MICROBIAL GENOME MINING IN UNDERSTANDING HUMAN BACTERIAL PATHOGENESIS AND CYANOBACTERIAL NATURAL PRODUCTS BIOSYNTHESIS

by

Daniel H. Kwak

Bachelor of Science, Carnegie Mellon University, 2011

Master of Science, University of Pittsburgh, 2015

Submitted to the Graduate Faculty of the

Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Daniel H. Kwak

It was defended on

November 20th, 2015

and approved by

Prof. Steve Weber, Professor, Chemistry

Prof. Alexander Deiters, Professor, Chemistry

Prof. Robert Shanks, Associate Professor, Microbiology and Molecular Genetics

Dissertation Advisor: Prof. Xinyu Liu, Assistant Professor, Chemistry

MICROBIAL GENOME MINING IN UNDERSTANDING HUMAN BACTERIAL PATHOGENESIS AND CYANOBACTERIAL NATURAL PRODUCTS BIOSYNTHESIS

Daniel H. Kwak, PhD

University of Pittsburgh, 2015

Recent accessibility of microbial genome sequencing data has enabled broad investigations into the nature of microbial physiology and their consequences on humans and their environment. Through an approach known as genome mining, an *in silico* technique that enables the identification of uncharacterized gene clusters based on sequence homology, the investigations described herein provide new insight into two important groups of microbes, namely human bacterial pathogens and cyanobacteria. In Part I of this dissertation, genome mining techniques have identified two evolutionarily-conserved cryptic biosynthetic operons in the human pathogens *Acinetobacter baumannii* and *Pseudomonas aeruginosa*. Interestingly, the findings demonstrate the significance of these gene clusters and their small molecule products to be important contributors to the pathogenesis of these organisms. As a result, the proteins encoded by these gene clusters are expected to be important targets in the development of next-generation antibiotics. In Part II, a novel platform for the investigation of natural products biosynthesis from cyanobacteria is described. Cyanobacteria have proven to be important yet relatively unexplored sources of bioactive compounds. Recent genome sequencing has indicated the substantial potential of these microbes to synthesize compounds that can be developed into new medicines.

The collective findings in this dissertation demonstrate the considerable utility of emerging microbial genome sequencing information and its future impact on human health and disease.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

My role models, colleagues, and family have been no less than an indispensible cornerstone of my professional and personal growth throughout my graduate studies. I will remember with fondess my advisor's devotion to engage our curiosity and inspire ingenuity. I also believe our research group to have engendered a rare balance of rigorous intellectualism and gracious character that promotes our best attributes as scientists and leaders. I am especially honored to have worked with postdoctoral associates Drs. Matthew Hillwig and Qin Zhu, graduate students Ani Sasmal, Kuljira Ittiamornkul, and Cihad Sigindere, and undergraduates Aditya Badeti, John Hong, Tyler Sevco, and Heather Fuhrman. I am also indebted to Profs. Robert Shanks, Yohei Doi, Sanford Asher, Russell Salter, and Jennifer Bomberger, whose guidance has been instrumental in my development as a researcher. Furthermore, I am privileged to have collaborated with Drs. Zhongyu Cai, Jian-Tao Zhang, Kim Brothers, Peter Keyel as well as Nick Stella and Becca Flitter. Without the unconditional support of my family, surely, the experience of graduate school would not have been as gratifying. I thank my parents, Ha, Esther, and Hanul.

1.0 PART I INTRODUCTION: THE DISCOVERY AND CHARACTERIZATION OF SECONDARY METABOLITE VIRULENCE FACTORS

There has been an unprecedented emergence of multidrug resistant (MDR) pathogenic bacteria that is taking place in hospitals worldwide. These "superbugs" have acquired resistance to the most commonly used antibiotics and continue to be the cause of rapidly increasing numbers of hospital-acquired infections. MDR bacterial strains have been identified from a number of different bacterial species with some of the most prevalent including methicillin-resistant *Staphylococcus aureus* (MRSA), vancomycin-resistant *Enterococcus faecium* (VRE) as well as pan-resistant *Acinetobacter baumannii* and *Pseudomonas aeruginosa*.[1] Pathogenic strains exhibiting pandrug resistance (PDR) are especially alarming, as they are resistant to all classes of antibiotics and represent the significant gap between antibiotics discovery and the treatment of MDR pathogens.

Traditional antibiotics function by inhibiting essential processes of bacterial metabolism (e.g., cell wall synthesis, protein synthesis, RNA transcription). This leads to the elimination of antibiotic sensitive strains and provides enhanced opportunities for the subpopulation of resistant and virulent strains to proliferate in the host. The introduction of new antibiotics to this microbial population leads to the continuous selection of increasingly resistant bacterial members.

