expanding the current catalog of antibiotic compounds.

In this thesis, I present genomic synBNP discovery efforts to expand the cilagicin family of very promising resistance evading antibiotics, as well as synthetic efforts to improve the bioactivity and bioavailability of cilagicin using a medicinal chemistry approach.

CHAPTER 2. STRUCTURE BASED GENOMIC DISCOVERY OF NATURALLY OCCURRING CILAGICIN ANALOG ANTIBIOTICS

2.1 Introduction

Cilagicin is an exceptional molecule because it binds two molecular targets, and it is shown to evade antibiotic resistance. It is both a very attractive lead compound for future clinical development, as well as the first antibiotic found to have this specific mechanism of dual binding polyprenyl phosphate targets. These unique characteristics motivated our efforts to try and identify if there are other uncharacterized BGCs in the sequenced bacterial genome that may produce molecules that share a similar structure and function with cilagicin.

The benefits of such a discovery effort are numerous. First, the identification of novel distinct antibiotic molecules is always positive for expanding the antibiotic catalog. The even greater benefit of discovering natural products that are analogous to one another, and that share consensus targets, is that having multiple structures that bind the same target can aid in clinical drug development⁶⁰. Analog-based drug discovery is very important for applied research, in which medicinal chemists will synthesize many analogs of a single lead compound to evaluate how slight alterations to regions of a molecule impact its desired activity⁶¹. In the case of natural products, Nature has done much of the troubleshooting work already. Thus, discovering multiple natural products of similar structure and target can give indication as to which regions of a molecule are most important for biological function, i.e. conserved regions, and which regions can be manipulated without impacting bioactivity when designing a molecule with a specific desired mechanism, an example of this is the conserved calcium binding motif identified in calcium-dependent antibiotics (CDAs)⁶². This information is immensely valuable for future drug development efforts of natural product inspired therapeutics.

To go about discovering natural analogs of cilagicin, we used the synBNP method of antibiotic discovery that was taken to originally discover cilagicin (Figure 1.6). However, we decided that for this study we would need to use an advanced search methodology comparing BGC product structure predictions rather than using gene sequence comparisons or phylogenetic approaches, as was employed in the initial cilagicin discovery study. This approach represents an expansion of the synBNP method that combines predicted BGC product structure comparison with total synthesis, yielding a highly adaptable and robust method for discovering naturally encoded molecular products with specific desirable features.

This investigation yielded three novel polyprenyl phosphate binding lipopeptide antibiotics that are active against multidrug-resistant Gram-positive pathogens, named paenilagicin, bacilagicin, and virgilagicin⁶³. Paenilagicin and virgilagicin, did not develop resistance even after prolonged antibiotic exposure. Furthermore, identification of these additional polyprenyl phosphate binding antibiotics allowed us to identify a conserved polypeptide motif that we believe may be important for target engagement in this class of antibiotics.

2.2 Results

2.2.1 Peptide Sequence Based Genome Library Search for Cilagicin Analog BGCs

The cilagicin (*cil*) BGC was originally selected as a synBNP target based on a phylogenetic analysis of condensation starter (C_{starter}) domains. The *cil* C_{starter} domain was related to domains from known antibiotic producing BGCs but was associated with a clade that did not contain sequences from any previously characterized BGCs (Figure 2.1).

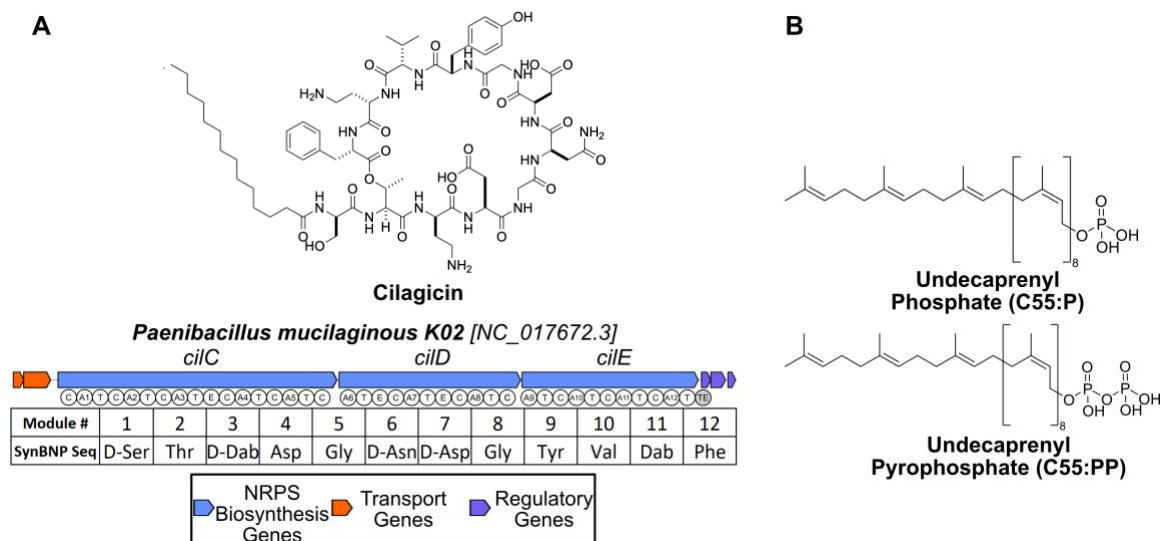


Figure 2.1. Cilagicin, a dual polyprenyl phosphate binding antibiotic. A) Molecular structure of cilagicin and the *cil* BGC. Substrate predictions for each A-domain in the *cil* BGC are shown. B) Structures of polyprenyl phosphates C55:P and C55:PP, the binding targets of cilagicin.

To identify structure predictions that could serve as synBNP targets to produce antibiotics that do not develop resistance, we searched bioinformatically analyzed NRPS-derived NP peptide sequence predictions for structures related to cilagicin. NRPS biosynthesis takes place in an assembly line fashion involving distinct modules containing sets of domains that build a product one amino acid at a time. A canonical NRPS extender module contains a minimum of three domains: a thiolation (T) domain that passes the growing polymer from one module to the next, an adenylation (A) domain that selects and activates a specific amino acid substrate and a condensation (C) domain that catalyzes the formation of an amide bond between the new amino acid and the previously assembled portion of the peptide (Figure 1.4)⁶⁴. The amino acid used by each adenylation (A) domain can be predicted based on 10 amino acid residues that line the substrate binding pocket³⁷.

For this analog search, our goal was to find NRPS BGC molecular products that are structurally similar to cilagicin, based on the assumption that molecules of similar structure will share a similar biological function. Therefore, to approach this project, we needed to adapt the standard gene-based search approach that is traditionally used in synBNP discovery projects to an approach that compared the predicted peptide sequence of NRPS BGC products.

Table 2.1. NRPS structure prediction database search results of BGCs ranked by percent identity match to cilagycin linear polypeptide sequence.

BGC ID	Organism	BGC Length	Best Prediction	Computational Prediction	Identical Residues	Percent Match
NC_017672.3.region1.4	Paenibacillus mucilaginosus K02	85023	Ser Thr 2,4-Dab Asp Gly Asn Asp Gly Tyr Val 2,4-Dab Phe	STJ DGN DGY V F	12	100.0
NC_015690.region1.4	Paenibacillus mucilaginosus KNP414	84964	Ser Thr 2,4-Dab Asp Gly Asn Asp Gly Tyr Val 2,4-Dab Phe	STJ DGN DGY V F	12	100.0
Ga0175260_11.region1.2	Bacillus cereus FORC_024	89309	Ser Thr Orn Asp Gly Asn Asp Gly Tyr Val Ser Leu	ST* DGN DGY V S L	9	75.0
Ga0349645_01.region1.2	Virgibacillus sp. Bac332	85393	Thr Orn Asp Gly Asn Asp Gly Tyr Ile Orn Ile	_T* DGN DGY * I	7	58.3
NZ_WUWM01000002.region56.1	Paenibacillus puerhi strain SJY2	83431	Thr Ser Asp Gly Asn Gly Gly Tyr Ile Orn Tyr	_TSDGN DGY * Y	7	58.3
c00001_NZ_JAGP...region28.1	Streptomyces sp. RK76	57117	Ser Thr Trp Asp Asp Hpg Asp Gly Asn Glu	STWDD6 DGN E_	5	41.7
c00002_NZ_JAHD...region2.1	Streptomyces sp. Tu10	72735	Thr Ala Asp Gly Gly Phe Pro Tyr Trp Gly Phe	_TADG GFP Y W G F	5	41.7
NZ_WWHF01000001.region1.10	Streptomyces sp. SID7813	80612	Ser Thr Trp Asp Asp Hpg Asp Gly Asn Glu Trp	STWDD6 DGN E _	5	41.7
NZ_CP039123.1.region1.22	Streptomyces sp. S552	77081	Ser Thr Trp Asp Gly Asn Glu Trp	STW DGN E _	5	41.7
NZ_QXD01000001.region1.19	Streptomyces sp. 19	80674	Ser Thr Trp Asp Asp Hpg Asp Gly Asn	STWDD6 DGN _	5	41.7
Ga0181086_11.region1.20	Streptomyces sp. 19	80674	Ser Thr Trp Asp Asp Hpg Asp Gly Asn Glu	STWDD6 DGN E _	5	41.7
Ga0314607_106.region6.1	Streptomyces coelicolor A(3)2 C31-mCherry-2	80654	Ser Thr Trp Asp Asp Hpg Asp Gly Asn Glu Trp	STWDD6 DGN E _	5	41.7
NC_022657.1.region1.5	Actinoplanes friuliensis DSM 7358	87947	Asp Pip Asp Asp Gly Asp Gly 2,4-Dab Val Pro Dpr	D h D D G D G V P _	4	33.3
Ga0069497_11.region1.1	Bacillus subtilis 916	67593	Thr Gln Asp Gly Tyr Val	_TQD GY _	4	33.3
NZ_CP033073.1.region1.17	Streptomyces dangxiongensis strain 2022	89331	Val Tyr Asp Gly Asp Gly Tyr Val Leu	_YD D G D G Y V _	3	25.0
c00001_NZ_JAGF...region43.1	Streptomyces sp. NEAU-YJ-81	55214	Asp Pip Gly Asp Gly Asp Gly Thr Val Pro	D h G D D G T V _	3	25.0
NZ_NTGX01000009.region168.1	Streptomyces sp. rh206	51165	Thr Ala Asp Gly Phe Pro Tyr Trp Gly Phe	_TADG FP Y W G F _	3	25.0
CP006272.region1.5	Actinoplanes friuliensis DSM 7358	87947	Asp Pip Asp Asp Gly Asp Gly 2,4-Dab Val Pro Dpr	D h D D G D G V P _	3	25.0
NC_009926.region2.1	Acaryochloris marina MBIC11017	91818	Ser Cys Gly Asp Dpr Gly Gly Tyr Val	_SCGD GG Y V _	3	25.0
NZ_BDBK01000001.region12.3	Nocardia inohanensis NBRC 100128	163108	Ser Thr Gly Thr Tyr Val Val Gly	STG TY V V G _	2	16.7
NZ_BILX01000003.region42.1	Paenibacillus ehimensis NBRC 15659	54299	Gly Val Thr Ser Asp Gly Asn Orn Gly Tyr	_GVTSDGN* GY	2	16.7
c00081_AA76DRA...region81.1	Paenibacillus ehimensis A2	54303	Gly Gln Thr Ser Asp Gly Asn Orn Gly Tyr	_GQTS DGN* GY	2	16.7
c00004_C245DRA...region4.1	Flavobacterium sp. WG21	55463	Ser Asp Asp Thr Asn Tyr Phe	_SDDT NYF	2	16.7
c00009_MXF1DRA...region9.2	Mycococcus xanthus DZF1	79375	Ser Gly 2,4-Dab Asn Glu Val	SG N DEV _	2	16.7
NZ_CP065050.1.region1.40	Streptomyces solisilvae strain HNM0141	86654	Asp Pip Gly Asp Gly Asp Gly Thr Val Pro Dpr	D h G D G D G T V P _	2	16.7

The modified search pipeline we developed for this study used complete NRPS BGCs collected from sequenced bacterial genomes found in the Joint Genome Institute and GenBank databases^{28, 65}. A-domain substrate binding pockets found in these NRPS systems were compared to a manually curated list of signature sequences collected from characterized NRPS BGCs to generate linear peptide product predictions from each BGC. This resulting database of predicted peptide sequences became the starting point for our structure-based analog search. This database was queried with the cilagycin linear peptide sequence and hits were ranked based on the number of positionally identical residues shared with the query sequence (Table 2.1). Two predicted products were identical to cilagycin, both of which were predicted from BGCs found in *Paenibacillus mucilaginosus* genomes. Only three additional predicted NRPS structures shared 7 or more (i.e. >50%) positionally identical residues with cilagycin. There were no predicted products that shared 6 positionally identical residues with cilagycin, and there was a large diverse collection of sequences that shared 5 or fewer positionally identical residues with cilagycin. We hypothesized that the small number of hits sharing >50% positionally identical residues with cilagycin would be the most likely structures to share a mode of action with cilagycin due to high structural homology. Therefore, we focused on these structures for the duration of this investigation. The BGCs from which the three potential antibiotics were predicted were found in the sequenced genomes of *Paenibacillus puerhi* (Accession no. NZ_WUWM01000006.1), *Bacillus cereus* (Accession no. CP068135.1), and *Virgibacillus sp. Bac332* (Accession no. NZ_CP033046.1) (Figure 2.2).

2.2.2 SynBNP Prediction and Synthesis of Cilagycin Analog BGCs

The three predicted BGCs that we identified from our structure-based homology search contained either 11 or 12 NRPS modules and are expected to encode previously uncharacterized, structurally unique undeca- and dodeca- peptides. Each linear peptide

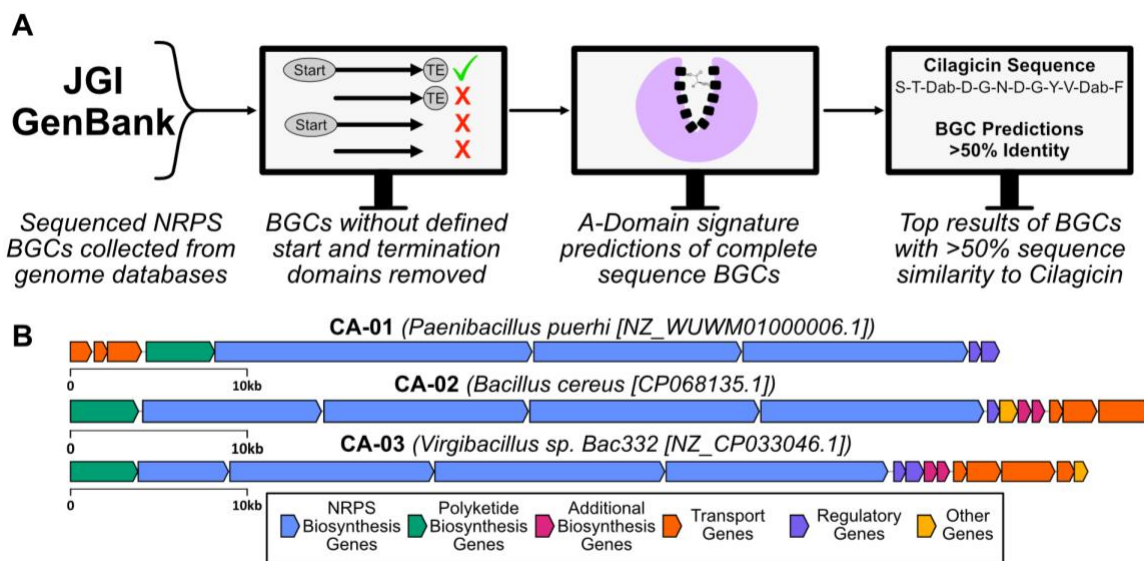


Figure 2.2. Bioinformatic search of predicted NRPS structures for antibiotics that evade resistance. A) Bioinformatic approach used to identify BGCs in this study. Sequenced NRPS BGCs were collected from publicly available genome databases. Collection was filtered to include only BGCs with start and termination domains. Linear peptide sequences from each complete NRPS BGCs were predicted using A-Domain signatures. Resulting predictions were ranked by their identity to the linear peptide sequence of cilgiclin. B) Three BGCs that were predicted to encode linear NRPS products that are >50% identical to the sequence of cilgiclin.

contains a threonine residue at either the first or second module position. Following the cyclization pattern observed in cilgiclin, we predicted that the NPs encoded by these BGCs are cyclized through their C-terminal carboxylates and this conserved N-terminal threonine (Figure 2.3). While the encoded linear peptides are different lengths, cyclization of each molecule through the threonine residue would generate an 11 amino acid cilgiclin-like macrocycle in each product. In the case of cilgiclin, the C_{starter} domain present in the *cilC* NRPS megaenzyme is predicted to add a long chain fatty acid to the N-terminus. Notably, the three BGCs identified in this study contain CoA-ligase (CAL) domains in place of a C_{starter} domain. Similar to C_{starter} domains, CAL-domains are predicted to append an N-terminal lipid onto NRPS encoded polypeptides⁶⁶. The absence of a C_{starter} domain in these three BGCs would explain why they were not identified in the original C_{starter} domain phylogenetic analysis that uncovered the BGC encoding cilgiclin. This realization highlights the importance of the structure-based search method employed in this study. While the three BGCs identified in this study are predicted to produce structural analogs of cilgiclin, the NRPS domain composition of their BGCs differ in significant ways. Therefore, searching for these analog clusters using a genetic homology approach would be ineffective for identifying these novel BGCs. This is a testament not only to the necessity of using different approaches when conducting searches in the sequenced bacterial genome, but also speaks to the adaptability of the synBNP method. So long as a BGC of interest can be identified, via whatever diverse search method is most applicable to one's study, molecular product predictions can be made and novel molecules can be synthesized.

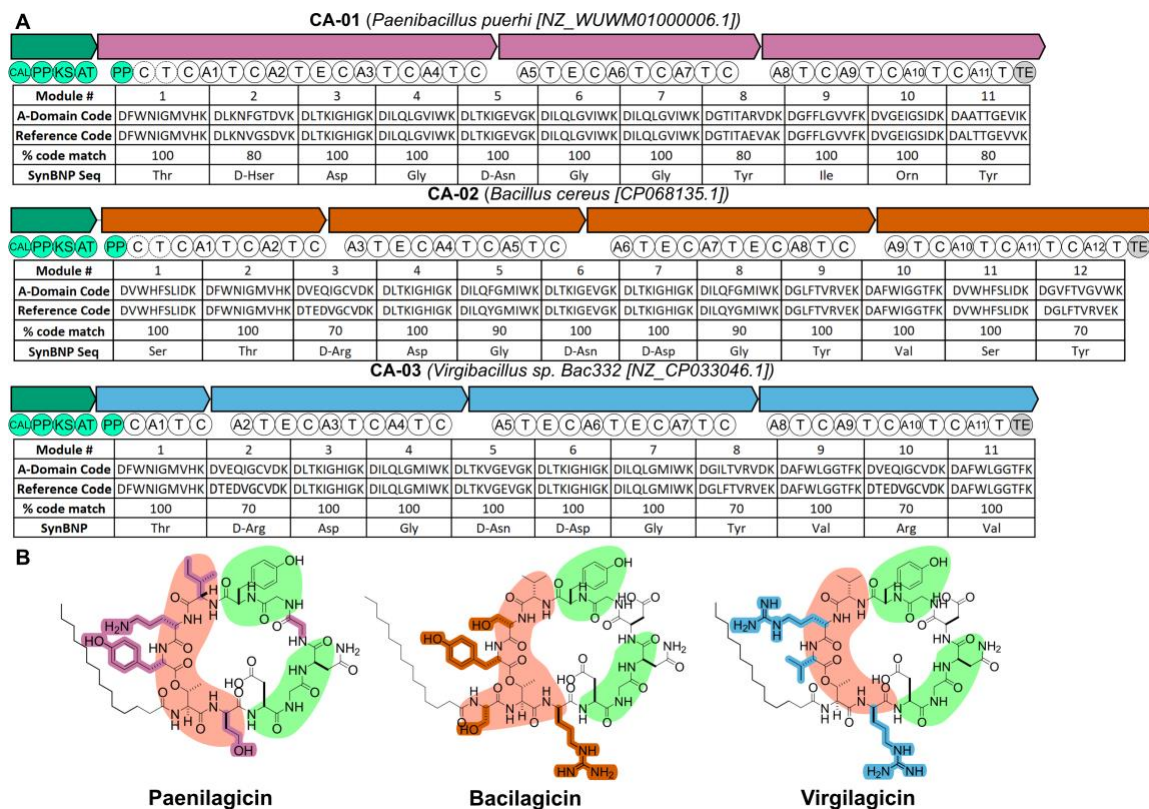


Figure 2.3. Polyprenyl phosphate binding synBNP BGCs and structure predictions. A) Domain composition and A-domain signature analysis of top result BGCs from predicted NRPS structure search [CoA-ligase (CAL), phosphopantetheine-binding (PP), ketosynthase (KS), acyl transferase (AT), adenylation (A), condensation (C), thiolation (T), epimerization (E), thioesterase (TE)]. B) Final structures predicted from each BGC. Variable and conserved regions are highlighted in red and green, respectively. Residues highlighted in respective BGC colors indicate where synBNPs vary from cilagicin.

NRPS-derived lipopeptides are often produced with a range of different lipids incorporated in to their final structure and bioinformatically predicting the exact fatty acid used in their biosynthesis remains a challenge⁶⁷. In our structure prediction analysis, we used myristic acid because it is one of the most common lipids found in lipopeptide secondary metabolites and it displayed potent activity in our original cilagicin study⁴⁶. Based on these bioinformatic arguments, the undeca- and dodeca- lipodepsipeptides predicted to arise from the three NRPS BGCs we identified are shown in Figure 2.3. In reference to the organisms in which these BGCs are found, we have named these structure predictions paenilagicin, bacilagicin, and virgilagicin, respectively.

Upon refining our structural predictions for paenilagicin, bacilagicin, and virgilagicin, we generated a synBNP of each using synthetic chemistry. Bioinformatically predicted linear peptides were synthesized using Fmoc-based solid phase peptide synthesis ending with a myristic acid on the N-terminus. Following linear assembly, ester bonds were formed on resin between the threonine side chain and the predicted amino acid from the last module in each BGC. Branched linear peptides were released from solid support by hexafluoroisopropanol (HFIP) cleavage. Each ring structure was completed in solution via amide coupling between the free amine of the branched amino acid and the carboxylic acid

formerly bound to the resin. Cyclized lipodepsipeptides were deprotected in 95% trifluoroacetic acid (TFA) and purified by high-pressure liquid chromatography (HPLC) to yield the final predicted molecular product (Figure 2.4). All structures were confirmed for identity and purity by HRMS and by ^1H and ^{13}C NMR (Appendix Figures 6.1-6.7).

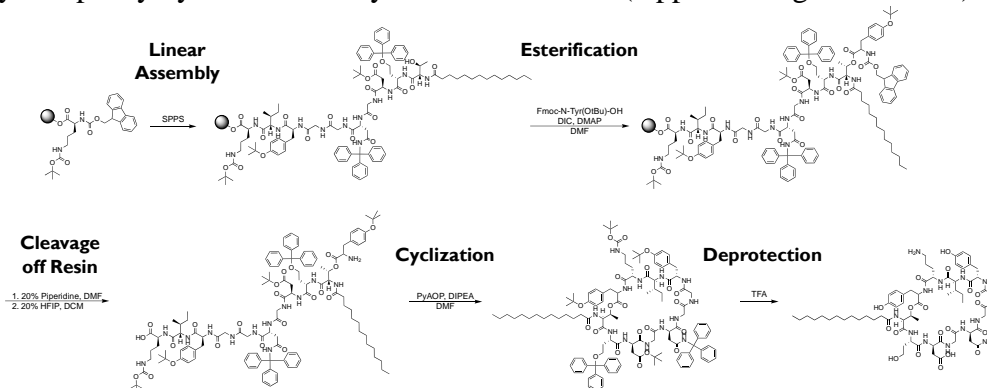


Figure 2.4. Synthetic scheme for synBNP lipodepsipeptides. General synthetic approach used for synBNPs in this study. Linear peptides are built on resin and capped with myristic acid. Ester bond is formed using DIC coupling before cleaving off resin with 20% HFIP. Peptide cyclization is performed in DMF with PyAOP. Final deprotection is performed in a solution of 95% TFA, 2.5% TIPS, and 2.5% H_2O .