Drug resistance arises due to random mutations within their genome or from the acquisition of drug resistance genes from other bacteria. These resistance genes function to

export the antibiotic outside of the cell, to inactivate the drug, to alter the drug target site, or to prevent the compound from entering the cell.[2, 3] The propagation of drug resistance genes and their phenotypes is facilitated by the exchange, transport, and acquisition of DNA, namely plasmids and other mobile DNA vectors.[4]

Because current antibiotics target vital components of the organism, a substantial selective pressure exists for the emergence of strains that can survive the effects of the drug. Consequently, antibiotic resistant strains can readily propagate and exhibit virulence. An inherent drawback of these types of antibiotics, therefore, is the continuous emergence of drug resistant bacterial strains. Therefore, there is expanding interest in next generation antibiotics that can selectively target pathogens without exerting substantial selective pressure for drug resistances.

1.1 NEXT GENERATION ANTIBIOTICS: ANTIVIRULENCE DRUGS

One promising approach is to develop drugs that inhibit targets that are non-essential to the overall survival of the cell but essential in its pathogenicity (i.e., virulence factors). Theoretically, the inhibition of bacterial virulence factors is nonlethal and places minimal pressure for the pathogen to develop resistance but is effective in combating bacterial pathogenesis. [5, 6] Some of these targets include genes and molecules that are involved in: adhesion (colonization and invasion), toxicity, biofilm formation, secretion and cell-to-cell communication (quorum sensing). Inhibitors directed against adhesion molecules, toxins, cell

communication systems, and secretory mechanisms have proven to be especially promising. Well-characterized microbial small-molecule virulence factors are shown in **Figure 1.1**.

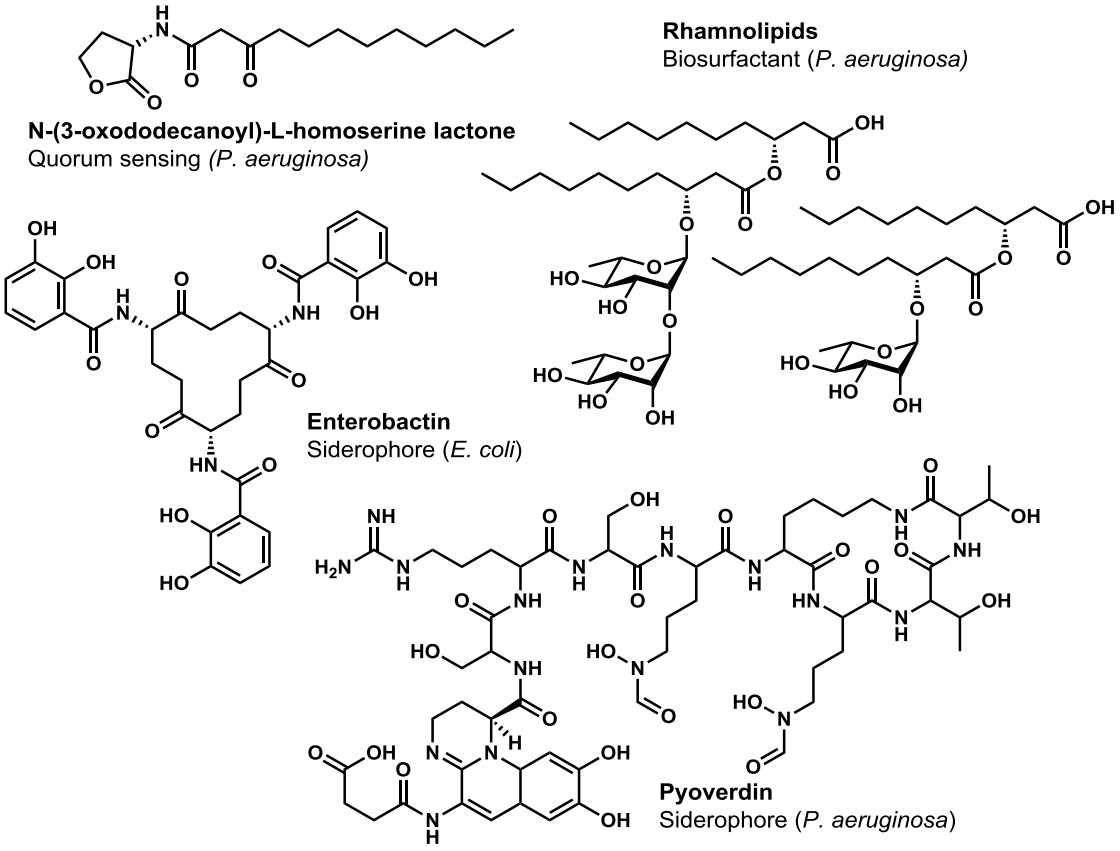


Figure 1.1 Microbial secondary metabolite virulence factors.

1.1.1 Pilus System

For example, proteinaceous hair-like cell surface appendages known as pili are the major structures of gram-negative bacteria responsible for adhesion and invasion of human cells and are also responsible for biofilm formation and macromolecular transport [7]. Uropathogenic *Escherichia coli* utilize these structures to cause urinary tract infections (UTI).[8] The design and development of several small molecule inhibitors of pili have been one promising route toward antivirulence approaches against UTI that circumvent the selection of antibiotic resistant pathogenic strains.[9-11]

1.1.2 Toxins

Toxins are another class of virulence factors that are attractive targets for drug development due to their direct effects on bacterial pathogenicity. Two well-characterized protein toxins include the Shiga toxin from *E. coli*[12] and the anthrax toxin from *Bacillus anthracis*[13]. Both exert their effects by entering into and then disrupting normal cell function in the infected human cell: the Shiga toxin interferes with intracellular signaling to ultimately cause over-accumulation of cellular fluid and cell death[14], while the anthrax toxin halts host protein synthesis, also resulting in cell death[15]. Successful antivirulence approaches to combat the Shiga and anthrax toxins have included the inhibition of toxin entry into the cell[16, 17] as well as the inhibition of protein-protein interactions or enzymatic activity[18, 19] that lead to cell death. Developments in these antivirulence therapeutics have included both antibodies as well as small molecule inhibitors targeting functional subunits of these toxins.

1.1.3 Biofilm

Biofilms are complex bacterial community structures that are composed of a matrix consisting of polysaccharides, proteins, nucleic acids, and various nutrients and small molecules. These structures impede the diffusion of molecules and has been shown to confer drug resistance to their microbial inhabitants. Biofilms are an important clinical challenge due to their persistence in a number of bacterial infections, including those of the middle ear due to *Haemophilus influenzae*[20], *Helicobacter pylori* biofilms of gastric ulcers[21], as well as biofilms of *P. aeruginosa* that contribute to lung infections of cystic fibrosis patients[22, 23]. Antivirulence approaches that target biofilms have included attenuation of quorum sensing systems as well as the inhibition of polysaccharide and protein biosynthesis that give rise to the extracellular matrix of the biofilm.[24] In addition to ubiquitous biological macromolecules, small molecules biosurfactants, such as the rhamnolipids from *P. aeruginosa*, have been demonstrated to be essential in the proper architectural development of these complex microstructures [25].

1.1.4 Secretory Systems

Type III secretion systems (TTSS) have been another bacterial target in the development of next generation antibiotics. TTSS are composed of protein complexes that enable the transport of toxins and other effectors from the pathogenic bacteria directly into the human cell to elicit virulence; they can also secrete molecules into the surrounding environment to facilitate microbial motility.[26, 27] Drug screening studies have identified several classes of compounds that can effectively inhibit these TTSS mechanisms from two prominent human pathogens,

Chlamydia trachomatis and *Yersinia pseudotuberculosis*, preventing their development as well as the progression of disease.[28, 29]

1.1.5 Quorum Sensing

Lastly, antivirulence approaches to target quorum sensing (QS) systems have also been explored. QS molecules are autoinducing signals that upregulate their own biosynthesis and is dependent on bacteria population densities. QS regulates gene expression and enables bacteria to coordinate diverse group behaviors, such as those involved in virulence. The pleiotrophic roles of QS in bacterial pathogenicity makes this target especially attractive, because multiple virulence factors could be silenced by inhibiting QS. Synthetic QS inhibitors, whose designs are based on their natively synthesized structures, have been developed and demonstrate inhibition against virulence phenotypes from human bacterial pathogens, such as in the production of toxic shock syndrome toxin-1, a toxin from *Staphylococcus aureus* [30], and as well in the production of small molecule virulence factors, pyocyanin and rhamnolipids, from *P. aeruginosa* [31]. Other inhibitors of the widely distributed QseC histidine sensor kinase QS system, found in over two dozen bacterial species, have been identified by library screening and is believed to be one ideal target as a broad spectrum antibiotic.[32]

Next generation antibiotics based on antivirulence approaches circumvent a significant shortcoming of current antimicrobial therapeutics: the substantial pressure placed on microbes to acquire drug resistances. As opposed to targeting essential processes, they are designed to inhibit mechanisms involved primarily in virulence, and ideally, render these microbes non-pathogenic. Although the antivirulence approach is still in nascent developmental stages, there is increasing

promise in this method of addressing human pathogen drug resistance. Moving forward, however, will require the elucidation of several unknown aspects. Although in theory antivirulence drugs will alleviate selective pressure for drug resistant strains, it is unknown whether or how rapidly resistant strains will emerge. Furthermore, it is unknown how such antivirulence drugs will affect the commensal population of the human microbiota and whether there will be detrimental effects toward the host. This latter concern could be addressed by the selective targeting of bacterial species (narrow spectrum) as opposed to targeting of general microbial mechanisms (broad spectrum). Narrow spectrum antibiotics are particularly ideal when the aetiological agent of the infection is known and when disturbance of the normal microbiota is unwanted.