2.2.3 Bioactive Characterization of Natural Cilagicin Analog synBNPs

Each synBNP structure was tested for antibiotic activity against Gram-positive and Gram-negative bacteria as well as human cells (Table 2.2). Culture conditions for all species are included in Appendix Figure 6.8. All three synBNPs exhibited activity against clinically relevant Gram-positive pathogens⁶⁸. Against most strains, these new compounds showed a slight reduction in potency compared to cilagicin. However, paenilagicin and virgilagicin were slightly more active than cilagicin against *Clostridium difficile*. Generally, they did not have Gram-negative activity, although paenilagicin, like cilagicin, showed mild activity against *Acinetobacter baumannii*. No synBNPs inhibited the growth of human cells at the highest concentration tested.

Table 2.2. Biological activity (MICs) of natural cilagicin analogs, reported in $\mu\text{g/mL}$.

	Cilagicin	Paenilagicin	Bacilagicin	Virgilagicin
Gram-Positive				
<i>Staphylococcus aureus</i>	1	2	2	4
<i>Enterococcus faecium</i>	2	2	4	8
<i>Enterococcus faecalis</i>	1	2	2	4
<i>Clostridium difficile</i>	4	2	4	2
<i>Streptococcus agalactiae</i>	1	2	4	4
Gram-Negative				
<i>Acinetobacter baumannii</i>	4	8	>64	>64
<i>Escherichia coli</i>	>64	>64	>64	>64
Human Cells				
HEK293 (IC_{50})	>64	>64	>64	>64

Cilagicin is a bi-functional antibiotic that is able to sequester both C55:P and C55:PP⁴⁶. Addition of excess C55:P or C55:PP to culture media suppresses the antibiotic activity of cilagicin by sequestering the antibiotic away from the cell wall of target bacteria. We explored the role of C55:P and C55:PP in the activity of each synBNP by determining the minimum inhibitory concentration (MIC) against *Staphylococcus aureus* USA300 in Lysogeny broth (LB) supplemented with varying ratios of antibiotic and polyprenyl phosphate (Figure 2.5). Minimum inhibitory concentration is defined as the lowest concentration of antibiotic required to completely stop bacterial replication in culture⁶⁹. Like cilagicin, the activity of paenilagicin and virgilagicin were suppressed by C55:P and C55:PP in a dose dependent manner, suggesting these structures retain both polyprenyl phosphates as molecular targets. The activity of bacilagicin was only suppressed by C55:P. In the case of C55:PP, the MIC of bacilagicin remained largely unchanged even when five-fold molar excess of C55:PP was added to the assay media, suggesting that bacilagicin does not sequester C55:PP (Figure 2.5A). The antibacterial activity of cilagicin was completely suppressed at less than a 2-fold molar excess of C55:P and C55:PP. Where suppression of antibacterial activity was observed for the new synBNPs, it required ~3-fold molar excess of a polyprenyl phosphate. This difference in polyprenyl phosphate induced growth inhibition implies that these new antibiotics may have a weaker affinity for C55:P and C55:PP than cilagicin, which could also explain the lower antibiotic potency observed in these new synBNPs.

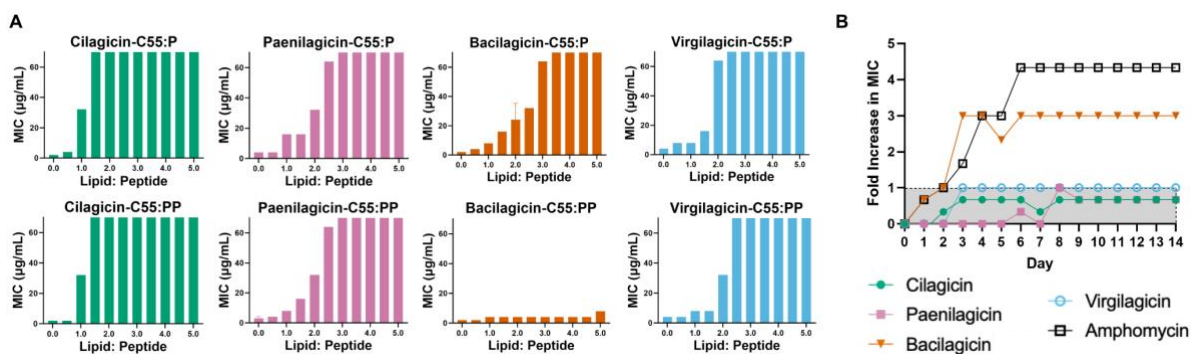


Figure 2.5. Suppression of antibiotic activity by polyprenyl phosphates and development of resistance in serially passed cultures. A) MICs of antibiotics against *S. aureus* USA300 in the presence of different molar ratios of C55:P or C55:PP. The highest concentration of peptide tested was 64 µg/mL. Average of two replicate experiments with error bars representing standard deviation. B) Fold change in MIC from Day 0 to Day 14 of synBNPs against serially passed *S. aureus* USA300 cultures exposed to 0.5x the previous day's MIC of the same antibiotic.

The ability to bind two molecular targets (C55:P and C55:PP) is what we believe enables cilagicin to avoid the development of antibiotic resistance even after prolonged exposure. Other antibiotics that bind a single polyprenyl phosphate molecule (e.g., amphotericin and bacitracin) typically develop resistance quickly^{70, 71}. As such, we expected that the single molecule binding synBNP, bacilagicin, would not be able to avoid resistance development during long-term exposure to a pathogen. To test this, we attempted to raise *S. aureus* USA300 antibiotic resistant mutants by daily serial passage for 14 days in the presence of sub-lethal (half-MIC) concentrations of paenilagicin, bacilagicin,

virgilagicin, cilagicin or amphomycin. As anticipated, cultures exposed to paenilagicin or virgilagicin, like those exposed to cilagicin, did not develop antibiotic resistance. By contrast, cultures exposed to bacilagicin, quickly showed a 4-fold increase in MIC, following a similar pattern to amphomycin (Figure 2.5B). These results reinforce our hypothesis that sequestration of both C55:P and C55:PP provides a unique antibacterial mechanism against which it is difficult for pathogens to develop resistance.

2.2.4 Identification of a Conserved Structural Motif Among Polyprenyl Phosphate Binding Lipopeptides

The structures of these polyprenyl phosphate binding antibiotics show the most variability in the region surrounding the site of cyclization (Figure 2.3B). This variable region includes the three C-terminal residues of each peptide and the residues adjacent to the conserved threonine. The central region of each macrocycle is more highly conserved. In fact, these 10-membered macrocycles all contain a “DGnxGY” motif that we predict is important for target engagement and therefore potentially useful for guiding the discovery of additional polyprenyl phosphate binding antibiotics in the future. A key difference between bacilagicin, which only binds C55:P, and the other structures in this family that bind both C55:P and C55:PP, is the absence of the positively charged residue at position 11. While a more extensive SAR study would be required to determine the role of this residue in target engagement, the extra positive charge may be important for binding the additional negative charge found on the pyrophosphate in C55:PP. We did not find any known antibiotics that share the “DGnxGY” motif. The closest match we found was in locillomycin; however, the stereochemistry of multiple residues is inverted in this structure.⁷² Locillomycin also contains a 9 membered macrocycle in place of the 11 membered ring seen in the polyprenyl phosphate binding synBNPs we have identified. Not only does locillomycin differ in structure from these antibiotics, but biosynthetically it is predicted to arise from the repetitive use of some NRPS modules making the BGC much smaller than those described here. To the best of our knowledge the molecular target of locillomycin has not been reported and therefore whether it binds one or both polyprenyl phosphates, or if it has a different molecular target, remains to be determined.

2.2.5 Discussion: Expansion of the Cilagicin Family of Polyprenyl Phosphate Binding Antibiotics

While cilagicin is a promising candidate for the development of antibiotics that can overcome resistance mechanisms plaguing our current arsenal of approved drugs, it, as is possible with any preclinical class of therapeutics, many will experience unforeseen issues during the development of the compound. With this in mind, we sought to expand the available structural diversity within the cilagicin family by bioinformatically screening sequenced bacterial genomes for BGCs predicted to encode cilagicin-like structures. This search led us to synthesize three additional members of this mechanistically novel class of antibiotics.

As seen with cilagicin, two of these structures do not develop resistance even after extended antibiotic exposure. These structures should provide alternative drug development candidates should they be needed in the future. These novel antibiotics also

provide insight into structural components that are conserved in this polyprenyl phosphate binding family, which is an important characteristic to understand for future structure-driven discovery studies, as well as for understanding the mechanism of action for target engagement of this family of antibiotics. We believe that this study demonstrates how coupling synBNP methods with the targeted search of databases comprised of bioinformatically predicted BGC product structures is now a straightforward and broadly applicable approach for identifying bioactive small molecules with specific desirable features.⁴³

CHAPTER 3. DESIGNING AN OPTIMIZED BIOACTIVE AND BIOAVAILABLE CLINICAL DEVELOPMENT CANDIDATE INSPIRED BY CILAGICIN

3.1 Introduction

There are several qualities of a molecule that must be evaluated when being considered as a clinical drug candidate. At the pre-clinical phase, a compound is assessed for its chemical properties, pharmacological properties, pharmacokinetics, and safety and toxicity⁷³. These properties ensure that the molecule in development is stable, has high affinity for its intended target, is not lethal to animals or humans, and has adequate bioavailability. Bioavailability meaning that the compound can travel into a living organisms circulatory system and reach its desired target⁷⁴.

When cilagicin was discovered, its impressive antibiotic activity and lack of cytotoxicity immediately made it an exciting lead drug development candidate. However, once cilagicin was evaluated in an *in vivo* mouse neutropenic thigh infection model, researchers observed that despite initially promising high plasma bioavailability, cilagicin did not reduce bacterial burden in the host organism⁴⁶. Investigation of this confusing result led to the determination that cilagicin has a strong affinity for serum, which was believed to be the cause of the lack of bacterial potency observed *in vivo*. Protein binding as it affects antibiotic bioavailability is of particular concern for developing antibiotics that will display the desired biological effect when needed⁷⁵⁻⁷⁷. Therefore, to address this, researchers sought to modify the original synBNP of cilagicin to reduce serum protein binding to ensure antibiotic activity *in vivo*.

The positive correlation between a compound's lipophilicity and serum protein binding is well established⁷⁸. Thus, when aiming to reduce the serum binding of cilagicin, the lipid tail was the first moiety that researchers went about manipulating. After attempting a few different cilagicin variations with different lipid tails, a variant with a biphenyl tail, named Cilagicin-BP was found to evade serum binding (Figure 3.1).

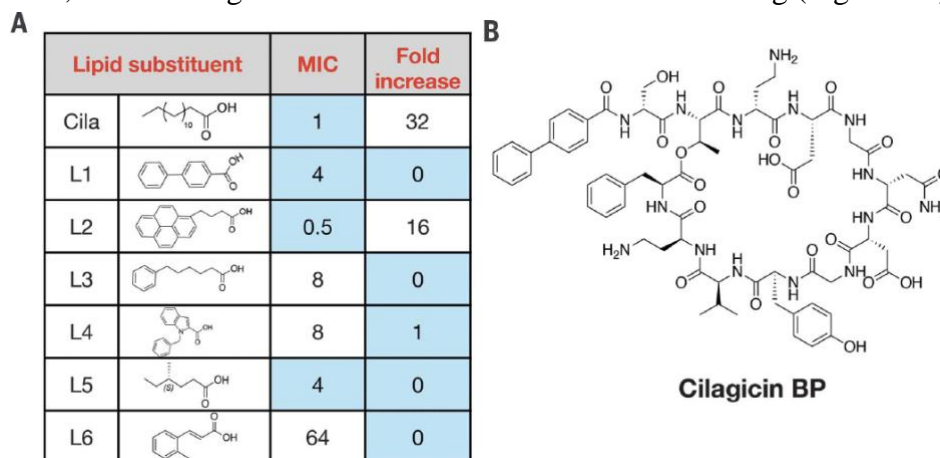


Figure 3.1. Lipid substituent library synthesized to reduced cilagicin serum binding. A) Cilagicin lipid tail variants with bioactivity (MIC) against *S. aureus* USA300 and MIC fold increase in the presence of 10% serum B) Structure of cilagicin-BP (figure adapted from Wang, Z. 2022).

Fortunately, cilagicin-BP achieved the goal of almost no serum binding in *in vitro* and *in vivo* antibiotic activity models. Unfortunately, this came at the sacrifice of antibiotic potency, with cilagicin-BP showing MIC values, at minimum, 4-times greater than those of original cilagicin and completely losing activity against some pathogens (Table 3.1).

Table 3.1. Antibiotic activity of cilagicin and cilagicin-BP in a panel of clinically relevant pathogens (table adapted from Wang, Z. 2022).

Pathogen	MIC $\mu\text{g/mL}$	
	Cilagicin	L1(BP)
<i>S. aureus</i> USA300	1	4
<i>S. aureus</i> USA300 +10% Serum	32	4
<i>E. faecium</i> Com15	0.5	>64
<i>S. pneumoniae</i> Tigr4	0.25	4
<i>S. pyrogens</i> ATCC 19615	0.125	1
<i>S. agalactiae</i> ATCC2675	1	16
<i>C. difficile</i> HM89	2	16

After the successful discovery of novel polyprenyl phosphate binding antibiotics described in Chapter 2, we were motivated to continue investigating this family of antibiotic resistance evading molecules to reach a consensus lead drug development candidate with optimal biochemical properties. Considering our expanded collection of polyprenyl phosphate binding antibiotics, cilagicin still seemed to be the best performing compound as far as antibiotic potency (Table 2.2). However, as previously described, if we were to continue developing a lead candidate from cilagicin, there are still pharmacological properties that need to be optimized to make a viable antibiotic. This goal of lead compound optimization is what motivated the research described in this chapter. We set out to determine if we could identify a variant of cilagicin that would balance strong antibiotic activity and minimal serum binding. We wanted to balance these two properties while maintaining the cilagicin mechanism of action of binding both C55:P and C55:PP, but paramount to all, we wanted to retain the ability to evade antibiotic resistance development

To go about doing this we employed techniques of medicinal chemistry to methodically manipulate the structure of cilagicin in incremental steps to streamline the structure, optimize our desired pharmacological traits, and identify regions of the antibiotic that are essential for function^{79, 80}. We addressed the two main regions of cilagicin; the lipid tail and the polypeptide core. We synthesized analogs of cilagicin individually manipulating these regions to identify optimal versions of each region of the molecule. We then combined our highest performing varieties of each region into a single molecule and further tweaked and altered different characteristics until we created a single molecule that encompassed all of the traits we were seeking.

Our final compound, called dodecacilagicin, incorporates the naturally encoded polypeptide core of cilagicin with a dodecanoic acid lipid tail. This compound shows improved antibiotic performance *in vitro* with minimal serum binding. It also maintains

the same polyprenyl phosphate targets and mechanism of action as original cilagicin and evades antibiotic resistance. Dodecacilagicin is an optimized clinical candidate derived from cilagicin that can be used as an improved lead compound for antibiotic drug development.

3.2 Results

This study involved a massive synthetic effort in which 91 variations of the original cilagicin synBNP were produced. This was accomplished as a collaborative effort between Adam Rosenzweig, Kaylyn Spotton, and Dr. Abir Bhattacharjee. Each molecule synthesized in this study has been assigned a numeric identifier and referred to as such in this chapter. The structure and numeric assignment of each molecule referenced in this chapter can be found in Appendix Figure 6.11. HRMS data for all variants synthesized in this study can be found in Appendix Figure 6.12. Full HRMS, ¹H and ¹³C NMR characterization of the final compound dodecacilagicin is in Appendix Figure 6.9 and 6.10.

3.2.1 Survey of Structurally Diverse Fatty Acid Containing Cilagicin Variants

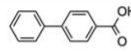

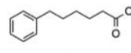
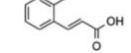
In our original cilagicin discovery study, we found that the lipid substituent appended to the compound's N-terminus had a significant impact on both antibiotic potency and molecular susceptibility to human serum binding. Although cilagicin-BP was less susceptible to human serum binding than cilagicin, it showed reduced potency as an antibiotic against several pathogens. Here, we sought to identify an acyl substituent that would not only avoid human serum binding, but also maintain potent activity against a broader range of pathogens.

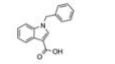
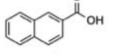
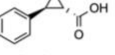
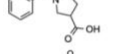
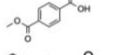


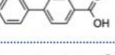
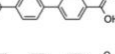
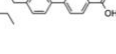
Based on the success of aromatic ring containing N-acyl substituents in the initial cilagicin discovery study (Figure 3.2A, left), we began our lipid tail optimization exploration by synthesizing a series of aromatic ring containing analogs to expand upon these original structures (Figure 3.2A, right). While most of these analogs showed poor antibacterial activity, compounds **12** and **13**, which have substituents more closely resembling the biphenyl structure of cilagicin-BP (**2**), retained activity. When tested in the presence of serum, the methyl biphenyl substituent (**13**) had the most promising bioactivity. In the presence of serum, **13** had an MIC of 1 µg/mL against *S. aureus* USA300, which was 8-fold lower than that observed with cilagicin-BP (**2**). To further explore this biphenyl scaffold, two additional analogs in this biphenyl series were synthesized, compounds **14** and **15**. However, **14** showed reduced antibacterial activity and **15** was highly susceptible to serum binding. Thus, the methyl biphenyl substituent (**13**) was selected for a wider spectrum of activity assays (Table 3.2). Across most of the pathogens we tested, **13** was more potent than cilagicin-BP with and without serum.

This collection of aromatic acyl tails was a promising start, but to ensure we were investigating more diverse structures of lipopeptide acyl substituents, we next systematically surveyed analogs with aliphatic (straight-chain) N-acyl substituents. In a first round of screening, we synthesized and tested the activity of structures with substituents that were 6 to 18 carbons in length to find the pharmacologically optimal carbon length (Figure 3.2B, left). The decanoic (C10, **20**), dodecanoic (C12, **21**) and

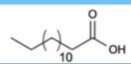
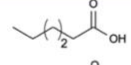
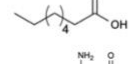

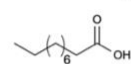
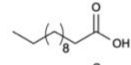
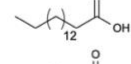
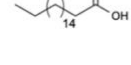
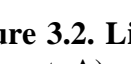
myristic (C14, **1**) acid analogs had the lowest MICs against *S. aureus* USA300, with the dodecanoic acid analog (**21**) displaying a formidable MIC of 0.125 $\mu\text{g/mL}$. This aliphatic

a.

Lipid Substituent	MIC ($\mu\text{g/mL}$)	Fold Change	Serum MIC
2 	4	2	8
3 	0.5	16	8
4 	4	2	8
5 	64		

Lipid Substituent	MIC ($\mu\text{g/mL}$)	Fold Change	Serum MIC
6 	4	1	4
7 	16		
8 	32		
9 	>64		
10 	>64		
11 	16		
12 	2	2	4
13 	0.5	2	1
14 	64		
15 	0.5	64	32

b.

Lipid Substituent	Chain Length	MIC ($\mu\text{g/mL}$)	Fold Change	Serum MIC
1 	14	1	16	16
16 	6	>64		
17 	8	16	1	16
18 	8	>64		
19 	8	8	2	16
20 	10	0.5	2	1
21 	12	0.125	8	1
22 	16	4	>16	>64
23 	18	8	>8	>64

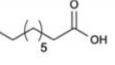
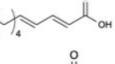
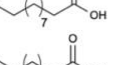
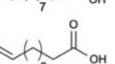
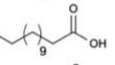
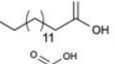
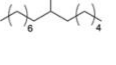
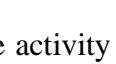
Lipid Substituent	Chain Length	MIC ($\mu\text{g/mL}$)	Fold Change	Serum MIC
24 	9	2	2	4
25 	10	2	2	4
26 	11	0.5	8	4
27 	11	0.5	2	1
28 	12	0.25	8	2
29 	13	0.25	16	4
30 	15	0.5	64	32
31 	15	1	16	16

Figure 3.2. Lipid substituent scans of ciligacin. The activity of ciligacin analogs with different A) aromatic or B) aliphatic acyl substituents determined against *S. aureus* USA300 in Lysogeny broth (LB) both in the absence and presence of 10% human serum, MIC reported in $\mu\text{g/mL}$. Lipid substituents with the lowest MICs in the absence or presence serum are highlighted in green.

scan proved significant for our study as it revealed that the activity of each analog in the presence of serum was highly dependent on the acyl chain length. Increased lipophilicity is often associated with higher serum binding of drugs^{81, 82}. We observed this same general trend with cilagycin analogs, as those containing shorter chain length fatty acids were the least affected by serum. However, we also found that lipid substituents with shorter chain lengths (≤ 8 carbons) led to a decrease in potency. We hypothesize that this is due to the nonpolar tail of these lipopeptides driving localization to the bacterial cell membrane where the polyprenyl phosphate binding targets reside. Thus, lower lipophilicity would result in decreased affinity for the cell membrane. The decanoic (**20**) and dodecanoic (**21**) analogs were the most active structures in the presence of serum and both showed 16-fold lower MICs than the myristic acid containing cilagycin (**1**). This led us to a narrow range of lipid chain lengths, between 10 and 14 carbons in length, that appeared to balance antibacterial activity with low serum binding (Figure 3.2B, left).

In the next round of synthesis and lipid screening we narrowed our focus to structures containing lipids with 9 to 15 carbons, focusing in on our previously determined optimal carbon range. Analogs with lipids containing 11 to 15 carbons (**26-30**) showed the lowest MICs (0.25-0.50 $\mu\text{g/mL}$) in *S. aureus* USA300 (Figure 3.2B, right). The 10-undecenoic acid containing analog (**27**) was the most active structure in the presence of serum with an MIC of 1 $\mu\text{g/mL}$ against *S. aureus* USA300. As we had done with the methyl biphenyl substituent (**13**), the three most potent analogs containing aliphatic lipids, decanoic (**20**), dodecanoic (**21**), and 10-undecenoic acid (**27**), were screened for activity against a larger panel of clinically relevant pathogens in normal culture conditions and in culture supplemented with 10% serum. In the presence and absence of serum, the dodecanoic acid variant (**21**) was consistently the most active analog against the pathogens we tested (Table 3.2). Additionally, in human cell culture, **21** did not display any cytotoxicity in the assayed concentration range.

Table 3.2. Spectrum of activity of analogs with different acyl substituents in the absence (-) and presence (+) of 10% serum.

Compound Acyl Substituent (R)	Cilagycin (1) C14		(2) Biphenyl		(13) Methyl Biphenyl		(20) C10		(21) C12		(27) C11(unsat.)	
	-	+	-	+	-	+	-	+	-	+	-	+
10% Human Serum	-	+	-	+	-	+	-	+	-	+	-	+
Pathogen												
<i>Staphylococcus aureus</i> USA300	1	16	4	8	0.5	1	0.5	1	0.125	1	0.5	1
<i>Enterococcus faecium</i> 802	1	32	32	64	8	32	4	16	1	16	4	16
<i>Enterococcus faecalis</i> AR785	1	16	8	64	4	32	2	16	0.25	8	2	16
<i>Enterococcus gallinarum</i> AR784	0.5	16	8	16	4	16	2	8	0.25	4	2	4
<i>Enterococcus casseliflavus</i> AR798	0.125	2	0.5	2	0.5	2	0.5	1	0.0625	0.5	0.125	1
<i>Streptococcus agalactiae</i> BAA2675	1	8	8	8	2	2	1	1	0.5	1	1	2
<i>Streptococcus pyogenes</i> ATCC19615	0.125	0.5	1	1	0.125	0.125	0.125	0.125	0.0625	0.125	0.125	0.125
<i>Clostridium difficile</i> HM89	1	16	8	8	2	2	1	1	0.5	2	2	2
Frequency of Lowest MIC	1	0	0	0	0	2	0	5	8	7	0	4

3.2.2 Orthogonal Scans of Cilagicin Polypeptide Core

After identifying a few preferred lipid substituents to include in our optimized structure, we next turned to the other essential moiety of the lipopeptide, the polypeptide core. While it was possible for us to systematically synthesize and test cilagicin analogs containing acyl substituents of different lengths and structures for improved activity, a full randomization of the cilagicin peptide core was not practical and would introduce too many simultaneous variables to assess the biochemical contribution of single residues of the polypeptide. Instead, we thought it more rational to conduct a series of orthogonal scans to investigate the depsipeptide structure (Figure 3.3). These scans would allow us to singularly interrogate the effect of altering different chemical properties of individual amino acid residues. Our scans included an alanine scan to investigate the impact of increasing hydrophobicity, a glutamic acid scan to investigate the effect of adding a formal charge, an N-methyl scan which allowed us to explore the effect of manipulating the polypeptide backbone, and a stereochemistry scan in which we replaced of each D-amino acid with its L-stereoisomer counterpart. For each of these scans, we built molecules containing the myristic acid (C14) acyl substituent initially included in cilagicin (**1**).

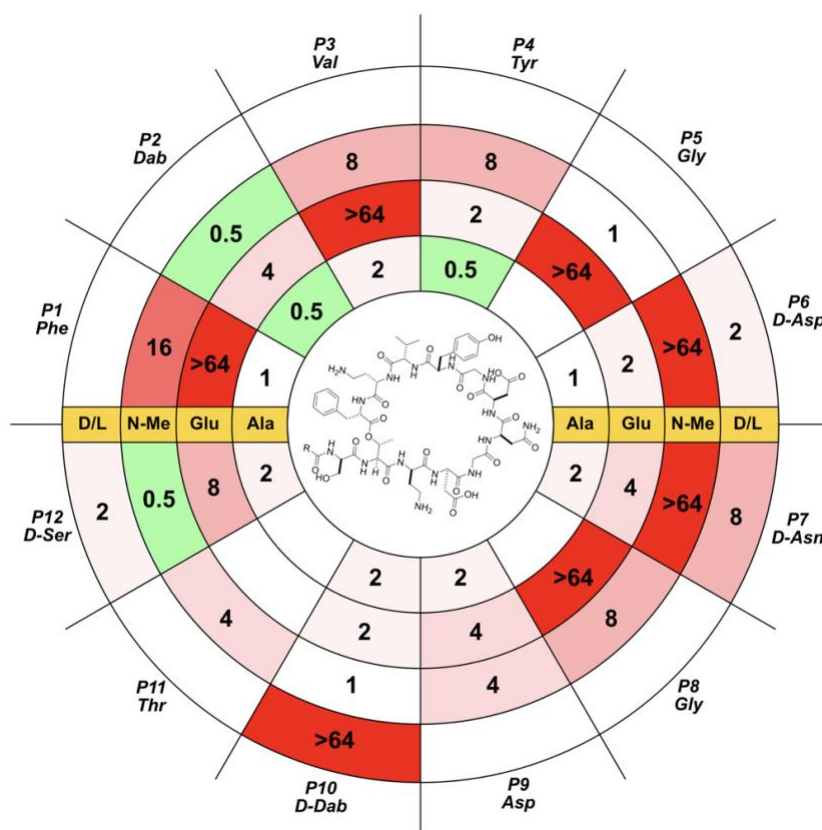


Figure 3.3. Orthogonal polypeptide scans of cilagicin. From inside to outside, each ring reports the *S. aureus* USA300 activity (MIC $\mu\text{g}/\text{mL}$) of cilagicin analogs with single alanine, glutamic acid, N-methyl or side chain D- to L-stereochemistry changes. Each sector corresponds with the change at one position. Improved (lowered) MICs are highlighted in green. Worsened (increased) MICs are highlighted in shades of red.

Alanine scan: In the alanine scan, we replaced each non-glycine residue with the appropriate stereoisomer of alanine. All analogs were then assayed for antibacterial activity against *S. aureus* USA300 (Figure 3.3, Ala ring, compounds **32-40**). Overall, no single alanine substitution resulted in a large change in MIC. The largest increase in MIC was a 2-fold increase to 2 $\mu\text{g/mL}$. Substitution of alanine at P2 or P4, which were originally diaminobutyric acid (Dab) and tyrosine, respectively, each resulted in 2-fold decreases in MIC (0.5 $\mu\text{g/mL}$). These beneficial substitutions would be kept in mind for further assessment once combined with the preferred optimized lipid substituents.

Glutamic acid scan: As no individual hydrophobic (alanine) substitutions led to a dramatic change in MIC, we next explored the effect of changing side chain charge on cilagicin's activity. Cilagicin natively contains two positively charged residues. As highly positively charged peptides are prone to induce cell lysis, and are therefore commonly toxic to human cells, we chose not to explore the addition of more positively charged residues to cilagicin⁸³. We instead explored the effect of adding negative charges by individually changing each residue to the appropriate stereoisomer of glutamic acid (Figure 3.3, Glu ring, compounds **41-51**). Glutamic acid substitutions at positions P1 (Phe), P3 (Val), P5 (Gly) or P8 (Gly) abrogated antibacterial activity completely with a reported MIC of $>64 \mu\text{g/mL}$. Notably, each of these replacements involved the addition of a negative charge where a nonpolar residue had originally been present. Interestingly, the molecular targets of cilagicin are negatively charged, and therefore we would expect the positive side chain residues of cilagicin to be important for target binding. However, we observed only a 2- to 4-fold reduction in antibacterial activity when replacing either positively charged Dab residue (P2 or P10) with glutamic acid. Exchange of the exocyclic Ser (P12) for glutamic acid resulted in an 8-fold increase in MIC, while nothing greater than a 4-fold increase was observed for substitutions of any of the remaining residues: P4 (Tyr), P6 (Asp), P7 (Asn) or P9 (Asp).

D-amino acid scan: A key feature of NRPS encoded natural products, that distinguishes them from ribosomally produced peptides, is the incorporation of D-amino acids⁶⁴. Each inverted residue arises from the presence of a specific epimerization, or condensation, domain capable of inverting the stereochemistry of the α -carbon of an amino acid substrate in the BGC⁸⁴. The retention of these epimerase domains implies that some advantage is conferred by incorporation of D-amino acids in the final structure. Cilagicin contains four D-residues: P6 (D-Asp), P7 (D-Asn), P10 (D-Dab), P12 (D-Ser). To test the importance of the stereochemistry of these D-residues, each was independently exchanged for its L-stereoisomer (Figure 3.3, D/L ring, compounds **64-67**). We found that the stereochemistry of the Dab at P10 was essential for antibiotic activity as its inversion resulted in an observed MIC of $>64 \mu\text{g/mL}$. Changing any of the remaining D-amino acids to their corresponding L-isomer resulted in only a 2- to 8-fold increase in MIC.

N-methyl backbone scan: Crystal structures of polyprenyl phosphate bound antibiotics indicate that coordination between the polypeptide amide backbone plays an essential role in target engagement⁸⁵. To evaluate if this same binding modality had a role in our lipopeptides, we individually replaced each residue with its N-methyl analog (Figure 3.3, N-Me ring, compounds, **52-63**). These changes had variable effects on antibacterial activity. N-methylation of P1 (Phe), P3 (Val), P4 (Tyr) or P8 (Gly) resulted in 8- to 16-fold increases in MIC, while N-methylation at P6 (D-Asp) or P7 (D-Asn)

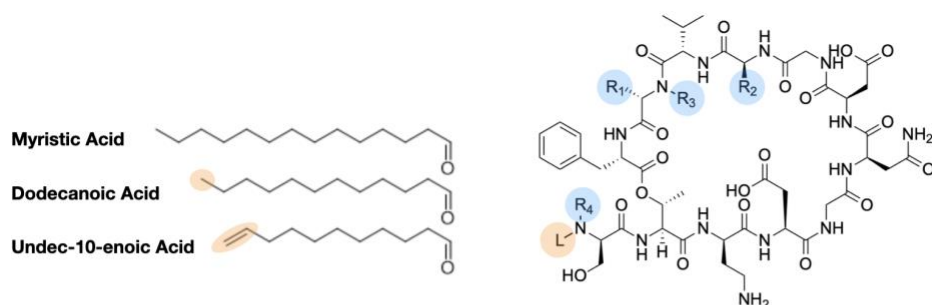
completely abolished activity (MIC >64 µg/mL). Conversely, N-methylation of P2 (Dab) or P12 (D-Ser) resulted in 2-fold improvements in MIC.

Taken together, our four scans of the cilagicin peptide structure suggest that side chain and backbone interactions together are likely important for target engagement as we observed abrogation of activity with both amino acid changes as well as backbone N-methylation. We previously observed that the "DGnxGY" motif of residues P4 through P9 of cilagicin is conserved in natural cilagicin congeners⁶³. While in contrast, residues P12 through P3 of these congeners show considerable variation. Similarly, in the present study, we found that residue positions P12 through P3 appear to be more amenable to change than the residues or backbone of positions P4 through P9. Interestingly, neither of the positively charged residues, P2 or P10, is essential for activity. Individually changing either residue to alanine or glutamic acid did not dramatically reduce the potency of the compound. Considering that the molecular targets of cilagicin are negatively charged, this finding was somewhat surprising but not completely unexpected as the presence of positively charged residues at these positions is not conserved among the natural cilagicin congeners identified in the study described in Chapter 2. To further explore the importance of the P2 and P10 positive charges, we synthesized a single analog where both residues were changed to alanine (**68**). This analog had an MIC of 1 µg/mL, confirming that no positive charges are required for antibiotic activity.

3.2.3 Combining Optimized Features and Final Compound Bioactivity Evaluation

With preferred lipid candidates identified and a better understanding of what regions of the polypeptide core are amenable to manipulation, the next phase of this study was to combine features that reduced the MIC of our optimal lipopeptide. The only core changes that improved potency were the introduction of an alanine at P2 (Dab) or P4 (Tyr) and the N-methylation of the backbone at P2 (Dab) or P12 (D-Ser). Various combinations of these core modifications were synthesized with three of the most promising lipids myristic acid (C14), dodecanoic acid C12, and 10-undecenoic acid (monounsaturated C11). The resulting structures were tested for *S. aureus* USA300 antibiotic activity with and without serum (Figure 3.4). Apart from the double alanine replacement of P2 and P4 with the original C14 acyl substituent (**77**), only analogs with the original or with a single modification to the peptide core resulted in MIC improvement. It appears that any more significant modifications to the depsipeptide core drastically reduces antibacterial activity. Among the final set of improved analogs we identified, **21** (dodecanoic acid with cilagicin core), **27** (10-undecenoic acid with cilagicin core) and **71** (dodecanoic acid with P4 Ala substituted cilagicin core) showed the lowest *S. aureus* USA300 MICs (1 µg/mL) in the presence of serum (Figure 3.4). In our original lipid scan analysis, we tested compounds **21** and **27** against a diverse collection of pathogens and found that compared to cilagicin they both showed improved activity in the presence of human serum (Table 3.2). Further testing of the activity spectrum of compound **71** in the presence of serum found that it showed little or no improvement in potency against most pathogens other than *S. aureus* USA300 when compared to cilagicin (**1**).

Among all analogs we examined, **21** (dodecanoic acid with cilagicin core), which we name dodecacilagicin, best exhibited the features we set out to identify, balancing high antibiotic potency and low serum binding. Cilagicin achieves its unique antibacterial



R1	R2	R3	R4	Myristic Acid		Dodecanoic Acid		Undec-10-enoic Acid				
				-	+	-	+	-	+			
Dab	Tyr	H	H	1	1	16	21	0.125	1	27	0.5	1
Ala	Tyr	H	H	33	0.5	8	69	0.25	2	70	2	
Dab	Ala	H	H	35	0.5	4	71	0.25	1	72	2	
Dab	Tyr	CH ₃	H	53	0.5	16	73	2		74	2	
Dab	Tyr	H	CH ₃	63	0.5	32	75	2		76	16	
Ala	Ala	H	H	77	0.25	2	78	2		79	8	
Dab	Tyr	CH ₃	CH ₃	80	8		81	4		82	>64	
Dab	Ala	CH ₃	CH ₃	83	4		84	8		85	>64	
Ala	Ala	H	CH ₃	86	4		87	16		88	>64	
Ala	Ala	CH ₃	CH ₃	89	4		90	8		91	>64	

Figure 3.4. Activity summary of combined feature ciligacin variants. Acyl substituents are shown with differences from myristic acid highlighted in orange. The locations of peptide core modifications are highlighted in blue. In the chart, each modification is highlighted in yellow. (+) and (-) indicate with and without serum, respectively. Serum (+) MICs ($\mu\text{g/mL}$) of the best performing structures in *S. aureus* USA300 are highlighted in green.

activity and low resistance through its ability to bind both C55:P and C55:PP. For dodecacyliligin to be a good development candidate it must still be able to engage with both targets and show low resistance development⁸⁶.

The activity of ciligacin can be suppressed by providing an excess of either polypropenyl phosphate in culture media because this extracellular pool of target will sequester the antibiotic, thus rendering it inactive. Polypropenyl phosphate suppression provides a facile means of checking that each improved analog retains the ability to interact with both molecular targets. To determine whether the activity of our best optimized lead compound was still suppressed in the presence of excess C55:P and C55:PP, we assayed for activity against *S. aureus* USA300 in media containing increasing ratios of antibiotic and polypropenyl phosphates. Similar to ciligacin, the activity of dodecacyliligin was dose-dependently suppressed by addition of either C55:P or C55:PP in the culture media (Figure 3.5A). This suggests that both molecular targets are maintained by dodecacyliligin.

To test for the development of resistance to dodecacyliligin (**21**), we attempted to raise resistant mutants by daily serial passage of *S. aureus* USA300 in the presence of sub-lethal (0.5x MIC) concentrations of antibiotic. As seen for ciligacin, no resistance developed (i.e. MIC was maintained at $<2 \mu\text{g/mL}$) to dodecacyliligin even after 10 days of continuous exposure (Figure 3.5B).

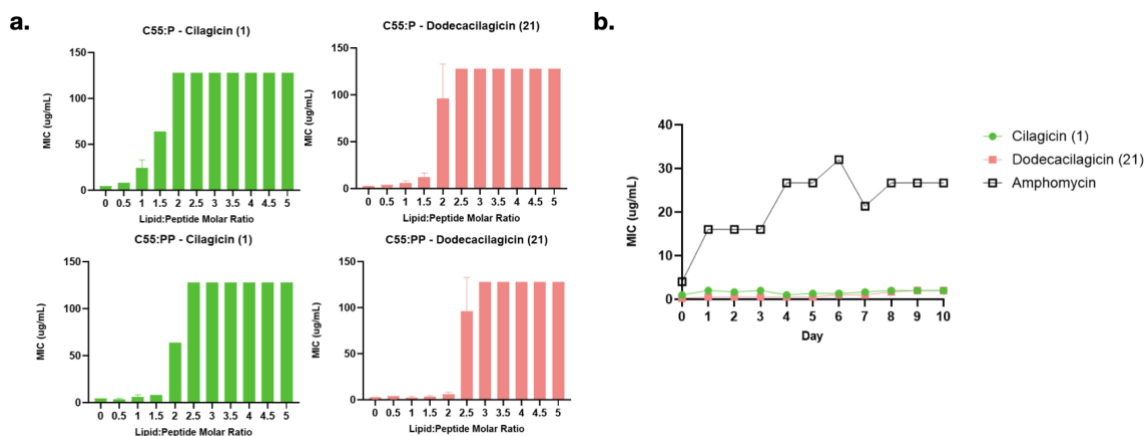


Figure 3.5. Mode of action and resistance development studies for dodecacilagin.

A) Suppression of antibiotic activity against *S. aureus* USA300 by addition of increasing molar ratios of polyprenyl phosphates, C55:P and C55:PP. Data plotted represents the average MICs of two replicate experiments with error bars of standard deviation. B) Resistance development during 10-day serial passaging of *S. aureus* USA300 in the presence of sub-lethal (0.5x MIC) levels of ciligacin, dodecacilagin or amphotericin.

3.2.4 Discussion: Identification of an Optimized Bioavailable Polyprenyl Phosphate Binding Antibiotic

Ciligacin was an appealing initial therapeutic development candidate because it showed activity against multidrug resistant clinical isolates. However, its high serum binding rendered it ineffective *in vivo*. Ciligacin-BP's modified acyl tail decreased serum binding which increased *in vivo* activity, but this change generally reduced potency. In our effort to optimize ciligacin, we found that modifications performed on the peptide core were largely ineffective at increasing potency. However, changes to the hydrophobicity of the aliphatic tail modulated both potency and serum binding. In both the presence and absence of serum, dodecacilagin (**21**) was consistently the most active analog against the assayed pathogen spectrum. In the presence of serum, dodecacilagin was 16 times more potent than ciligacin. Additionally, dodecacilagin evades antibiotic resistance development even after constant long-term exposure. Of the analogs we synthesized, we believe dodecacilagin (**21**) is the most appealing candidate for further therapeutic development.

When using the synBNP approach, one must consider that due to the limitations of existing bioinformatic algorithms, synBNPs may not always be perfect copies of what is produced by a BGC *in vivo*. With the technology currently available, it is especially difficult to predict the naturally incorporated lipid substituent in lipopeptides. Therefore, a critical step in any synBNP study is post hit optimization. These efforts are crucial to identifying optimized synBNP structures because, as we prove here, not only does dodecacilagin represent an improved clinical candidate due to its unique balance of potency and serum binding, but its optimization process demonstrates a successful improvement of the synthetic natural product inspired by the *cil* BGC.

CHAPTER 4. CONCLUSION AND FUTURE DIRECTIONS

4.1 Conclusion

As emphasized in Chapter 1, illnesses and deaths related to antibiotic resistance infections remain a critical public health risk in the United States and in the world at large⁸⁷. To combat this threat, we must maintain an arsenal of novel antibiotic chemical agents that can be used clinically to prevent the spread of such infections. For nearly 100 years, bacterial natural products have been the go-to source for antibiotic molecules with unique modes of action^{7, 9, 11}. While currently most of these molecules have pathogenic strains that have developed resistance to their antibiotic effects, the pool of genetically encoded molecular natural products from which we can discover new molecules remains largely untapped. Scientists have scratched the surface of natural product discovery by historically only isolating those molecules that are highly expressed in bacterial culture that can be grown in laboratory conditions^{12, 14, 16, 26}. However, the new frontier of genomic sequencing and bioinformatic genome mining has revolutionized this field^{27, 88-90}. Using bioinformatic tools we can identify BGCs and predict the structures of their biosynthesized natural product molecules *in silico* and follow up this prediction with total synthesis to build these molecules independent of a bacterial culture^{42, 45, 91}. This synBNP pipeline has been an efficient and fruitful means to discover novel antibiotic molecules from silent and cryptic BGCs that can become lead candidates for drug development and clinical trials. The synBNP approach is what led to the discovery of the dual polyprenyl phosphate binding Gram-positive resistance evading antibiotic cilagicin⁴⁶. It is this innovative discovery approach and this promising novel antibiotic that has motivated the research presented in this thesis.

In Chapter 2, I described a discovery study in which we sought to identify additional naturally occurring members of the cilagicin family of antibiotics, with the intention of expanding this promising family of molecules with a unique mechanism of action that endows these molecules with the ability to evade antibiotic resistance development⁶³. We did this using the synBNP method, though we adapted our search parameters to drive BGC discovery based on final molecular product structure rather than BGC nucleotide sequence as was used initially to discover cilagicin. We built our database for BGC discovery by pulling whole sequenced bacterial genomes from publicly available databases, then using bioinformatic algorithms to identify NRPS BGCs within those genomes. Additional algorithms were used to make predictions of the linear peptide sequences of the NRPS product made by the BGCs. This list of polypeptide predictions became the database we then queried with the linear peptide sequence of cilagicin to identify previously uncharacterized BGCs predicted to produce molecules that were >50% similar in sequence to cilagicin. This bioinformatic pipeline yielded three BGCs that encoded three structurally unique previously unidentified molecules we called paenilagicin, bacilagicin, and virgilagicin. Using SPPS to synthesize synBNPs of these predictions, we then assayed their bioactivity, molecular target binding ability, and resistance evasion. We found that paenilagicin and virgilagicin bound C55:P and C55:PP as cilagicin does, while bacilagicin only binds C55:P. This difference in target binding affinity was reflected in each antibiotic's ability to evade resistance development. Paenilagicin and virgilagicin evaded antibiotic resistance as cilagicin does, which was

expected for dual binding compounds. Meanwhile bacilagicin showed resistance development after a few days of serial culture passage, as is consistent with other single polyprenyl binding antibiotics^{70, 71}.

This discovery study identified new members of the cilagicin family of antibiotics, which will provide variations of polyprenyl binding structures that can be considered when designing an optimal clinical candidate from this family of molecules. Additionally, identifying multiple members of this family allowed us to identify the conserved “DGnxGY” motif that we believe is important for polyprenyl phosphate target engagement, and can be used as a lead sequence in future discovery studies. Overall, this study exemplified the flexibility of the synBNP methodology, as adapting our BGC search approach to compare structural output of NRPS BGCs, rather than nucleotide sequenced based homology search, allowed us to find the clusters responsible for the molecules we identified in this study which we would not have identified otherwise.

synBNP is a robust and adaptable means of natural product discovery, but based on current algorithmic predictive capabilities, there are limitations to the predictions that can be made. Currently, our predictive capabilities for polypeptides are highly accurate, however prediction for other chemical moieties such as lipid substituents remains sub-optimal. In Chapter 3, I describe a medicinal chemistry optimization approach in which we synthesized a library of cilagicin analogs incrementally modifying single molecular moieties of cilagicin, as it is the molecule with the highest antibiotic activity in its family. This optimization study was motivated by the goal of finding a variation of cilagicin that would show high antibiotic bioactivity and high *in vivo* bioavailability by demonstrating minimal serum protein binding affinity. We did this optimization study in two parts, first analyzing the effects of structurally diverse lipid tail substituents on our desired biological properties. We then did the same for the polypeptide region of cilagicin, in which we singularly replaced amino acids in a series of orthogonal scans to analyze the effects of changing different chemical properties in different regions of the cyclized polypeptide core. After combining lipid and peptide modified elements, we identified our optimized molecule and called it dodecacilagicin. This compound maintains the same polypeptide core as cilagicin with a dodecanoic acid tail. Biological assays confirmed dodecacilagicin maintains C55:P and C55:PP as binding targets and, like its predecessor, evades antibiotic resistance. This study not only presents an optimized clinical drug development candidate from the cilagicin family of resistance evading antibiotics, but it demonstrates the need for post-prediction optimization of synBNP molecules that is required due to current limitations in the predictive capabilities of bioinformatic structure prediction algorithms.

Overall, the research presented in this thesis demonstrates the use of cutting-edge natural product discovery methods to expand a novel family of antibiotics with a unique mechanism of action. Additionally, we have identified an optimized lead compound that has been derived from this family and has strong potential for further development into an antibiotic clinical candidate.

4.2 Future Directions

The discoveries made in this body of work will continue to be carried forward in future research projects. At the time of writing this thesis, plans are underway to conduct

animal model studies with dodecacilagicin to advance its candidacy into the drug development and clinical trial pipeline. Such advancement would be tremendously meaningful as it is an antibiotic with potential to be a lifesaving treatment that could hopefully maintain clinical efficacy for a long time.

A significant portion of the work presented here resulted from the evolution and advancement of the synBNP method. Continued development and improvement upon this natural products discovery approach is the most promising factor when it comes to the future directions of this work. Early research with synBNPs only produced linear peptides comprised of canonical proteinogenic amino acid residues^{34, 92, 93}. Advancement in BGC identification and natural product prediction enabled the discovery of more complex products such as cyclized lipopeptides, like those presented in this thesis, and products containing non-proteinogenic amino acids^{44, 45}. With each generation of technological improvement, the discovery capabilities of the synBNP approach grow, and we expect this trend to continue.

The synBNP method relies on three key steps; bioinformatic BGC identification, molecular prediction, and total chemical synthesis. Each of these steps is reliant on technological advancements that will only improve the synBNP pipeline as they themselves develop. Increases in the speed and accuracy of sequencing technologies will build and refine the bacterial genome databases from which we derive our raw sequencing data³². Improved bioinformatic algorithms will increase the accuracy and complexity of BGCs that we can identify in bacterial genomes, as well as improving the quality of our molecular product predictions, within and beyond just polypeptide moieties in NRPS/NRPS-hybrid BGCs. And finally, continually improving chemical synthesis techniques will enhance our synthetic abilities to produce structurally complex and challenging molecular predictions. Our lab has already begun developing synBNP discovery pipelines for identifying BGCs responsible for producing bioactive natural products with complex aromatic lipid tail moieties beyond the straight chain lipid tails found on lipopeptides published in previous discovery studies.

All of these elements together speak very encouragingly to the longevity and the improvement of the synBNP approach. We are very optimistic that this robust methodology will evolve alongside other technologies and will enable discovery of increasingly novel and complex bioactive natural products that were once believed to be inaccessible.

CHAPTER 5. METHODS

5.1 Methods for “Structure Based Genomic Discovery of Naturally Occurring Cilagicin Analog Antibiotics”

5.1.1 General Experimental Procedures and Materials

All reagents and solvents were purchased from commercial sources and used without further purification. Solvents used for chromatography were HPLC grade or higher. Preparative HPLC was performed on a CombiFlash EZ Prep purification system with UV detection and equipped with a Phenomenex Luna 5 μ m C18 prepHPLC column using a dual solvent system (A/B: water/acetonitrile, supplemented with 0.1% (v/v) formic acid). HRMS and MS/MS data were acquired on a SCIEX ExionLC UPLC coupled to an X500R QTOF mass spectrometer, equipped with a Phenomenex Kinetex PS C18 100 Å column (2.1 x 50 mm, 2.6 μ m) and operated by SCIEXOS software. ¹H NMR and ¹³C NMR spectra were acquired at room temperature on a Bruker Avance DMX 600 MHz spectrometer (The Rockefeller University, New York, NY) equipped with cryogenic probes and spectra were analyzed using MestReNova software (version 14.3.0-30573). Chemical shift values were reported in ppm and referenced to residual solvent signals, for ¹H NMR: DMSO-d₆ = 2.54 ppm; for ¹³C NMR: DMSO-d₆ = 40.45 ppm.

5.1.2 Identification and bioinformatic analysis of natural cilagicin biosynthetic gene clusters

Sequenced nonribosomal peptide synthetase (NRPS) BGCs were collected from the bacterial genome databases JGI and GenBank. BGCs without clearly defined starting (condensation start (C_{start}) or CoA Ligase (CAL)) and ending (thioesterase (TE)) domains were removed from the collection. For the remaining complete sequenced NRPS BGCs, the 10 amino acids that make up each adenylation domain binding pocket (*i.e.*, amino acids 235, 236, 239, 278, 299, 301, 322, 330, 331, and 517) were identified using a curated list of A-Domain substrate signatures to predict the substrate of each BGC A-domain. These A-Domain signatures allowed us to make a linear polypeptide sequence prediction for each NRPS BGC in the collection. Using the linear polypeptide sequence of cilagicin as a query term, we ranked NRPS BGCs by their linear polypeptide sequence similarity to cilagicin. BGCs in the resultant list that shared $\geq 50\%$ polypeptide sequence similarity to cilagicin we deemed congeners. This search yielded two BGCs from *Paenibacillus mucilaginosus* genomes that shared 100% polypeptide sequence identity to cilagicin. Three BGCs were found to share 7 or more ($\geq 58\%$) of the 12 amino acid positions in cilagicin. The remaining BGCs from the search shared 5 or fewer ($\leq 42\%$) amino acid positions with cilagicin. The three BGCs with $\geq 50\%$ polypeptide sequence similarity to cilagicin were carried forward as the congener BGCs investigated in this study.

5.1.3 Solid Phase Peptide Synthesis

Natural cilagicin analogs characterized in this study were synthesized using standard Fmoc-based solid-phase peptide synthesis (SPPS) methods on 2-chlorotrityl chloride resin. All peptides were synthesized starting from the penultimate module's amino acid, for paenilagicin this was ornithine, for bacilagicin this was serine, and for and virgilagicin this was arginine. 2-chlorotrityl resin pre-loaded with the appropriate amino acid was swollen in DCM for 30 minutes at room temperature then drained and washed with DMF (3 mL, 3x). Subsequent couplings were carried out using Fmoc-protected amino acids (or a fatty acid) (3 equiv. relative to resin loading) mixed with HATU (3 equiv.) and DIPEA (3 equiv.) in DMF (5 mL). Each coupling reaction was carried out for 45 minutes at room temperature then washed with DMF (5 mL). Fmoc deprotection was carried out by treating resin-bound peptide with 20% piperidine in DMF (5 mL) for 5 minutes (2x). After deprotection, the resin was then washed with DMF (5 mL, 2x), DCM (5 mL, 2x), and DMF (5 mL, 2x). These steps were repeated for each amino acid and fatty acid to construct the linear peptides.

Ester bond formation: Ester bonds were formed between the unprotected threonine hydroxyl and the carboxylic acid of the final module's amino acid. For paenilagicin and bacilagicin this amino acid is tyrosine. For virgilagicin this amino acid is valine. The resin-bound peptide with a free hydroxyl group was mixed with the appropriate Fmoc-AA (15 equiv.) and DIC (15 equiv.) in 7mL DMF. DMAP (0.5 equiv.) was added to the solution, and gently shaken for ~16 hours at room temperature.

Peptide cyclization: Resin-bound linear peptides were cleaved by treating with 20% hexafluoroisopropanol (HFIP) in DCM for 1 hour (2x). Crude linear peptides were then collected by filtration and dried under reduced pressure. The cleaved linear peptides were cyclized without purification by resuspending in DMF to 0.002M and then mixing with PyAOP (7 equiv.) and DIPEA (20 equiv.). After 2 hours, reaction was transferred to a separatory funnel and ethyl acetate (2.5x volume of DMF) was added. This organic layer was washed with saturated brine (4x), then dried over sodium sulfate. Dried organic layers were filtered and concentrated under reduced pressure to yield crude cyclized peptide.

Bulk deprotection: Peptides were dissolved in 6 mL of cleavage cocktail (95% (v/v) TFA, 2.5% (v/v) TIPS and 2.5% (v/v) water) for 1.25 hours. Cleavage cocktail was evaporated under air flow to yield crude deprotected peptides.

Peptide purification: Crude peptides were purified on an Phenomenex Luna 5 μ m C18 prepHPLC column attached to a CombiFlash EZ Prep purification system using a dual solvent system (A/B: water/acetonitrile, supplemented with 0.1% (v/v) formic acid). Peptide purity and identity were confirmed by UPLC, HRMS, and NMR.

5.1.4 Minimum inhibitory concentration (MIC) assay

MIC assays were conducted using the protocol recommended by the Clinical and Laboratory Standards Institute.⁶⁹ Culture conditions (temperature, medium) are detailed in Supplementary Table S6. All compounds were dissolved in sterile DMSO (ATCC, USA) to give a concentration of 6.4 mg/mL. Tested compounds were serially diluted 2-fold in DMSO from a maximum stock concentration of 6.4 mg/mL to 0.006 mg/mL. In a 96-well plate filled with 49 μ L fresh growth medium, 1 μ L of compound stock dilution was added

across wells in a row. An overnight culture of an assay strain was diluted 5,000-fold in fresh medium. 50 μL of this inoculum dilution was added into each well, giving a final volume of 100 μL per well. Final assayed concentrations of test compounds ranged from 64 $\mu\text{g}/\text{mL}$ to 0.06 $\mu\text{g}/\text{mL}$. MIC values were recorded as the minimum concentration at which no bacterial growth appeared, based on visual inspection, after 16 hours of static incubation at 37 °C. *Clostridium difficile* plates were statically incubated under anaerobic conditions (Vinyl anaerobic chamber, 37 °C, 5% H₂, 5% CO₂, 90% N₂). MICs were performed in technical duplicate (n=2) and repeated three independent times (n=3).

5.1.5 Cytotoxicity assay

The cytotoxicity of natural cilagicin analogs were tested using an MTT (3-(4,5-Dimethyl-2-thiazolyl)-2,5-diphenyl-2H-tetrazolium bromide) assay. HEK293 cells were seeded in a 96-well plate with a density of 5,000 cells/well and cultured in Dulbecco's Modified Eagle Medium (DMEM) without phenol red and supplemented with 10% fetal bovine serum, 1% Pen/Strep and 1% glutamate for 24 hours at 37 °C with 5% CO₂. Serially diluted compounds were added into each well at a final concentration ranging from 64 $\mu\text{g}/\text{mL}$ to 0.06 $\mu\text{g}/\text{mL}$. After 48 hours of incubation, the media was removed and 15 μL of freshly prepared MTT solution (5 mg/mL in DPBS) was added to each well. The plates were incubated for 3 hours at 37 °C with 5% CO₂ after which the MTT solution was removed by aspiration. Precipitated formazan crystals were dissolved by addition of 100 μL of solubilization solution (40% DMF, 16% SDS and 2% acetic acid in H₂O). The absorbance of each well was measured at OD570nm using a microplate reader (Infinite 200 PRO, Tecan). All experiments were performed in duplicate (n=2) and repeated three independent times (n=3).

5.1.6 Undecaprenyl phosphate feeding assay

The effect of cell wall phospholipids undecaprenyl phosphate (C55:P) and undecaprenyl pyrophosphate (C55:PP) on natural cilagicin analogs' antibacterial activity was evaluated by co-drying peptide and lipid at molar ratios from 0x to 5x, at 0.5x increments, in plastic tubes *in vacuo* for 2 hours to completely remove all organic solvent. After drying, compounds were resuspended in scant 2.5 μL methanol, then in 50 μL fresh LB to bring the peptide concentration to 128 $\mu\text{g}/\text{mL}$, followed by vigorous sonication and vortexing. 25 μL of this solution was transferred in duplicate to a 384 well plate and serially diluted 2-fold in LB medium from 128 $\mu\text{g}/\text{mL}$ to 0.012 $\mu\text{g}/\text{mL}$. An overnight culture of *S. aureus* USA300 was diluted 5,000-fold in fresh LB medium. 12.5 μL of this inoculum dilution was added to each well, giving a final volume of 25 μL per well. Final assayed concentrations of test compounds ranged from 64 $\mu\text{g}/\text{mL}$ to 0.06 $\mu\text{g}/\text{mL}$. MIC values were recorded as the minimum concentration at which no bacterial growth appeared, based on visual inspection, after 16 hours of static incubation at 37°C. All assays were run in duplicate (n=2) and repeated two independent times (n=2).

5.1.7 Evaluating antibiotic resistance by serial passage in liquid broth

A single colony of *S. aureus* USA300 was inoculated in 5 mL LB broth and grown overnight at 37 °C with continuous shaking (200 rpm). The overnight culture was then diluted 1:5,000 into fresh LB medium. 50 µL aliquots of dilute cells were transferred into individual wells of 96-well plates containing 50 µL of serially diluted cilagicin, natural cilagicin analogs, and amphomycin (Cayman Chemical Company, USA), in accordance with the standard MIC assay set up described above. *Note: stock dilutions of test compounds were prepared fresh daily.* Plates were statically incubated at 37 °C. After 24 hours, the MIC was recorded. For the next round of assays, an aliquot from the culture well at half of the MIC from the previous day's MIC plate was diluted 5000-fold in fresh LB and mixed with serially diluted antibiotics. The MIC was determined as described above. This process was repeated daily for 14 days. For amphomycin, LB medium was supplemented with 100 µg/mL CaCl₂·2H₂O. Experiments were performed three independent times (n=3).

5.2 Methods for “Designing an Optimized Bioactive and Bioavailable Clinical Development Candidate Inspired by Cilagicin”

5.2.1 General Experimental Procedures and Materials

All reagents and solvents were purchased from commercial sources and used without further purification. Solvents used for chromatography were HPLC grade or higher. Preparative HPLC was performed on a CombiFlash EZ Prep purification system with UV detection and equipped with a Phenomenex Luna 5 µm C18 prepHPLC column using a dual solvent system (A/B = water/acetonitrile, supplemented with 0.1% (v/v) formic acid). High-resolution mass spectrometry (HRMS) and MS/MS data were acquired on a SCIEX ExionLC UPLC coupled to an X500R QTOF mass spectrometer, equipped with a Phenomenex Kinetex PS C18 100 Å column (2.1 mm × 50 mm, 2.6 µm), and operated by SCIEXOS software. ¹H NMR and ¹³C NMR spectra were acquired at room temperature on a Bruker Avance DMX 600 MHz spectrometer (The Rockefeller University, New York, NY) equipped with cryogenic probes, and spectra were analyzed using MestReNova software (version 14.3.0-30573). Chemical shift values were reported in ppm and referenced to residual solvent signals: for ¹H NMR, DMSO-d₆ = 2.50 ppm; for ¹³C NMR, DMSO-d₆ = 39.52 ppm.

5.2.2 Solid-Phase Peptide Synthesis

Fmoc-based solid phase peptide synthesis (SPPS) was used to build all molecules evaluated in this study. Using 2-chlorotrityl chloride resin, each linear peptide was synthesized starting with the amino acid in position 2 pre-loaded on to the resin. Resin was swelled in dichloromethane (DCM) for 30 minutes at room temperature, then drained and repeatedly washed with dimethylformamide (DMF) (3x, 3 mL). Linear assembly of Fmoc-protected amino acids 3 through 12 and the terminal lipid tail were conducted using the following method: 3 equivalents (equiv) (relative to resin) Fmoc-amino acid or fatty acid and hexafluorophosphate azabenzotriazole tetramethyl uronium (HATU) (3

equiv) were dissolved in DMF (3 mL), then N,N-diisopropylethylamine (DIPEA) (3 equiv) was added. This coupling mixture was left to activate for 5 min. The activated coupling mixture was poured over resin and mixed at room temperature for 45 min. After coupling, the N-Fmoc of the most recently coupled amino acid was removed by mixing resin with 3 mL 20% piperidine in DMF for 5 min (2x). After deprotection, resin was washed thoroughly with DMF (3x, 3 mL), then DCM (2x, 3 mL), and again with DMF (2x, 3 mL). After washing, the next coupling reaction was carried out.

Ester bond formation: To form an ester bond, the amino acid in position 1 was coupled onto the free side chain of the threonine residue at position 11 of the completed acylated linear peptide still bound to resin. The appropriate position 1 Fmoc-amino acid (15 equiv) was dissolved in DMF (6 mL) and N,N'-diisopropylcarbodiimide (DIC) (15 equiv) was added. This solution was added to the resin-bound peptide, along with a catalytic amount of 4-dimethylaminopyridine (DMAP) (0.5 equiv). The reaction was allowed to shake at room temperature (22 °C) overnight (16 hour). The terminal N-Fmoc group was removed after esterification as described above.

Peptide cyclization: Esterified linear peptides were removed from resin using 20% hexafluoroisopropanol (HFIP) in DCM and mixing for 1 hour (2x). Crude peptides were obtained by drying the flow through under reduced pressure. To cyclize the linear peptides, the crude peptide and 7-azabenzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyAOP) (7 equiv) were dissolved in DMF to a concentration of 2 mM. DIPEA (20 equiv) was then added and the reaction was left for 2 hours at room temperature with stirring. The cyclized peptide was extracted from the reaction using ethyl acetate (2.5x reaction volume). The organic layer was washed with brine (5x), collected, and dried over sodium sulfate. The organic layer was filtered and concentrated under reduced pressure to yield the crude, protected cyclized peptide.

Trifluoroacetic acid (TFA) deprotection: Crude cyclized peptides were dissolved in 10 mL of a solution of 95% (v/v) TFA, 2.5% (v/v) triisopropylsilane (TIPS) and 2.5% (v/v) water for 1.25 hours to remove all acid labile side chain protecting groups. The solution was evaporated under air flow until completely dry.

Final product purification: Completed crude cyclic peptides were purified using a Phenomenex Luna 5 μ m C18 prep HPLC column attached to a CombiFlash EZ Prep purification system with a dual solvent system (A/B: water/acetonitrile, supplemented with 0.1% (v/v) formic acid). Compound purity and identity were confirmed by UPLC, HRMS, and/or NMR.

5.2.3 Determination of Synthetic Peptide Purity

Purified peptides were analyzed on a Waters Acquity H-Class UPLC. UV diode traces at 214nm were used to determine purity of peptide samples by calculating area under the curve for the main peptide sample peak then dividing by the area under the entire trace.

5.2.4 Minimum inhibitory concentration (MIC) assay

All MIC assays were executed following the protocol outlined by the Clinical and Laboratory Standards Institute⁶⁹. Synthetic peptides were dissolved in biology grade

DMSO (ATCC, USA) to a concentration of 6.4 mg/mL. Each stock solution was serially diluted 2x in DMSO from a maximum concentration of 6.4 mg/mL to a minimum concentration of 0.006 mg/mL. Assays were carried out in sterile clear 96-well plates. For each assay, 49 μ L of an appropriate bacterial growth medium was added to each well in a plate, and 1 μ L of a serially diluted stock solution was added across a row of the plate. An overnight culture of the bacterial pathogen being evaluated was diluted 5000x in appropriate growth medium and 50 μ L of this diluted inoculum was added to each well for a total assay volume of 100 μ L. Assay plates were statically incubated overnight (16 hour) at 37 °C. NOTE: *Clostridium difficile* assay plates were statically incubated under anaerobic conditions (vinyl anaerobic chamber, 37 °C, 5% H₂, 5% CO₂, 90% N₂). After incubation, MIC values were determined by visual inspection as the lowest concentration of the antibiotic at which there was no observed bacterial growth. All MICs were repeated in triplicate (n=3) with technical duplicates on each assay plate (n=2).

5.2.5 Human Serum Binding (MIC) assay

10% human serum (Sigma-Aldrich, USA) was added into the appropriate growth medium for each pathogen being assayed. MIC assays were executed using the assay protocol described above.

5.2.6 Undecaprenyl phosphate feeding assay

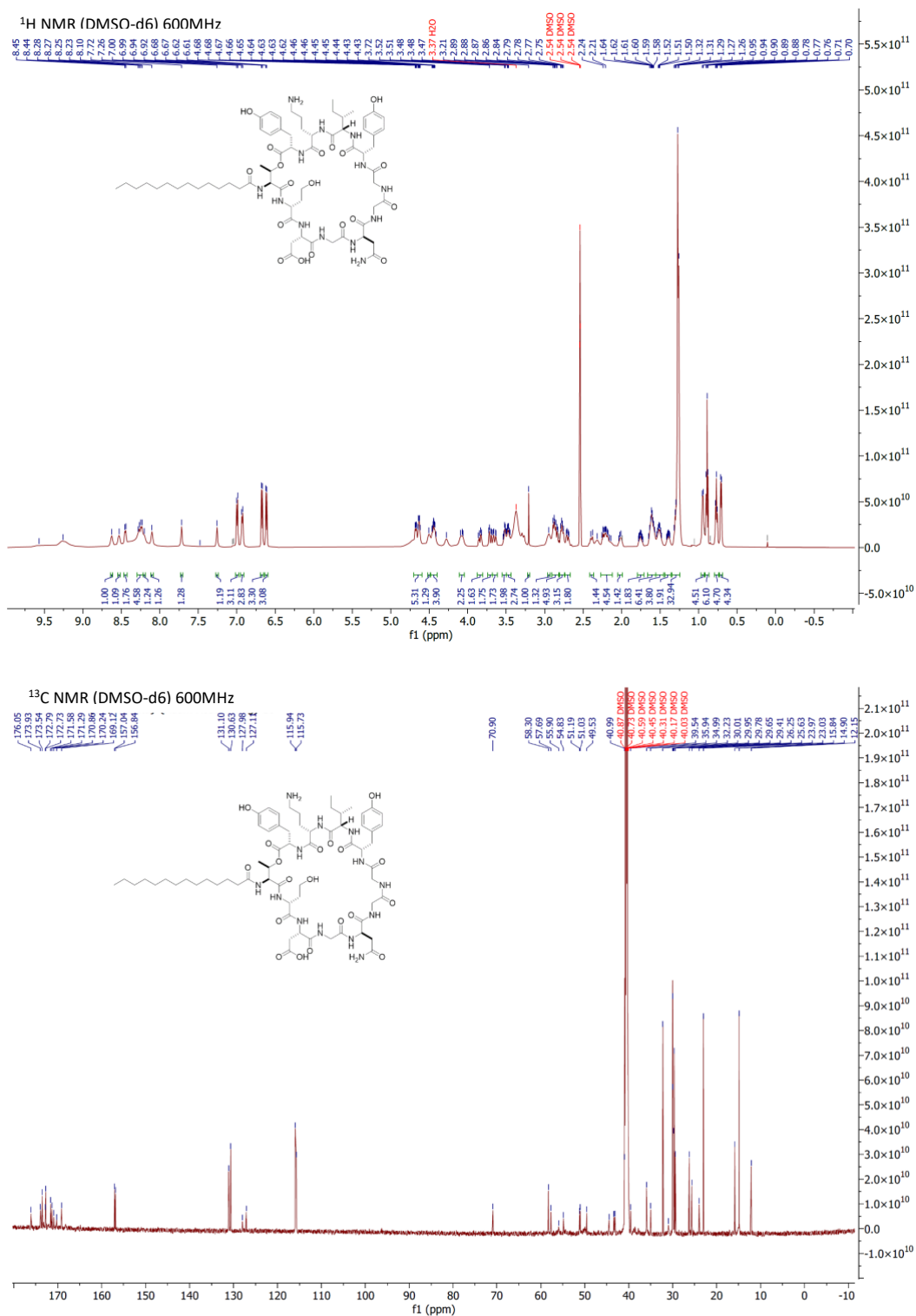
To investigate the interaction between undecaprenyl phosphates C55:P and C55:PP and an antibiotic, we included exogenous undecaprenyl phosphates into the MIC assay. To do this, a polyphosphorylated phosphate and an antibiotic were mixed at molar ratios ranging from 0 to 5 at intervals of 0.5. These mixtures were then dried in plastic tubes under vacuum for 2 hours to remove all organic solvents. Each dried mixture was resuspended in 2.5 μ L methanol and then 50 μ L Lysogeny broth (LB) was added to give a solution with a final antibiotic concentration of 128 μ g/mL. Resuspended solutions were aggressively sonicated and vortexed to ensure solubilization of all solution components. 25 μ L of these solutions were aliquoted into individual wells of a sterile 384-well microtiter plate. This stock mixture was serially diluted across the plate 2-fold in LB to give a final antibiotic dilution range of 128 μ g/mL to 0.012 μ g/mL. An overnight culture of *S. aureus* USA300 was diluted 5000x into fresh LB. 12.5 μ L of inoculum was distributed to each well of the assay plate, giving a final volume of 25 μ L and an antibiotic concentration range of 64 μ g/mL to 0.06 μ g/mL. 0 μ g/mL antibiotic was always included as a negative control. Assay plates were statically incubated overnight (16 hour) at 37 °C. After incubation, MIC values were determined by visual inspection as the lowest concentration of antibiotic at which there was no observed bacterial growth. All MICs were repeated in duplicate (n=2) with technical duplicates on each assay plate (n=2).

5.2.7 Evaluating resistance development by serial passaging in liquid broth

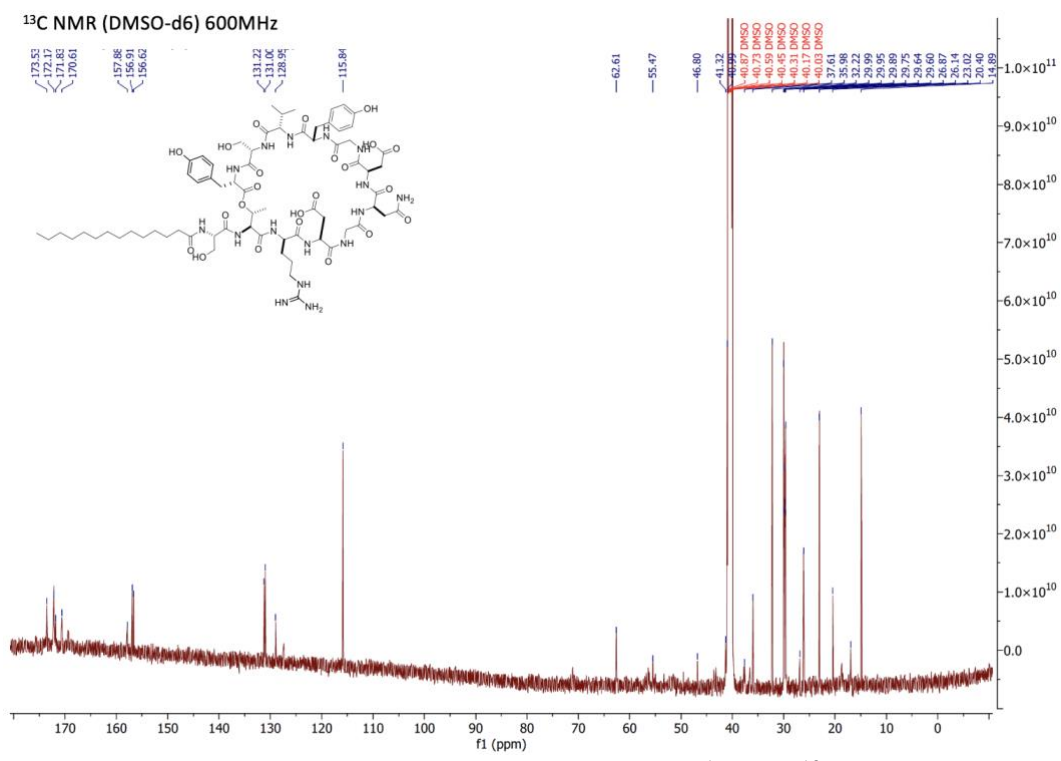
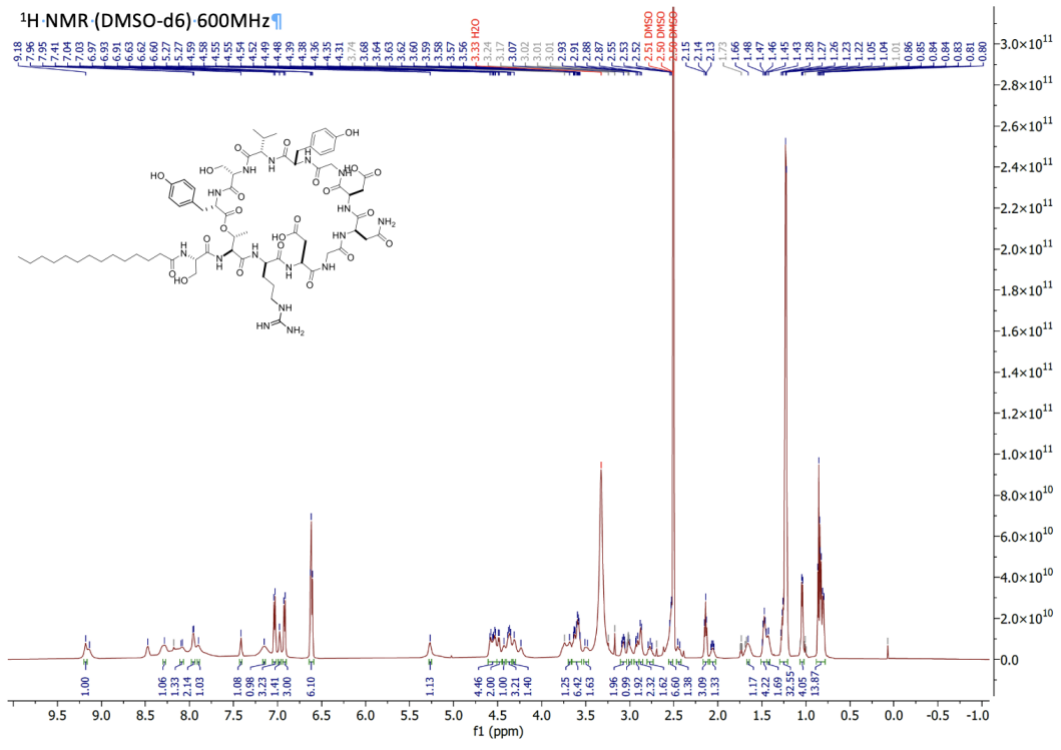
An overnight culture of *S. aureus* USA300 was grown by inoculating a single colony in 5 mL of Lysogeny broth (LB) and incubating overnight at 37 °C with continuous shaking (200 rpm). An initial MIC assay was then set up using cilagycin, dodecacilagycin, or amphomycin (Cayman Chemical Company, USA), following the standard MIC procedure described above. Stock dilutions of all compounds were prepared fresh daily. Plates were statically incubated at 37 °C. Each day, for the duration of the time course, we recorded the MIC (lowest antibiotic concentration where no bacterial growth was detectable), then from the well of highest concentration where cell growth was detectable (which we refer to as half-MIC) some of the bacterial culture was taken to be diluted 5000-fold in fresh Lysogeny broth and used as the inoculum for the next day's MIC with same serial dilution range in the appropriate antibiotics (cilagycin, dodecacilagycin, or amphomycin) following the same MIC assay protocol. This process was repeated daily for 10 days. For amphomycin, LB medium was supplemented with 100 µg/mL CaCl₂·2H₂O. Experiments were performed three independent times (n=3).

CHAPTER 6. APPENDIX

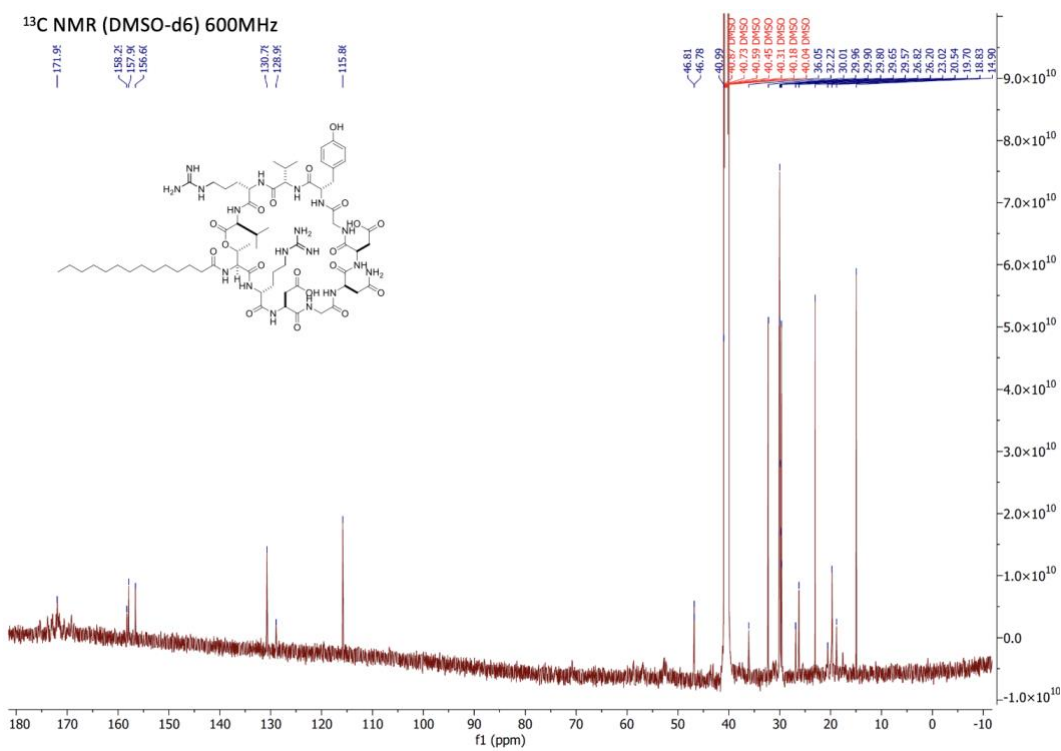
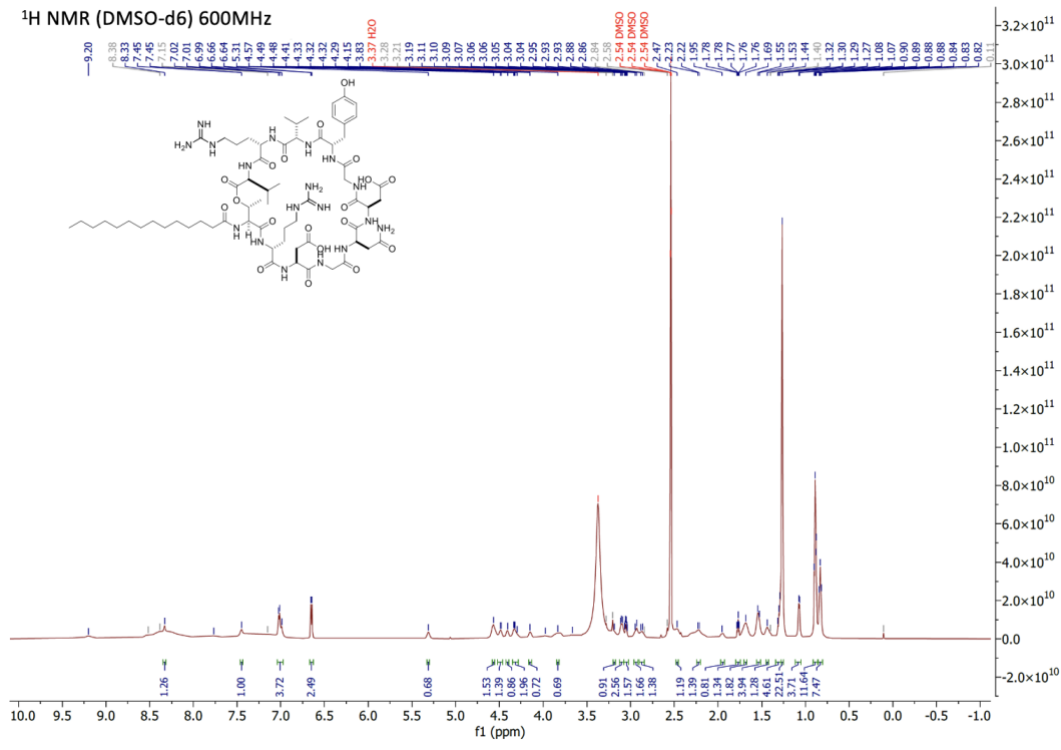
6.1 Appendix Figures for Chapter 2



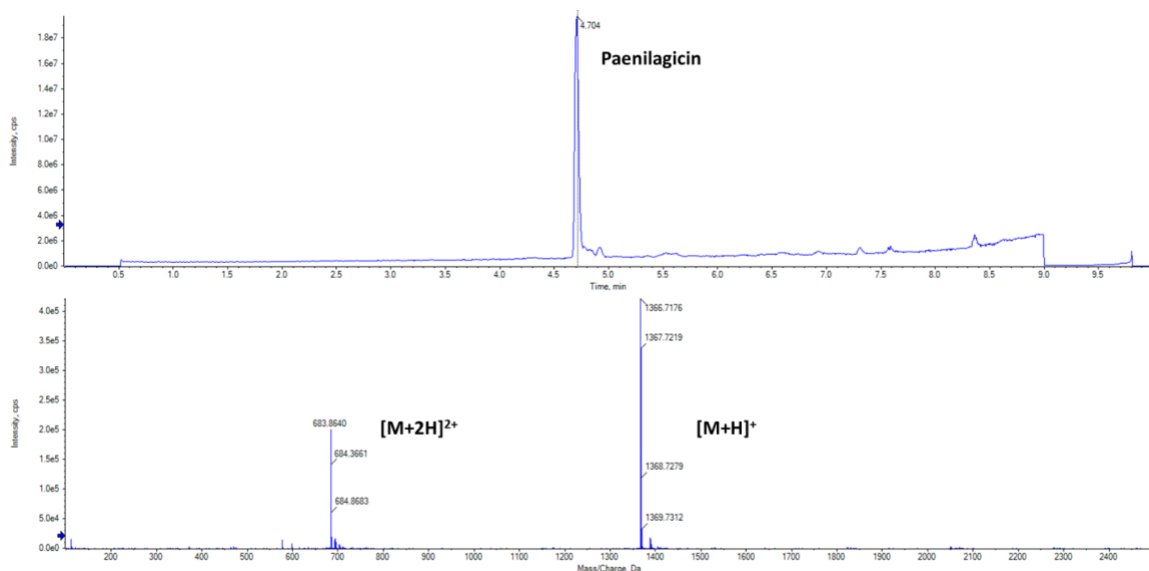
Appendix Figure 6.1. NMR spectra for paenilagicin. ¹H and ¹³C NMR taken in DMSO-d₆ at 600Hz



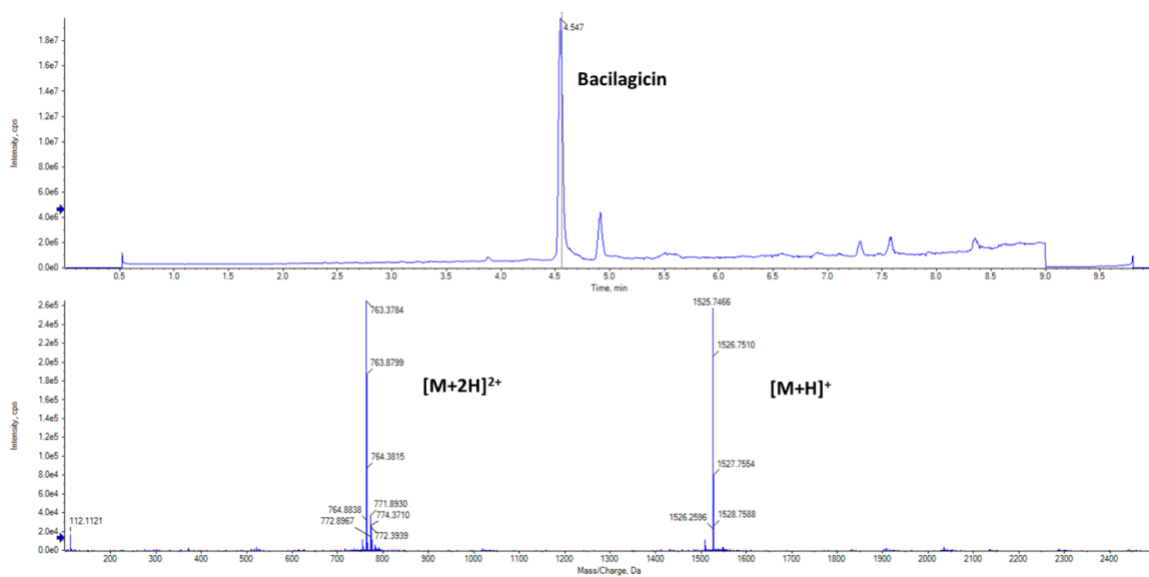
Appendix Figure 6.2. NMR spectra for baciligin. ¹H and ¹³C NMR taken in DMSO-d₆ at 600Hz



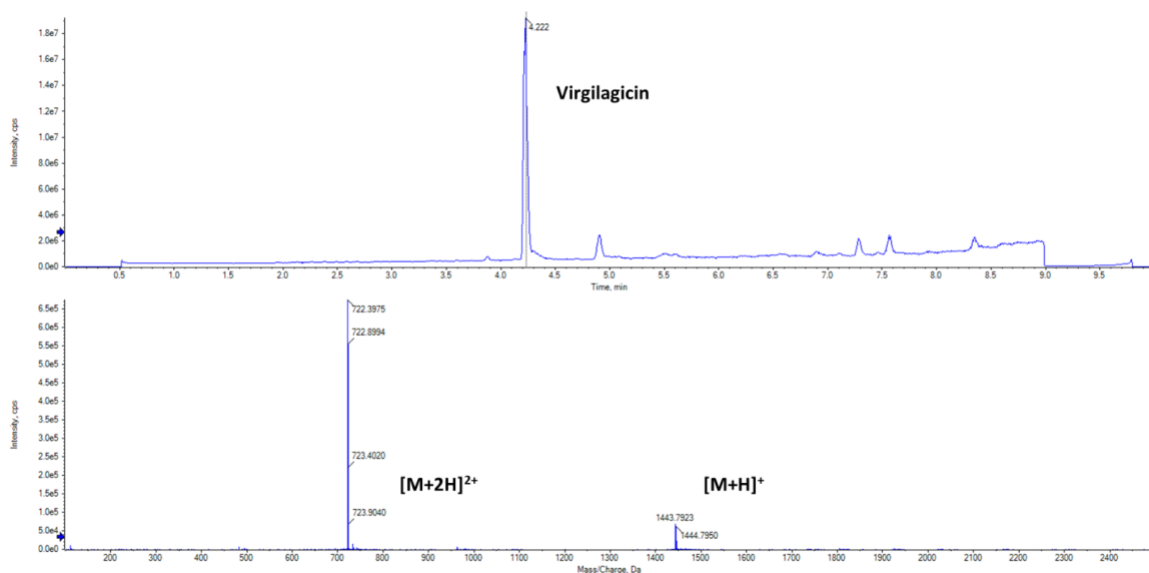
Appendix Figure 6.3. NMR spectra for virgilagicin. ¹H and ¹³C NMR taken in DMSO-d₆ at 600Hz



Appendix Figure 6.4. HRMS analysis of paenilagicin. HPLC chromatogram (top) and High-resolution mass spectra (bottom).



Appendix Figure 6.5. HRMS analysis of bacilagicin. HPLC chromatogram (top) and High-resolution mass spectra (bottom).



Appendix Figure 6.6. HRMS analysis of virgilagicin. HPLC chromatogram (top) and High-resolution mass spectra (bottom).

Molecules	Chemical Formula	Theoretical [M+H] ⁺	Observed [M+H] ⁺	Error (ppm)
Paenilagicin	C ₆₅ H ₉₉ N ₁₃ O ₁₉	1366.7258	1366.7176	1.7
Bacilagicin	C ₆₉ H ₁₀₄ N ₁₆ O ₂₃	1525.7538	1525.7466	2.2
Virgilagicin	C ₆₅ H ₁₀₆ N ₁₈ O ₁₉	1443.7960	1443.7923	1.5

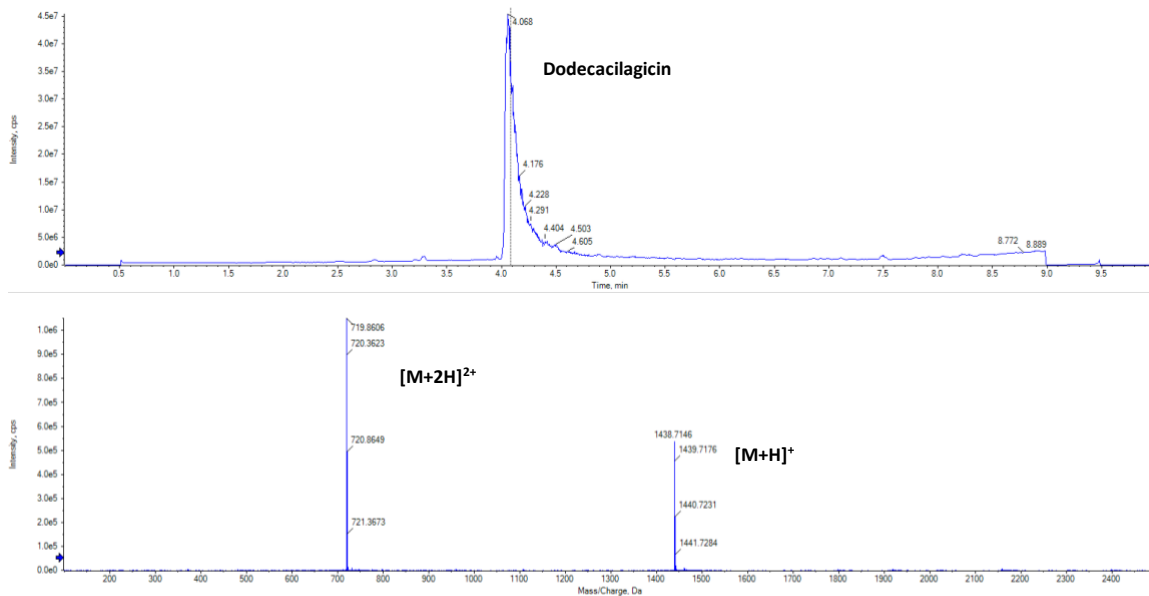
*All HRMS data were collected in positive ionization mode with a mass range from m/z 200 to 2500.

Appendix Figure 6.7: High-resolution mass spectrometry* data for natural cilagicin analogs

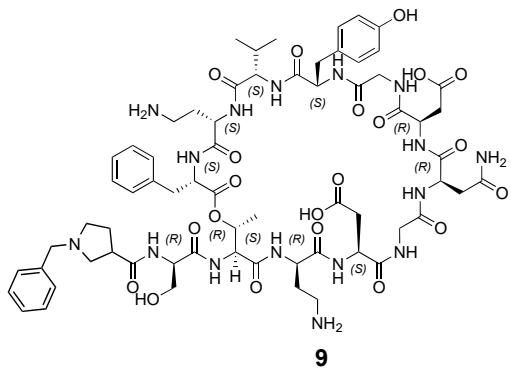
	Name of Cell Type	Strain	Media	Culture Condition
Bacterial Species	<i>Staphylococcus aureus</i>	USA300	LB	37 °C, aerobic
	<i>Enterococcus faecium</i>	AR807	LB	37 °C, aerobic
	<i>Enterococcus faecalis</i>	AR785	LB	37 °C, aerobic
	<i>Streptococcus agalactiae</i>	BAA2675	LB	37 °C, aerobic
	<i>Acinetobacter baumannii</i>	ATCC17978	LB	37 °C, aerobic
	<i>Escherichia coli</i>	ATCC25922	LB	37 °C, aerobic
	<i>Clostridium difficile</i>	HM746	LMB	37 °C, 5% H ₂ , 5% CO ₂ , 90% N ₂
	Human Cells	<i>Human Kidney Cells</i>	HEK293	DMEM, 10% Fetal Bovine Serum, L-Glutamine, Penicillin/ Streptavidin

*LBM media was a brain heart infusion media derivate which supplemented with 5 ug/mL hemin, 1 mg/mL maltose, 1 mg/mL cellobiose and 500 ug/mL L-cysteine.

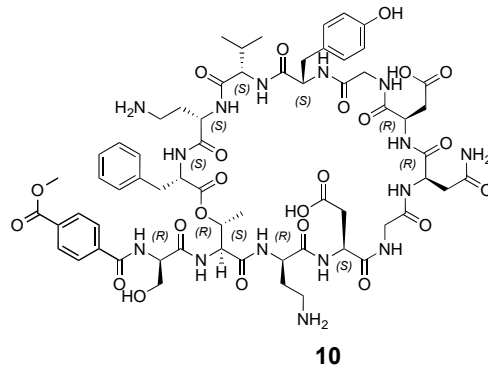
Appendix Figure 6.8: All species experimental culture conditions. Bacterial strains, human cell lines and corresponding culture conditions



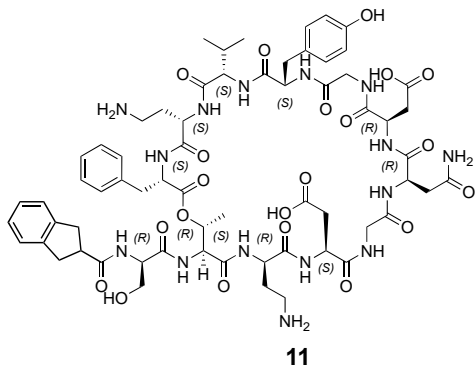
Appendix Figure 6.10. HRMS analysis of dodecacilagin. HPLC chromatogram (top) and High-resolution mass spectra (bottom).



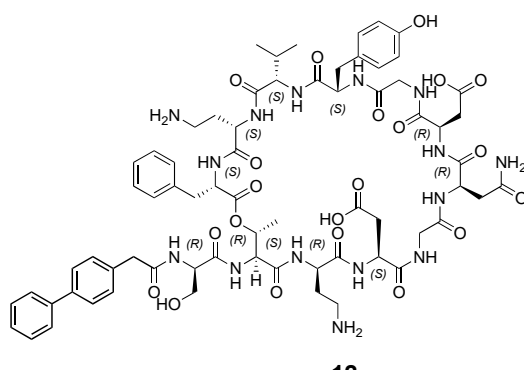
9



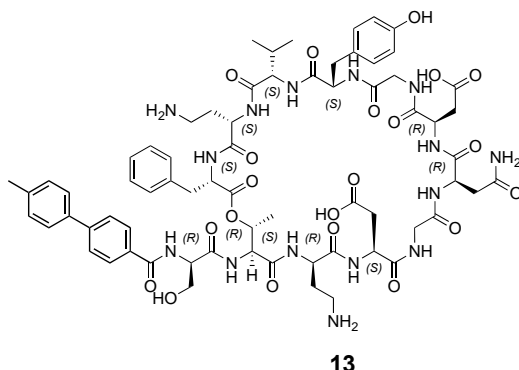
10



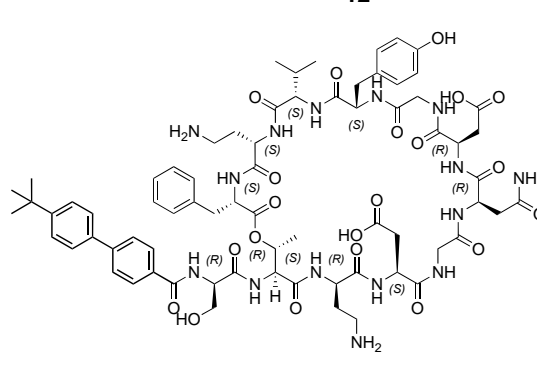
11



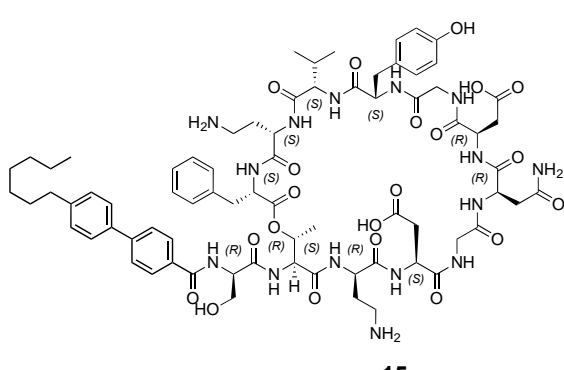
12



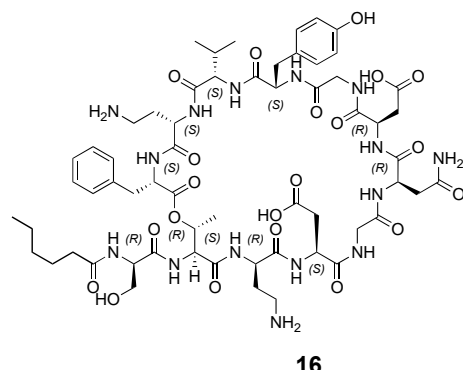
13



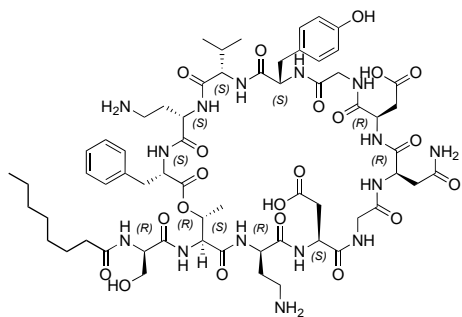
14



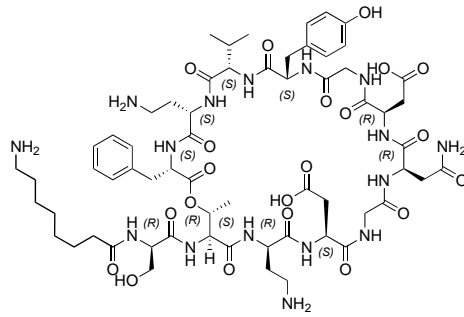
15



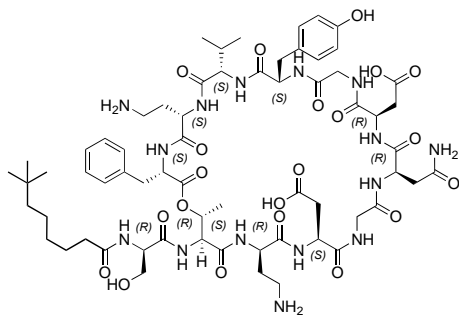
16



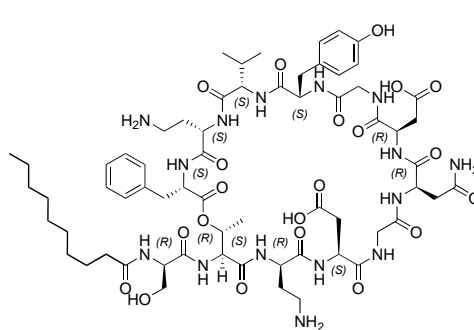
17



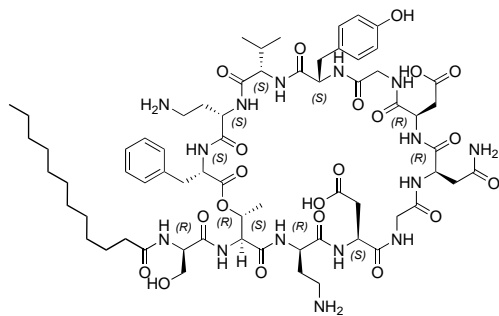
18



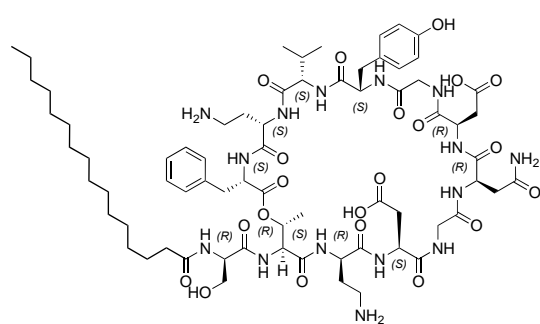
19



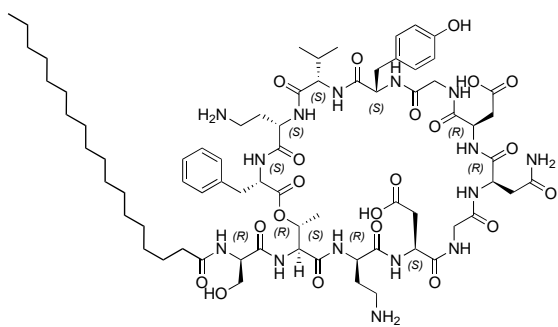
20



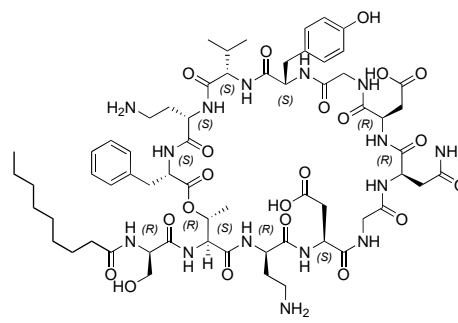
21



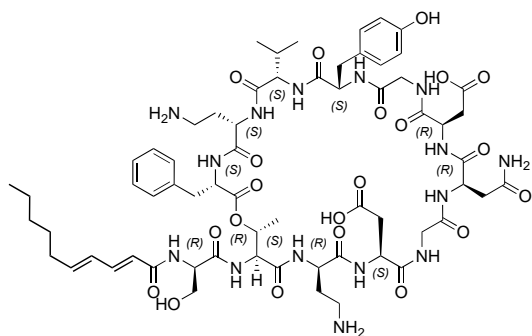
22



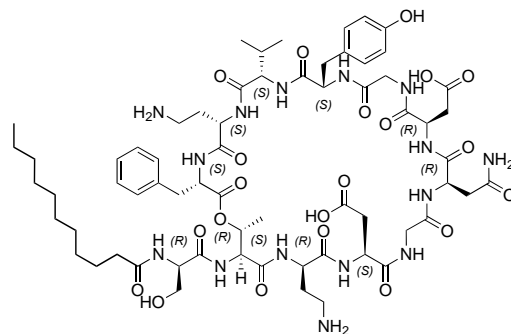
23



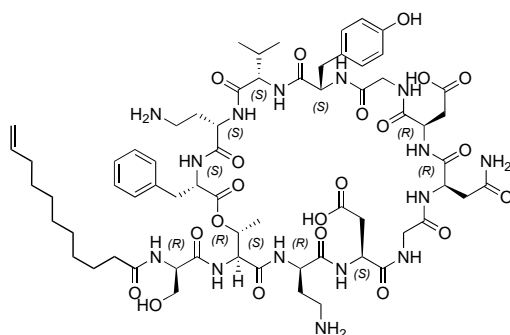
24



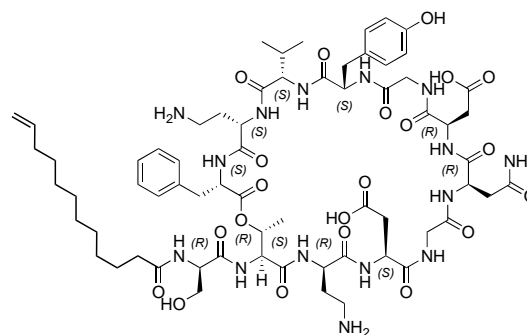
25



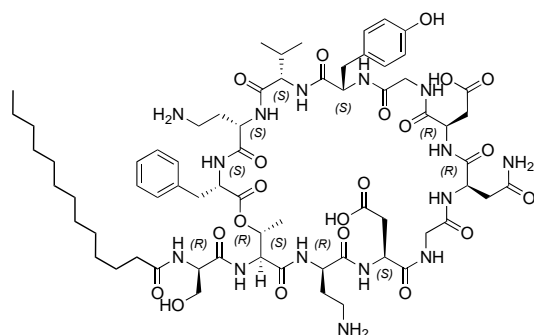
26



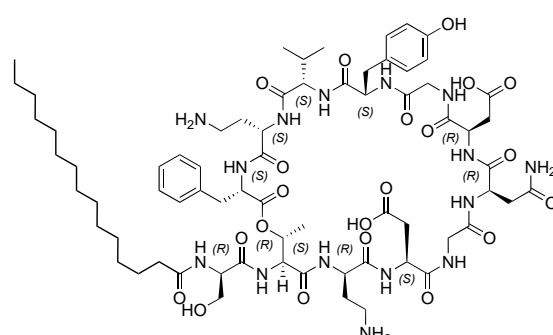
27



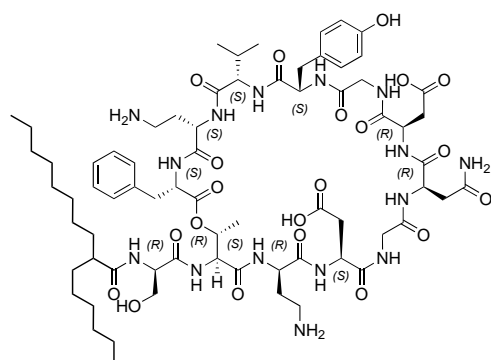
28



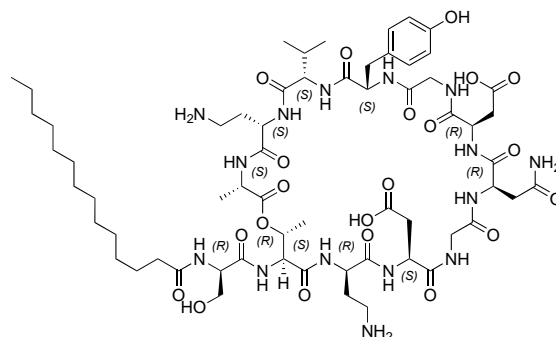
29



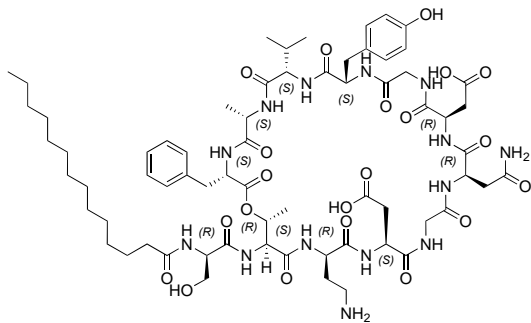
30



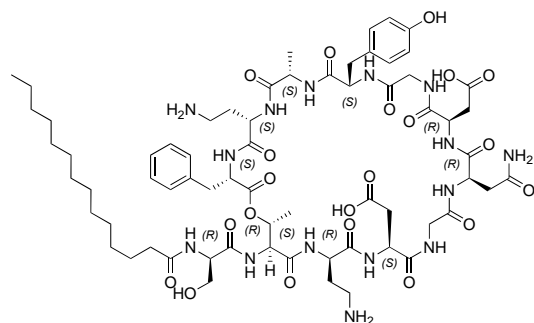
31



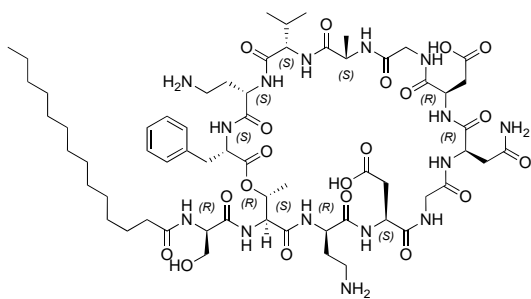
32



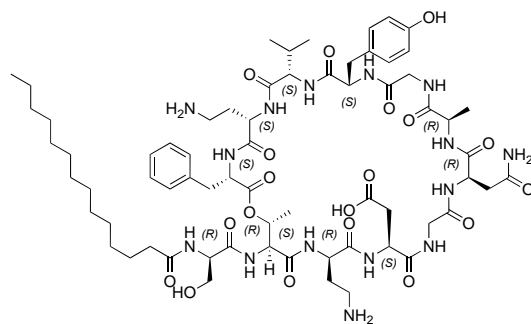
33



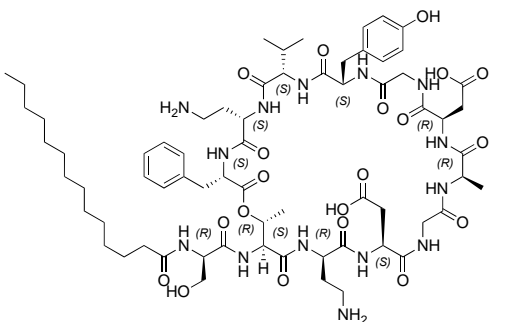
34



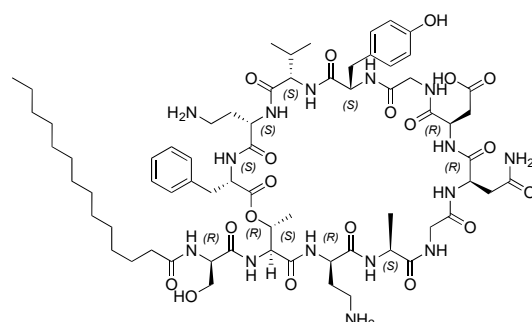
35



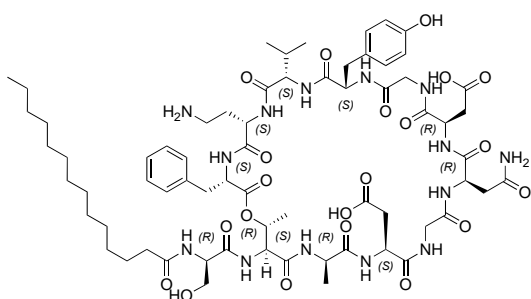
36



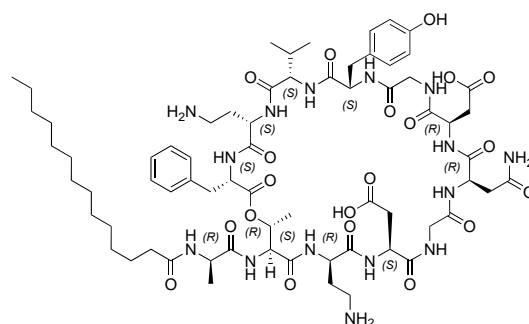
37



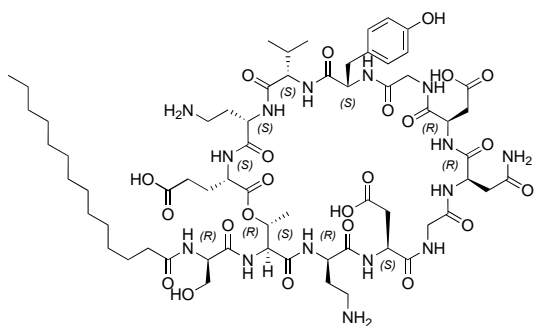
38



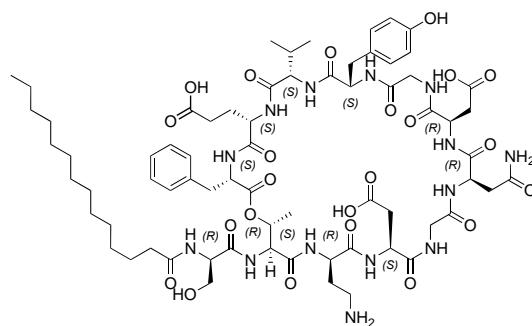
39



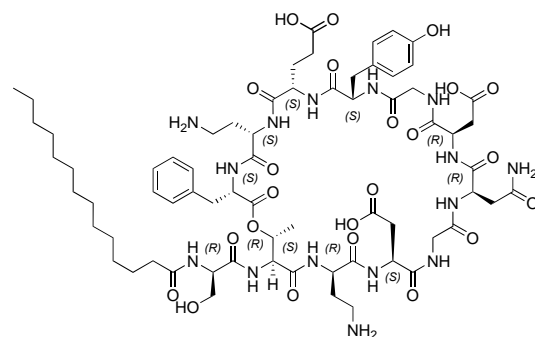
40



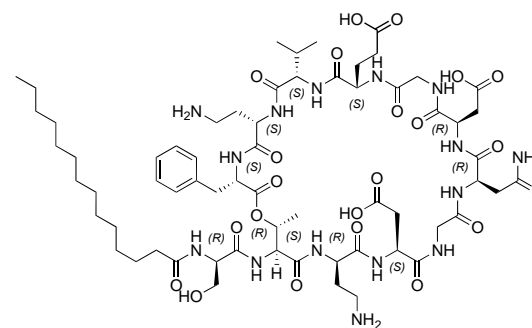
41



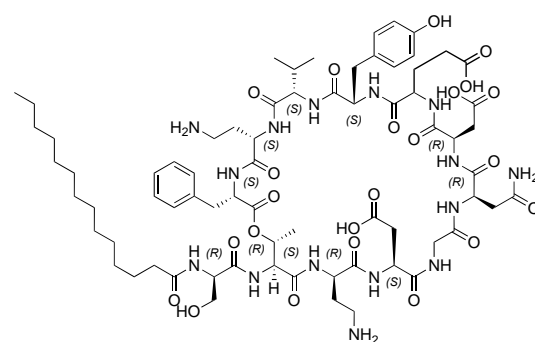
42



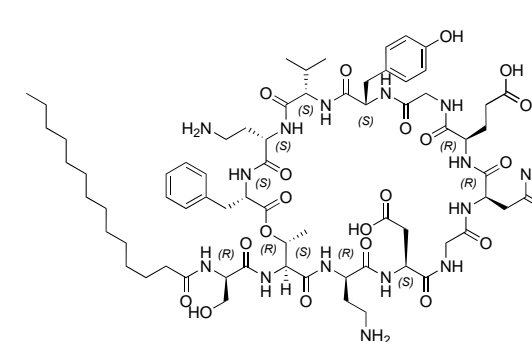
43



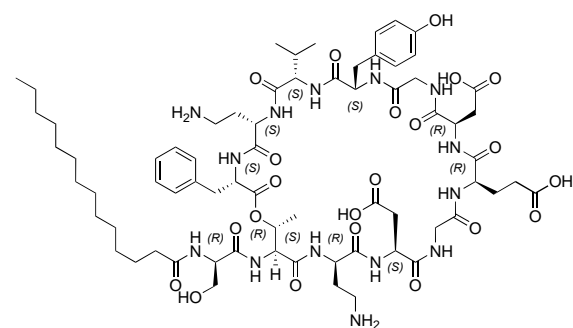
44



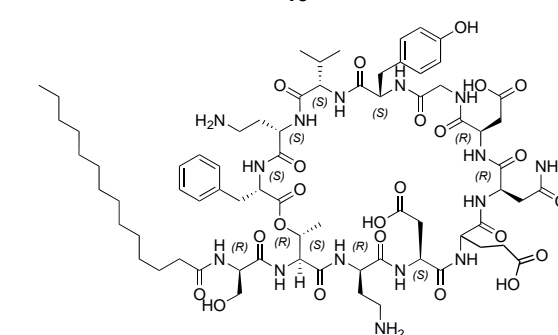
45



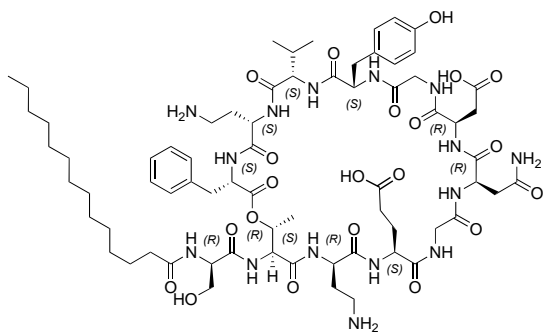
46



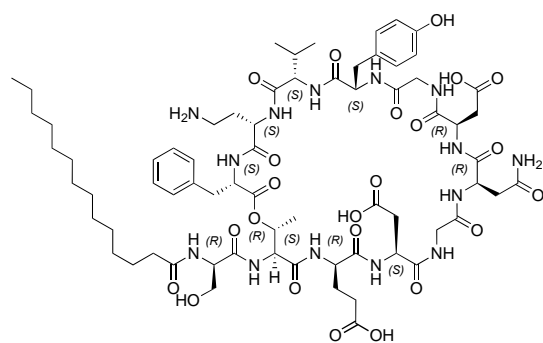
47



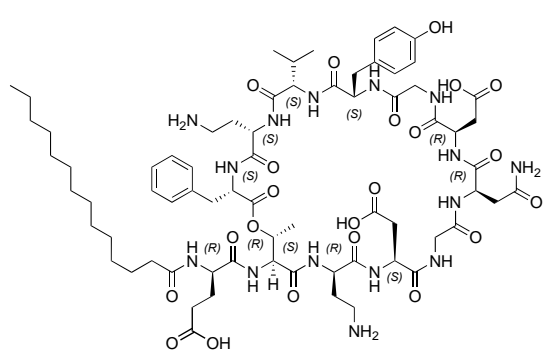
48



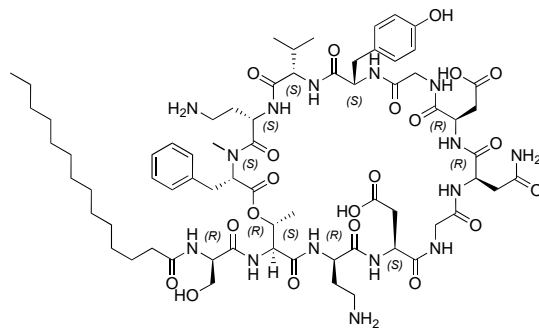
49



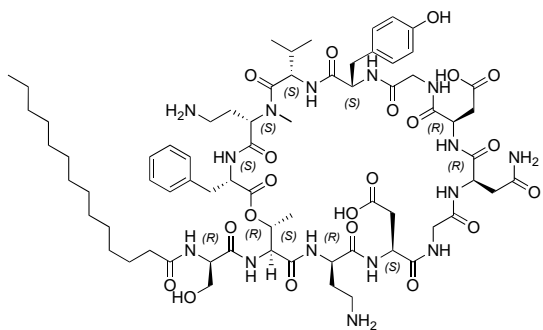
50



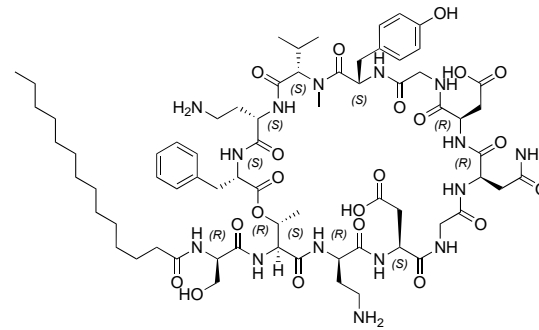
51



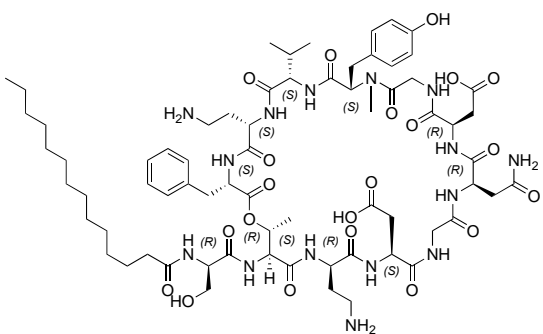
52



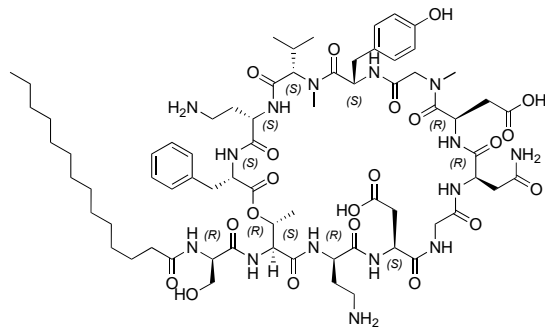
53



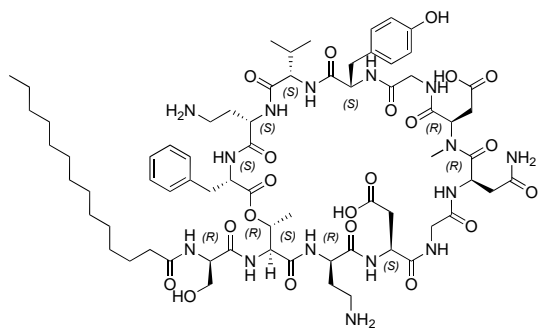
54



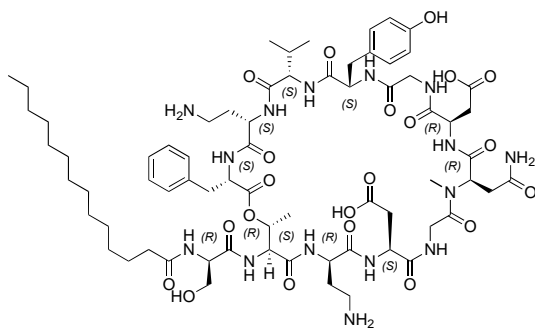
55



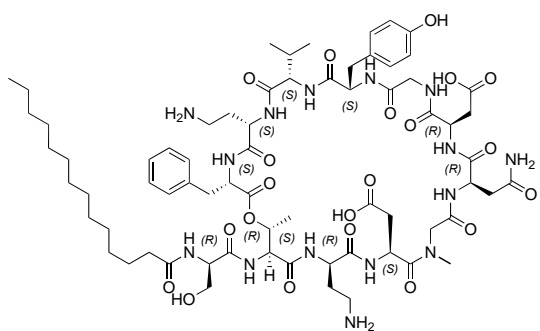
56



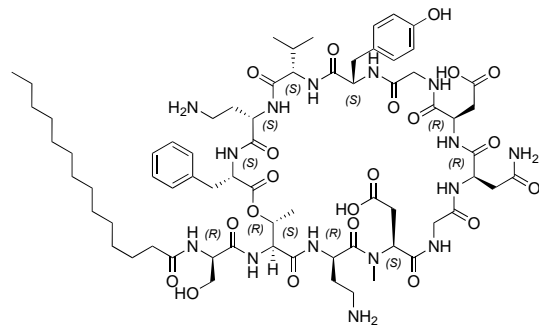
57



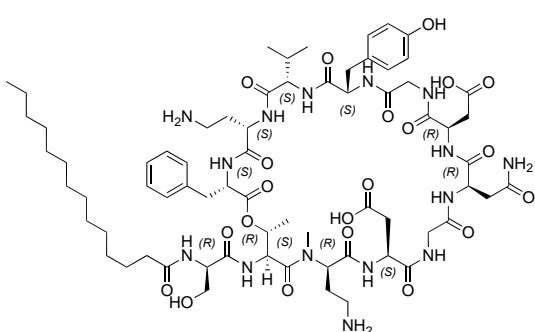
58



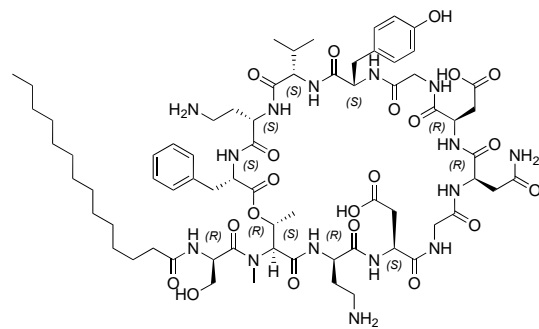
59



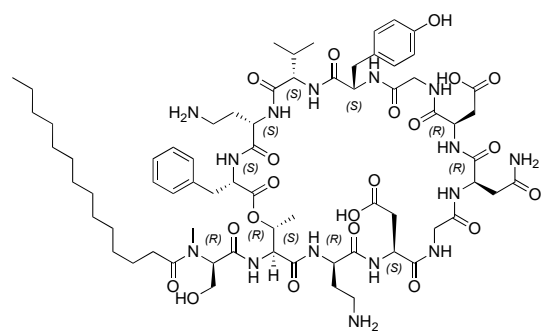
60



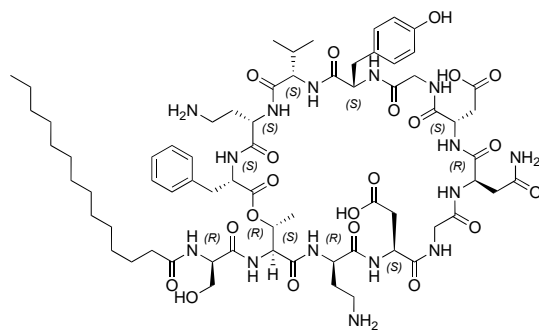
61



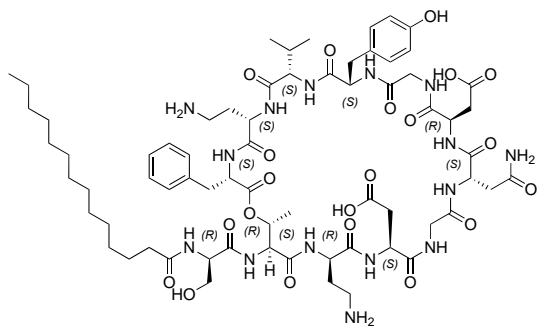
62



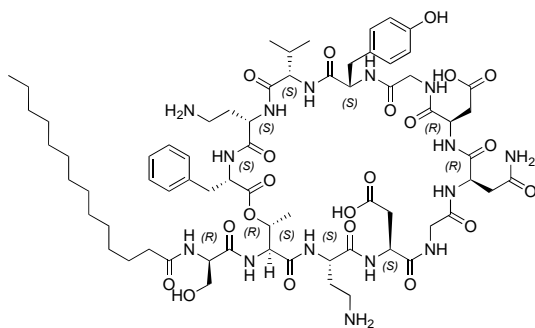
63



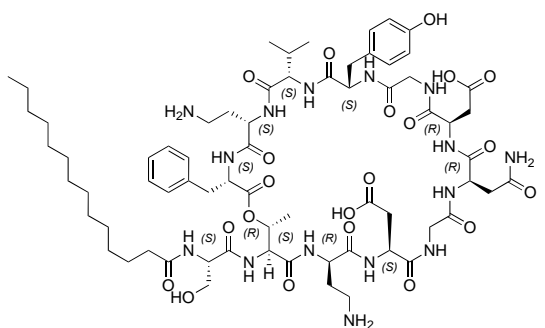
64



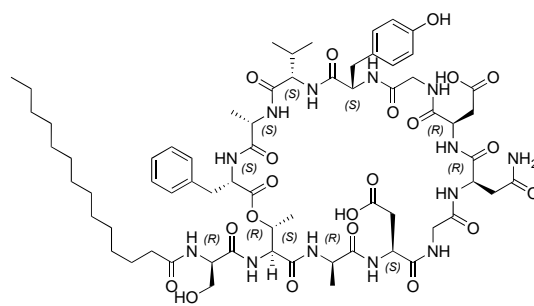
65



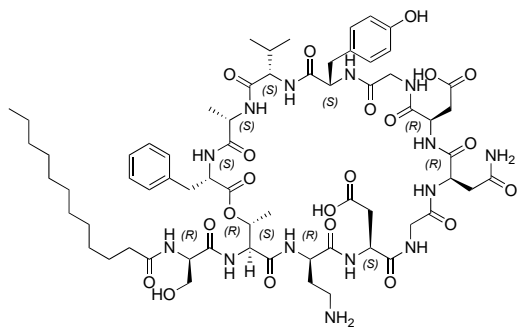
66



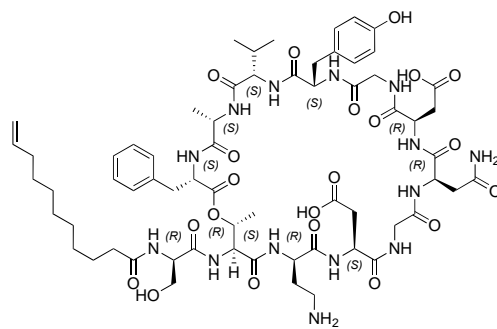
67



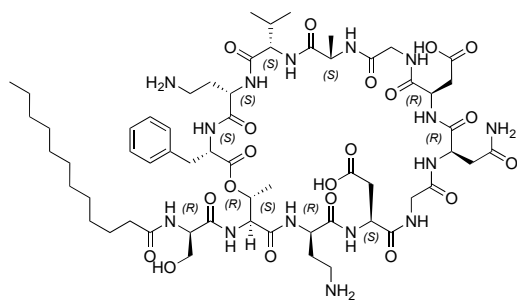
68



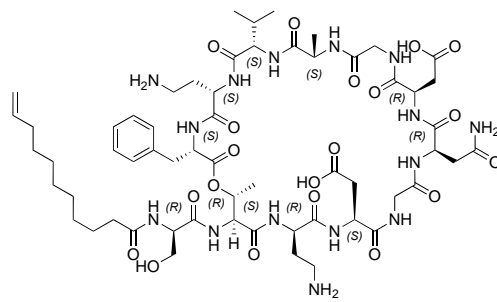
69



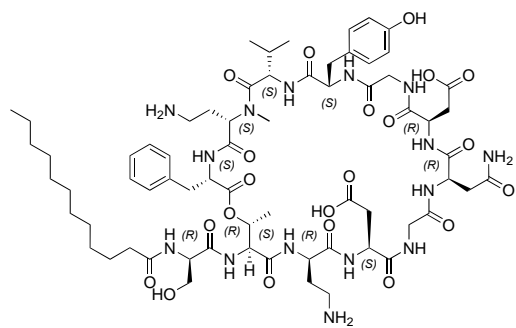
70



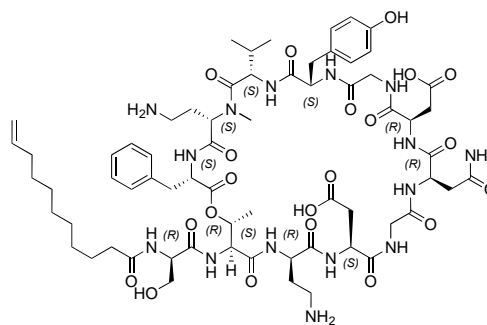
71



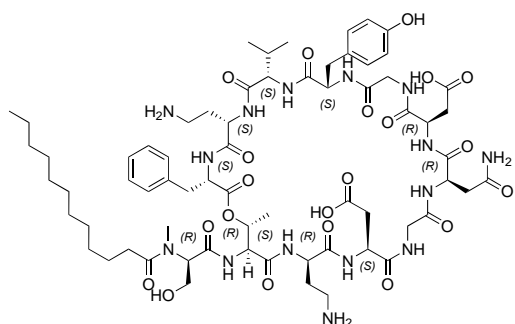
72



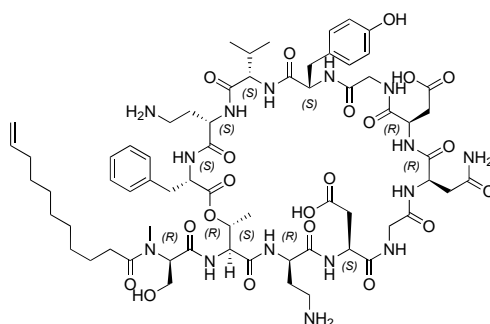
73



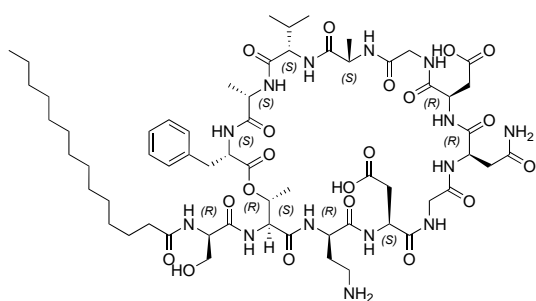
74



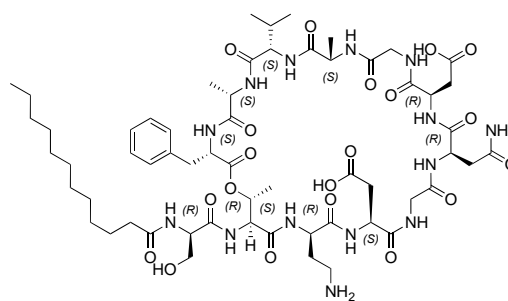
75



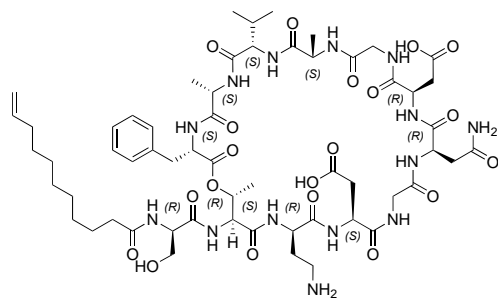
76



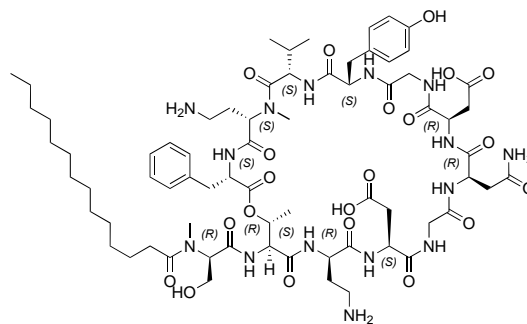
77



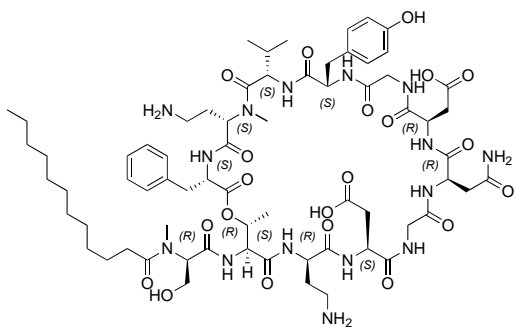
78



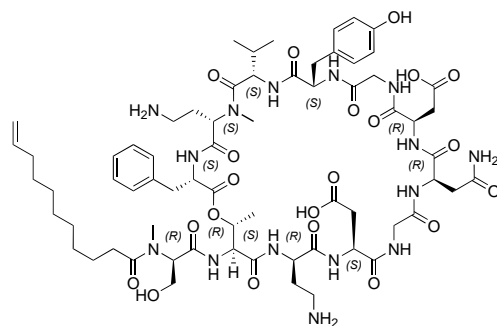
79



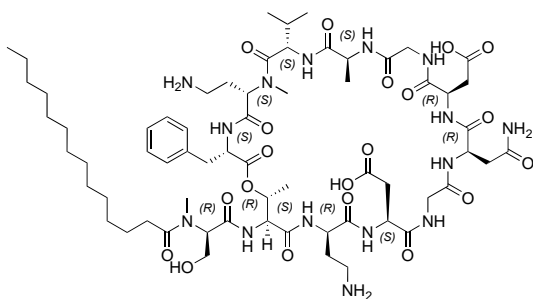
80



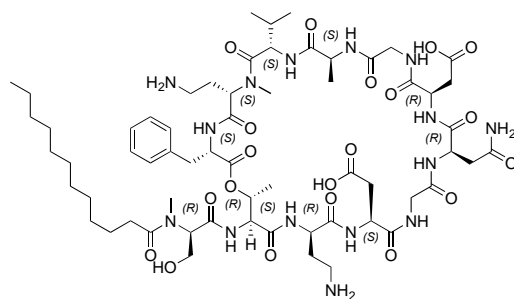
81



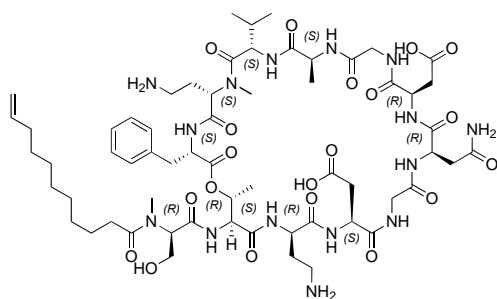
82



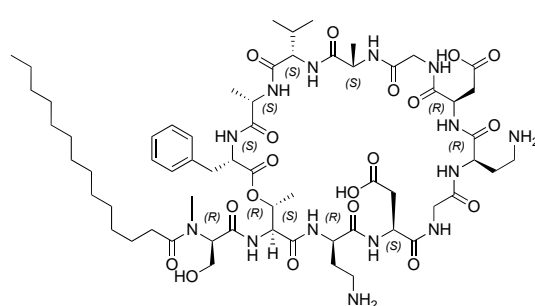
83



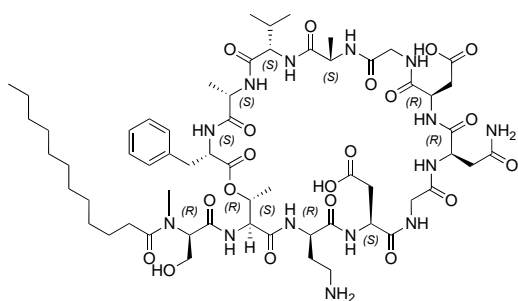
84



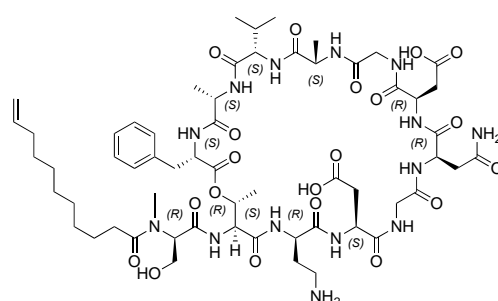
85



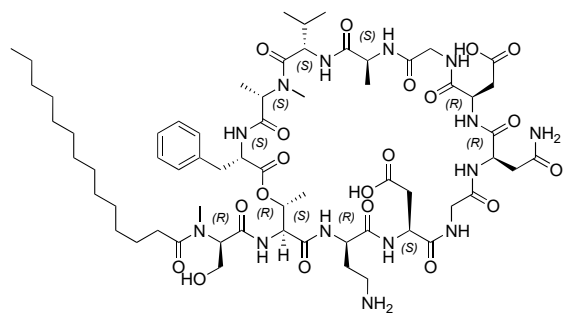
86



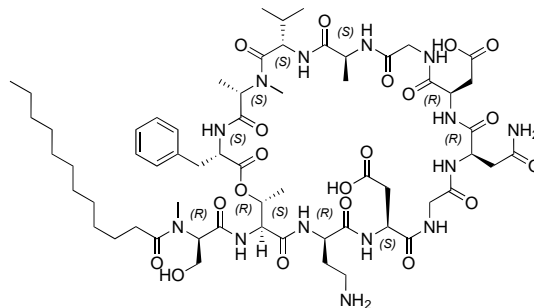
87



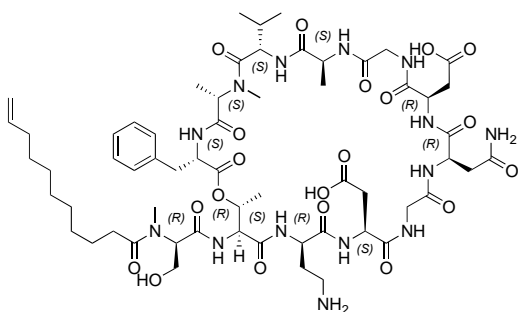
88



89



90



91

Appendix Figure 6.11. Structures and number assignments of synthesized cilgacin variants

Molecules	Chemical Formula	Calculated (M+H)	Observed (M+H)	Error (ppm)	Purity (%)	Made by
1	C68H103N15O21	1466.7531	1466.7422	5.3	95	AR
2	C67H85N15O21	1436.6122	1436.6047	3.7	98	AB
3	C74H91N15O21	1526.6592	1526.6552	3.4	98	AB
4	C66H91N15O21	1430.6592	1430.6557	4.0	>99	AB
5	C64H85N15O21	1400.6123	1400.611	0.7	89 [‡]	AB
6	C70H88N16O21	1489.6388	1489.6347	3.2	97	AB
7	C65H83N15O21	1410.5966	1410.5925	1.8	99	AB
8	C64H85N15O21	1400.6123	1400.6067	2.4	>99	AB
9	C66H90N16O21	1443.6545	1443.6521	1.5	>99	AB
10	C63H83N15O23	1418.5864	1418.5817	1.2	99	AB
11	C64H85N15O21	1400.6123	1400.6074	1.9	99	AB
12	C68H87N15O21	1450.6279	1450.6231	2.1	95	AB
13	C68H87N15O21	1450.6279	1450.6253	3.7	>99	AB
14	C71H93N15O21	1492.6749	1492.676	4.0	97	AB
15	C74H99N15O21	1534.7218	1534.7173	1.8	>99	AB
16	C60H87N15O21	1354.6279	1354.6227	2.0	>99	AB
17	C62H91N15O21	1382.6592	1382.6518	1.3	>99	AB
18	C62H92N16O21	1397.6701	1397.6595	7.5	93 [‡]	AB
19	C64H95N15O21	1410.6905	1410.6826	5.2	97	AB
20	C64H95N15O21	1410.6905	1410.6818	5.8	96	AB
21	C66H99N15O21	1438.7218	1438.7145	3.8	>99	AB
22	C70H107N15O21	1494.7844	1494.7802	0.1	>99	AB
23	C72H111N15O21	1522.8157	1522.807	2.0	99	AB
24	C63H93N15O21	1396.6749	1396.6645	3.9	99	AB
25	C64H91N15O21	1406.6592	1406.6545	3.2	>99	AB
26	C65H97N15O21	1424.7062	1424.7026	1.8	>99	AB
27	C65H95N15O21	1422.6905	1422.6892	0.6	99	AB
28	C66H97N15O21	1436.7062	1436.7006	0.4	98	AB
29	C67H101N15O21	1452.7375	1452.7322	1.5	99	AB
30	C69H105N15O21	1480.7688	1480.7676	5.1	>99	AB
31	C70H107N15O21	1494.7844	1494.7772	1.9	>99	AB
32	C62H99N15O21	1390.7218	1390.7173	1.9	99	AR
33	C67H100N14O21	1437.7266	1437.7179	1.5	98	AR
34	C66H99N15O21	1438.7218	1438.7169	2.2	>99	AR
35	C62H99N15O20	1374.7269	1374.7216	1.2	>99	AR
36	C67H103N15O19	1422.7633	1422.7565	2.5	>99	AR
37	C67H102N14O20	1423.7473	1423.7397	0.2	99	AR
38	C67H103N15O19	1422.7633	1422.7633	2.3	96	AR

39	C67H100N14O21	1437.7266	1437.7195	0.3	95	AR
40	C68H103N15O20	1450.7582	1450.7485	1.0	>99	AR
41	C64H101N15O23	1448.7273	1448.7239	2.7	97	KS
42	C69H102N14O23	1495.7321	1495.7199	6.8	>99	KS
43	C68H101N15O23	1496.7273	1496.7167	2.2	96	KS
44	C64H101N15O22	1432.7324	1432.7243	4.0	95	KS
45	C71H107N15O23	1538.7742	1538.7688	0.8	96	KS
46	C69H105N15O21	1480.7688	1480.7663	4.3	99	KS
47	C69H104N14O22	1481.7528	1481.7519	1.3	96	KS
48	C71H107N15O23	1538.7742	1538.7702	0.1	96	KS
49	C69H105N15O21	1480.7688	1480.7668	4.6	97	KS
50	C69H102N14O23	1495.7321	1495.7263	2.5	>99	KS
51	C70H105N15O22	1508.7637	1508.7655	3.6	>99	KS
52	C69H105N15O21	1480.7688	1480.7701	6.8	99	AR
53	C69H105N15O21	1480.7688	1480.7635	2.4	99	KS
54	C69H105N15O21	1480.7688	1480.7619	1.3	96	AR
55	C69H105N15O21	1480.7688	1480.7642	2.8	99	KS
56	C69H105N15O21	1480.7688	1480.7606	0.4	97	AR
57	C69H105N15O21	1480.7688	1480.76	0.0	99	KS
58	C69H105N15O21	1480.7688	1480.7586	0.9	98	KS
59	C69H105N15O21	1480.7688	1480.7602	0.1	98	AR
60	C69H105N15O21	1480.7688	1480.7646	3.1	99	KS
61	C69H105N15O21	1480.7688	1480.7618	1.2	99	KS
62	C69H105N15O21	1480.7688	1480.7656	3.8	99	KS
63	C69H105N15O21	1480.7688	1480.7586	0.9	97	AR
64	C68H103N15O21	1466.7531	1466.7437	4.3	99	KS
65	C68H103N15O21	1466.7531	1466.7439	4.2	96	KS
66	C68H103N15O21	1466.7531	1466.7439	4.2	94 [‡]	KS
67	C68H103N15O21	1466.7531	1466.7437	4.3	>99	KS
68	C66H97N13O21	1408.7000	1408.6955	3.2	>99	AR
69	C65H96N14O21	1409.6953	1409.692	1.4	>99	KS
70	C64H92N14O21	1393.6640	1393.6526	5.3	98	KS
71	C60H95N15O20	1346.6956	1346.6922	1.6	98	KS
72	C59H91N15O20	1330.6643	1330.6592	0.6	>99	KS
73	C67H101N15O21	1452.7374	1452.7301	0.1	>99	KS
74	C66H97N15O21	1436.7062	1436.6987	0.9	>99	KS
75	C67H101N15O21	1452.7374	1452.7277	1.6	>99	KS
76	C66H97N15O21	1436.7062	1436.7046	3.2	97	KS
77	C61H96N14O20	1345.7004	1345.6929	5.3	>99	KS
78	C59H92N14O20	1317.6691	1317.6593	0.5	98	KS

79	C58H88N14O20	1301.6378	1301.6256	3.4	97	KS
80	C70H107N15O21	1494.7844	1494.7784	1.1	>99	KS
81	C68H103N15O21	1466.7531	1466.7461	2.7	99	KS
82	C67H99N15O21	1450.7218	1450.7181	1.3	85 [‡]	KS
83	C64H103N15O20	1402.7582	1402.7528	2.0	>99	KS
84	C62H99N15O20	1374.7269	1374.7203	0.2	99	KS
85	C61H95N15O20	1358.6956	1358.6918	1.3	94 [‡]	KS
86	C62H98N14O20	1359.7160	1359.7102	0.1	>99	KS
87	C60H94N14O20	1331.6847	1331.6749	3.8	99	KS
88	C59H90N14O20	1315.6534	1315.6456	3.3	>99	KS
89	C63H100N14O20	1373.7317	1373.7186	8.3	98	KS
90	C61H96N14O20	1345.7004	1345.6848	11.3	96	KS
91	C60H92N14O20	1329.6691	1329.6627	2.0	69 [‡]	KS

*All HRMS data were collected in positive ionization mode with a mass range from m/z 200 to 2500.

[‡]Compounds with <95% purity were inactive.

Appendix Figure 6.12. High-resolution mass spectrometry* data for all synthetic cilagicin variants. UPLC determined peptide purity included. Individual responsible for synthesis noted [AR = Adam Rosenzweig, AB = Abir Bhattacharjee, KS = Kaylyn Spotton]

REFERENCES

- (1) Singh, S. B.; Pelaez, F. Biodiversity, chemical diversity and drug discovery. *Prog Drug Res* **2008**, *65*, 141, 143-174.
- (2) Williams, D. H.; Stone, M. J.; Hauck, P. R.; Rahman, S. K. Why are secondary metabolites (natural products) biosynthesized? *J Nat Prod* **1989**, *52* (6), 1189-1208.
- (3) Firm, R. D.; Jones, C. G. A Darwinian view of metabolism: molecular properties determine fitness. *J Exp Bot* **2009**, *60* (3), 719-726.
- (4) Cavalier-Smith, T. Origins of secondary metabolism. *Ciba Found Symp* **1992**, *171*, 64-80; discussion 80-67.
- (5) Demain, A. L.; Fang, A. The natural functions of secondary metabolites. *Adv Biochem Eng Biotechnol* **2000**, *69*, 1-39.
- (6) Vivek-Ananth, R. P.; Sahoo, A. K.; Baskaran, S. P.; Samal, A. Scaffold and Structural Diversity of the Secondary Metabolite Space of Medicinal Fungi. *ACS Omega* **2023**, *8* (3), 3102-3113.
- (7) Atanasov, A. G.; Zotchev, S. B.; Dirsch, V. M.; International Natural Product Sciences, T.; Supuran, C. T. Natural products in drug discovery: advances and opportunities. *Nat Rev Drug Discov* **2021**, *20* (3), 200-216.
- (8) Yuan, H.; Ma, Q.; Ye, L.; Piao, G. The Traditional Medicine and Modern Medicine from Natural Products. *Molecules* **2016**, *21* (5).
- (9) Dias, D. A.; Urban, S.; Roessner, U. A historical overview of natural products in drug discovery. *Metabolites* **2012**, *2* (2), 303-336.
- (10) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J Nat Prod* **2020**, *83* (3), 770-803.
- (11) Demain, A. L. Importance of microbial natural products and the need to revitalize their discovery. *J Ind Microbiol Biotechnol* **2014**, *41* (2), 185-201.

- (12) Hutchings, M. I.; Truman, A. W.; Wilkinson, B. Antibiotics: past, present and future. *Curr Opin Microbiol* **2019**, *51*, 72-80.
- (13) Davies, J. Where have All the Antibiotics Gone? *Can J Infect Dis Med Microbiol* **2006**, *17* (5), 287-290.
- (14) Katz, L.; Baltz, R. H. Natural product discovery: past, present, and future. *J Ind Microbiol Biotechnol* **2016**, *43* (2-3), 155-176.
- (15) Kapoor, G.; Saigal, S.; Elongavan, A. Action and resistance mechanisms of antibiotics: A guide for clinicians. *J Anaesthesiol Clin Pharmacol* **2017**, *33* (3), 300-305.
- (16) Durand, G. A.; Raoult, D.; Dubourg, G. Antibiotic discovery: history, methods and perspectives. *Int J Antimicrob Agents* **2019**, *53* (4), 371-382.
- (17) Santos-Aberturas, J.; Vior, N. M. Beyond Soil-Dwelling Actinobacteria: Fantastic Antibiotics and Where to Find Them. *Antibiotics (Basel)* **2022**, *11* (2).
- (18) Landeta, C.; Mejia-Santana, A. Union Is Strength: Target-Based and Whole-Cell High-Throughput Screens in Antibacterial Discovery. *J Bacteriol* **2022**, *204* (4), e0047721.
- (19) Liu, R.; Li, X.; Lam, K. S. Combinatorial chemistry in drug discovery. *Curr Opin Chem Biol* **2017**, *38*, 117-126.
- (20) Ventola, C. L. The antibiotic resistance crisis: part 1: causes and threats. *P T* **2015**, *40* (4), 277-283.
- (21) Stasiak, M.; Mackiw, E.; Kowalska, J.; Kucharek, K.; Postupolski, J. Silent Genes: Antimicrobial Resistance and Antibiotic Production. *Pol J Microbiol* **2021**, *70* (4), 421-429.
- (22) Rappe, M. S.; Giovannoni, S. J. The uncultured microbial majority. *Annu Rev Microbiol* **2003**, *57*, 369-394.
- (23) Fleischmann, R. D.; Adams, M. D.; White, O.; Clayton, R. A.; Kirkness, E. F.; Kerlavage, A. R.; Bult, C. J.; Tomb, J. F.; Dougherty, B. A.; Merrick, J. M.; et al. Whole-

genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **1995**, 269 (5223), 496-512.

(24) Hopwood, D. A. The Leeuwenhoek lecture, 1987. Towards an understanding of gene switching in *Streptomyces*, the basis of sporulation and antibiotic production. *Proc R Soc Lond B Biol Sci* **1988**, 235 (1279), 121-138.

(25) Stone, M. J.; Williams, D. H. On the evolution of functional secondary metabolites (natural products). *Mol Microbiol* **1992**, 6 (1), 29-34.

(26) Jensen, P. R. Natural Products and the Gene Cluster Revolution. *Trends Microbiol* **2016**, 24 (12), 968-977.

(27) Bode, H. B.; Muller, R. The impact of bacterial genomics on natural product research. *Angew Chem Int Ed Engl* **2005**, 44 (42), 6828-6846.

(28) Sayers, E. W.; Bolton, E. E.; Brister, J. R.; Canese, K.; Chan, J.; Comeau, D. C.; Farrell, C. M.; Feldgarden, M.; Fine, A. M.; Funk, K.; Hatcher, E.; Kannan, S.; Kelly, C.; Kim, S.; Klimke, W.; Landrum, M. J.; Lathrop, S.; Lu, Z.; Madden, T. L.; Malheiro, A.; Marchler-Bauer, A.; Murphy, T. D.; Phan, L.; Pujar, S.; Rangwala, S. H.; Schneider, V. A.; Tse, T.; Wang, J.; Ye, J.; Trawick, B. W.; Pruitt, K. D.; Sherry, S. T. Database resources of the National Center for Biotechnology Information in 2023. *Nucleic Acids Res* **2023**, 51 (D1), D29-D38.

(29) Arita, M.; Karsch-Mizrachi, I.; Cochrane, G. The international nucleotide sequence database collaboration. *Nucleic Acids Res* **2021**, 49 (D1), D121-D124.

(30) Covington, B. C.; Xu, F.; Seyedsayamdost, M. R. A Natural Product Chemist's Guide to Unlocking Silent Biosynthetic Gene Clusters. *Annu Rev Biochem* **2021**, 90, 763-788.

(31) Garfias-Gallegos, D.; Zirion-Martinez, C.; Bustos-Diaz, E. D.; Arellano-Fernandez, T. V.; Lovaco-Flores, J. A.; Espinosa-Jaime, A.; Avelar-Rivas, J. A.; Selem-Mojica, N. Metagenomics Bioinformatic Pipeline. *Methods Mol Biol* **2022**, 2512, 153-179.

(32) Rosenzweig, A. F.; Burian, J.; Brady, S. F. Present and future outlooks on environmental DNA-based methods for antibiotic discovery. *Curr Opin Microbiol* **2023**, 75, 102335.

- (33) Charlop-Powers, Z.; Milshteyn, A.; Brady, S. F. Metagenomic small molecule discovery methods. *Curr Opin Microbiol* **2014**, *19*, 70-75.
- (34) Chu, J.; Vila-Farres, X.; Inoyama, D.; Ternei, M.; Cohen, L. J.; Gordon, E. A.; Reddy, B. V.; Charlop-Powers, Z.; Zebroski, H. A.; Gallardo-Macias, R.; Jaskowski, M.; Satish, S.; Park, S.; Perlin, D. S.; Freundlich, J. S.; Brady, S. F. Discovery of MRSA active antibiotics using primary sequence from the human microbiome. *Nat Chem Biol* **2016**, *12* (12), 1004-1006.
- (35) Challis, G. L.; Naismith, J. H. Structural aspects of non-ribosomal peptide biosynthesis. *Curr Opin Struct Biol* **2004**, *14* (6), 748-756.
- (36) Robbins, T.; Liu, Y. C.; Cane, D. E.; Khosla, C. Structure and mechanism of assembly line polyketide synthases. *Curr Opin Struct Biol* **2016**, *41*, 10-18.
- (37) Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol* **1999**, *6* (8), 493-505.
- (38) Blin, K.; Shaw, S.; Kloosterman, A. M.; Charlop-Powers, Z.; van Wezel, G. P.; Medema, M. H.; Weber, T. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research* **2021**, *49* (W1), W29-W35.
- (39) Terlouw, B. R.; Blin, K.; Navarro-Munoz, J. C.; Avalon, N. E.; Chevrette, M. G.; Egbert, S.; Lee, S.; Meijer, D.; Recchia, M. J. J.; Reitz, Z. L.; van Santen, J. A.; Selem-Mojica, N.; Topping, T.; Zaroubi, L.; Alanjary, M.; Aleti, G.; Aguilar, C.; Al-Salihi, S. A. A.; Augustijn, H. E.; Avelar-Rivas, J. A.; Avitia-Dominguez, L. A.; Barona-Gomez, F.; Bernaldo-Aguero, J.; Bielinski, V. A.; Biermann, F.; Booth, T. J.; Bravo, V. J. C.; Castelo-Branco, R.; Chagas, F. O.; Cruz-Morales, P.; Gavriilidou, A.; Gayraud, D.; Gutierrez-Garcia, K.; Haslinger, K.; Helfrich, E. J. N.; van der Hoof, J. J. J.; Jati, A. P.; Kalkreuter, E.; Kalyvas, N.; Kang, K. B.; Kautsar, S.; Kim, W.; Kunjapur, A. M.; Li, Y. X.; Lin, G. M.; Loureiro, C.; Louwen, J. J. R.; Louwen, N. L. L.; Lund, G.; Parra, J.; Philmus, B.; Pourmohsenin, B.; Pronk, L. J. U.; Rego, A.; Rex, D. A. B.; Robinson, S.; Rosas-Becerra, L. R.; Roxborough, E. T.; Schorn, M. A.; Scobie, D. J.; Singh, K. S.; Sokolova, N.; Tang, X. Y.; Udworthy, D.; Vigneshwari, A.; Vind, K.; Vromans, S. P. J. M.; Waschulin, V.; Williams, S. E.; Winter, J. M.; Witte, T. E.; Xie, H. L.; Yang, D.; Yu, J. W.; Zdouc, M.; Zhong, Z.; Collemare, J.; Linington, R. G.; Weber, T.; Medema, M. H. MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Research* **2022**.

- (40) Skinnider, M. A.; Merwin, N. J.; Johnston, C. W.; Magarvey, N. A. PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Research* **2017**, *45* (W1), W49-W54.
- (41) Skinnider, M. A.; Johnston, C. W.; Gunabalasingam, M.; Merwin, N. J.; Kieliszek, A. M.; MacLellan, R. J.; Li, H.; Ranieri, M. R. M.; Webster, A. L. H.; Cao, M. P. T.; Pfeifle, A.; Spencer, N.; To, Q. H.; Wallace, D. P.; Dejong, C. A.; Magarvey, N. A. Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. *Nat Commun* **2020**, *11* (1), 6058.
- (42) Chu, J.; Vila-Farres, X.; Brady, S. F. Bioactive Synthetic-Bioinformatic Natural Product Cyclic Peptides Inspired by Nonribosomal Peptide Synthetase Gene Clusters from the Human Microbiome. *J Am Chem Soc* **2019**, *141* (40), 15737-15741.
- (43) Li, L.; Koirala, B.; Hernandez, Y.; MacIntyre, L. W.; Ternei, M. A.; Russo, R.; Brady, S. F. Identification of structurally diverse menaquinone-binding antibiotics with in vivo activity against multidrug-resistant pathogens. *Nat Microbiol* **2022**, *7* (1), 120-131.
- (44) Wang, Z.; Forelli, N.; Hernandez, Y.; Ternei, M.; Brady, S. F. Lapcin, a potent dual topoisomerase I/II inhibitor discovered by soil metagenome guided total chemical synthesis. *Nat Commun* **2022**, *13* (1), 842.
- (45) Chu, J.; Koirala, B.; Forelli, N.; Vila-Farres, X.; Ternei, M. A.; Ali, T.; Colosimo, D. A.; Brady, S. F. Synthetic-Bioinformatic Natural Product Antibiotics with Diverse Modes of Action. *J Am Chem Soc* **2020**, *142* (33), 14158-14168.
- (46) Wang, Z.; Koirala, B.; Hernandez, Y.; Zimmerman, M.; Brady, S. F. Bioinformatic prospecting and synthesis of a bifunctional lipopeptide antibiotic that evades resistance. *Science* **2022**, *376* (6596), 991-996.
- (47) Ledger, E. V. K.; Sabnis, A.; Edwards, A. M. Polymyxin and lipopeptide antibiotics: membrane-targeting drugs of last resort. *Microbiology (Reading)* **2022**, *168* (2).
- (48) Rausch, C.; Hoof, I.; Weber, T.; Wohlleben, W.; Huson, D. H. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol Biol* **2007**, *7*, 78.
- (49) Greber, K. E.; Zielinska, J.; Nierzwicki, L.; Ciura, K.; Kawczak, P.; Nowakowska, J.; Baczek, T.; Sawicki, W. Are the short cationic lipopeptides bacterial membrane

disruptors? Structure-Activity Relationship and molecular dynamic evaluation. *Biochim Biophys Acta Biomembr* **2019**, *1861* (1), 93-99.

(50) Mohapatra, S. S.; Dwibedy, S. K.; Padhy, I. Polymyxins, the last-resort antibiotics: Mode of action, resistance emergence, and potential solutions. *J Biosci* **2021**, *46* (3).

(51) Romaniuk, J. A.; Cegelski, L. Bacterial cell wall composition and the influence of antibiotics by cell-wall and whole-cell NMR. *Philos Trans R Soc Lond B Biol Sci* **2015**, *370* (1679).

(52) Miller, W. R.; Bayer, A. S.; Arias, C. A. Mechanism of Action and Resistance to Daptomycin in *Staphylococcus aureus* and Enterococci. *Cold Spring Harb Perspect Med* **2016**, *6* (11).

(53) Workman, S. D.; Strynadka, N. C. J. A Slippery Scaffold: Synthesis and Recycling of the Bacterial Cell Wall Carrier Lipid. *J Mol Biol* **2020**, *432* (18), 4964-4982.

(54) El Ghachi, M.; Howe, N.; Huang, C. Y.; Olieric, V.; Warshamanage, R.; Touze, T.; Weichert, D.; Stansfeld, P. J.; Wang, M.; Kerff, F.; Caffrey, M. Crystal structure of undecaprenyl-pyrophosphate phosphatase and its role in peptidoglycan biosynthesis. *Nat Commun* **2018**, *9* (1), 1078.

(55) Storm, D. R.; Strominger, J. L. Complex formation between bacitracin peptides and isoprenyl pyrophosphates. The specificity of lipid-peptide interactions. *J Biol Chem* **1973**, *248* (11), 3940-3945.

(56) Schneider, T.; Gries, K.; Josten, M.; Wiedemann, I.; Pelzer, S.; Labischinski, H.; Sahl, H. G. The lipopeptide antibiotic Friulimicin B inhibits cell wall biosynthesis through complex formation with bactoprenol phosphate. *Antimicrob Agents Chemother* **2009**, *53* (4), 1610-1618.

(57) Hernandez-Rocamora, V. M.; Otten, C. F.; Radkov, A.; Simorre, J. P.; Breukink, E.; VanNieuwenhze, M.; Vollmer, W. Coupling of polymerase and carrier lipid phosphatase prevents product inhibition in peptidoglycan synthesis. *Cell Surf* **2018**, *2*, 1-13.

(58) Reygaert, W. C. An overview of the antimicrobial resistance mechanisms of bacteria. *AIMS Microbiol* **2018**, *4* (3), 482-501.

- (59) Aslam, B.; Wang, W.; Arshad, M. I.; Khurshid, M.; Muzammil, S.; Rasool, M. H.; Nisar, M. A.; Alvi, R. F.; Aslam, M. A.; Qamar, M. U.; Salamat, M. K. F.; Baloch, Z. Antibiotic resistance: a rundown of a global crisis. *Infect Drug Resist* **2018**, *11*, 1645-1658.
- (60) Dangi, M.; Khichi, A.; Jakhar, R.; Chhillar, A. K. Growing Preferences towards Analog-based Drug Discovery. *Curr Pharm Biotechnol* **2021**, *22* (8), 1030-1045.
- (61) Fischer, J. n.; Ganellin, C. R.; Rotella, D. P. *Analogue-based drug discovery III*; Wiley-VCH, 2013.
- (62) Wood, T. M.; Martin, N. I. The calcium-dependent lipopeptide antibiotics: structure, mechanism, & medicinal chemistry. *Medchemcomm* **2019**, *10* (5), 634-646.
- (63) Rosenzweig, A. F.; Wang, Z.; Morales-Amador, A.; Spotton, K.; Brady, S. F. A Family of Antibiotics That Evades Resistance by Binding Polyprenyl Phosphates. *ACS Infect Dis* **2023**, *9* (12), 2394-2400.
- (64) Miller, B. R.; Gulick, A. M. Structural Biology of Nonribosomal Peptide Synthetases. *Methods Mol Biol* **2016**, *1401*, 3-29.
- (65) Nordberg, H.; Cantor, M.; Dusheyko, S.; Hua, S.; Poliakov, A.; Shabalov, I.; Smirnova, T.; Grigoriev, I. V.; Dubchak, I. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Res* **2014**, *42* (Database issue), D26-31.
- (66) Winn, M.; Richardson, S. M.; Campopiano, D. J.; Micklefield, J. Harnessing and engineering amide bond forming ligases for the synthesis of amides. *Curr Opin Chem Biol* **2020**, *55*, 77-85.
- (67) Meena, K. R.; Kanwar, S. S. Lipopeptides as the antifungal and antibacterial agents: applications in food safety and therapeutics. *Biomed Res Int* **2015**, *2015*, 473050.
- (68) Santajit, S.; Indrawattana, N. Mechanisms of Antimicrobial Resistance in ESKAPE Pathogens. *Biomed Res Int* **2016**, *2016*, 2475067.
- (69) Clinical; Laboratory Standards, I.; Weinstein, M. P. *Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically*; Clinical Laboratory Standards Institute, 2018.

- (70) Radeck, J.; Gebhard, S.; Orchard, P. S.; Kirchner, M.; Bauer, S.; Mascher, T.; Fritz, G. Anatomy of the bacitracin resistance network in *Bacillus subtilis*. *Mol Microbiol* **2016**, *100* (4), 607-620.
- (71) Diehl, A.; Wood, T. M.; Gebhard, S.; Martin, N. I.; Fritz, G. The Cell Envelope Stress Response of *Bacillus subtilis* towards Laspartomycin C. *Antibiotics (Basel)* **2020**, *9* (11).
- (72) Luo, C.; Liu, X.; Zhou, X.; Guo, J.; Truong, J.; Wang, X.; Zhou, H.; Li, X.; Chen, Z. Unusual Biosynthesis and Structure of Locillomycins from *Bacillus subtilis* 916. *Appl Environ Microbiol* **2015**, *81* (19), 6601-6609.
- (73) Hefti, F. F. Requirements for a lead compound to become a clinical candidate. *BMC Neurosci* **2008**, *9 Suppl 3* (Suppl 3), S7.
- (74) Griffin, J. P.; Posner, J.; Barker, G. R. *The textbook of pharmaceutical medicine*; Wiley-Blackwell, 2013.
- (75) Wise, R. Protein binding of beta-lactams: the effects on activity and pharmacology particularly tissue penetration. I. *J Antimicrob Chemother* **1983**, *12* (1), 1-18.
- (76) Wise, R. Protein binding of beta-lactams: the effects on activity and pharmacology particularly tissue penetration. II. Studies in man. *J Antimicrob Chemother* **1983**, *12* (2), 105-118.
- (77) Craig, W. A.; Kunin, C. M. Significance of serum protein and tissue binding of antimicrobial agents. *Annu Rev Med* **1976**, *27*, 287-300.
- (78) Pitkin, D. H.; Mico, B. A.; Sitrin, R. D.; Nisbet, L. J. Charge and Lipophilicity Govern the Pharmacokinetics of Glycopeptide Antibiotics. *Antimicrob Agents Ch* **1986**, *29* (3), 440-444.
- (79) Gioiello, A.; Piccinno, A.; Lozza, A. M.; Cerra, B. The Medicinal Chemistry in the Era of Machines and Automation: Recent Advances in Continuous Flow Technology. *J Med Chem* **2020**, *63* (13), 6624-6647.
- (80) Wang, S.; Dong, G.; Sheng, C. Structural simplification: an efficient strategy in lead optimization. *Acta Pharm Sin B* **2019**, *9* (5), 880-901.

(81) Deschamps-Labat, L.; Pehourcq, F.; Jagou, M.; Bannwarth, B. Relationship between lipophilicity and binding to human serum albumin of arylpropionic acid non-steroidal anti-inflammatory drugs. *J Pharm Biomed Anal* **1997**, *16* (2), 223-229.

(82) Ermondi, G.; Lorenti, M.; Caron, G. Contribution of ionization and lipophilicity to drug binding to albumin: a preliminary step toward biodistribution prediction. *J Med Chem* **2004**, *47* (16), 3949-3961.

(83) Lei, J.; Sun, L.; Huang, S.; Zhu, C.; Li, P.; He, J.; Mackey, V.; Coy, D. H.; He, Q. The antimicrobial peptides and their potential clinical applications. *Am J Transl Res* **2019**, *11* (7), 3919-3931.

(84) Wheadon, M. J.; Townsend, C. A. Evolutionary and functional analysis of an NRPS condensation domain integrates beta-lactam, D-amino acid, and dehydroamino acid synthesis. *Proc Natl Acad Sci U S A* **2021**, *118* (17).

(85) Kleijn, L. H. J.; Vlieg, H. C.; Wood, T. M.; Toraño, J. S.; Janssen, B. J. C.; Martin, N. I. A High-Resolution Crystal Structure that Reveals Molecular Details of Target Recognition by the Calcium-Dependent Lipopeptide Antibiotic Laspartomycin C. *Angew Chem Int Edit* **2017**, *56* (52), 16546-16549.

(86) Hughes, D.; Karlen, A. Discovery and preclinical development of new antibiotics. *Ups J Med Sci* **2014**, *119* (2), 162-169.

(87) CDC. *Antibiotic Resistance Threats in the United States*; Center for Disease Control and Prevention, Atlanta, GA, 2019. DOI: Dec. 2019.

(88) Chavali, A. K.; Rhee, S. Y. Bioinformatics tools for the identification of gene clusters that biosynthesize specialized metabolites. *Brief Bioinform* **2018**, *19* (5), 1022-1034.

(89) Chen, R.; Wong, H. L.; Burns, B. P. New Approaches to Detect Biosynthetic Gene Clusters in the Environment. *Medicines (Basel)* **2019**, *6* (1).

(90) Luo, Y.; Cobb, R. E.; Zhao, H. Recent advances in natural product discovery. *Curr Opin Biotechnol* **2014**, *30*, 230-237.

(91) Hover, B. M.; Kim, S. H.; Katz, M.; Charlop-Powers, Z.; Owen, J. G.; Ternei, M. A.; Maniko, J.; Estrela, A. B.; Molina, H.; Park, S.; Perlin, D. S.; Brady, S. F. Culture-

independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. *Nat Microbiol* **2018**, 3 (4), 415-422.

(92) Vila-Farres, X.; Chu, J.; Inoyama, D.; Ternei, M. A.; Lemetre, C.; Cohen, L. J.; Cho, W.; Reddy, B. V.; Zebroski, H. A.; Freundlich, J. S.; Perlin, D. S.; Brady, S. F. Antimicrobials Inspired by Nonribosomal Peptide Synthetase Gene Clusters. *J Am Chem Soc* **2017**, 139 (4), 1404-1407.

(93) Vila-Farres, X.; Chu, J.; Ternei, M. A.; Lemetre, C.; Park, S.; Perlin, D. S.; Brady, S. F. An Optimized Synthetic-Bioinformatic Natural Product Antibiotic Sterilizes Multidrug-Resistant *Acinetobacter baumannii*-Infected Wounds. *mSphere* **2018**, 3 (1).