1.1.6 Siderophores

A notable class of microbial secondary metabolites that are widely distributed amongst bacterial species and have essential roles (e.g., DNA synthesis, cellular respiration) in survival include iron-chelating molecules, known as siderophores. Iron is a micronutrient that serves as a cofactor for a variety of enzymatic functions within the cell. Although abundant in the environment, the bioavailability of iron to the microbe is limited, as it is most commonly present as the insoluble Fe^{3+} ion [33]. Therefore, access to iron requires a specialized carrier and transport system.

The *E. coli* siderophore, enterobactin, and its transport system serve as the archetype for analogous systems in other microbial species. Enterobactin is biosynthesized by the *ent* operon and is regulated by the presence of free iron [34-36]; this siderophore demonstrates a binding affinity for Fe^{3+} that is not atypical for siderophores ($K_a \approx 10^{52}\ M^{-1}$) [37]. Once enterobactin has

sequestered soluble Fe^{3+}, it can be recognized by outer and inner membrane protein systems (Fep and Ton-Exb systems) to transport the chelated iron into the cell for utilization [38]. Given the essential role of iron for the cellular growth and survival of the cell, its uptake inhibition has been associated with decreased bacterial viability [39, 40]. Their essential roles have also highlighted these important targets in the development of antivirulence therapeutics [41].

1.1.7 Discovery of Novel Secondary Metabolite Virulence Factors Quorum Sensing

The increasing availability of whole microbial genome information in recent years has enabled the unprecedented bioinformatic discovery of enzymatic machinery purported in the biosynthesis of secondary metabolites [42]. Metagenome information has been utilized to identify pathogenicity islands (DNA sequences predicted to encode virulence traits) [43]. Genome mining approaches have been employed in the discovery of novel microbial secondary metabolite virulence factors from such human bacterial pathogens as *S. aureus* [44], *P. aeruginosa* [45], and *Burkholderia pseudomallei* [46]. The utilization of complete genome sequencing data to characterize microbial virulence factors is particularly advantageous over traditional methods in the discovery of microbial virulence factors, which rely on the laborious and time-consuming generation and phenotypic screening of numerous transposon mutants [47].

In addition to bioinformatic analyses of factors that contribute to microbial pathogenesis, access to genome sequencing information can also enable the rapid and facile characterization of cryptic, virulent secondary metabolites. Specifically, novel genes with sequence homology to enzymes involved in the generation of nonribosomal peptides (NRPs), polyketides (PKs), and

ribosomally synthesized and posttranslationally modified peptides (RiPPs) can be identified via genome-mining approaches. Biosynthetic enzymes responsible for the generation of NRPs, PKs, or RiPPs are highly similar, and therefore, new genes responsible for their generation can be identified through comparative sequence analysis to known genes within these enzyme families. Genome-guided discovery of microbial virulence factors allows de-replication of small molecule discovery and the targeted elucidation of novel small molecules. In contrast, traditional approaches require bioassay-guided purification efforts, which are non-systematic and prone to the re-discovery of known compounds. The rapid characterization of large operons encoding virulence factors can be facilitated by employing cloning techniques that are capable of co-assembling biosynthetic entire gene clusters. Such techniques proved to be successful in the following chapters, investigating microbial virulence factors. This research will discuss the genome-guided discovery and characterization of cryptic secondary metabolites from human bacterial pathogens *Acinetobacter baumannii* and *P. aeruginosa*.

2.0 CRYPTIC ACINETOBACTER TOXIN: BIOGENESIS, FUNCTION AND STRUCTURE

Multidrug resistant (MDR) *Acinetobacter baumannii* has emerged in recent years as the hallmark "superbug": a pathogen difficult or impossible to treat with traditional clinical antibiotics. Its persistence as a human pathogen can be attributed to its rapid ability to acquire antibiotic resistance phenotypes and to thrive in diverse environments of the hospital setting for extended periods of time. [48] Specifically, MDR *A. baumannii* has been implicated in a number of infections (including respiratory, skin, and bloodstream) of immunocompromised patients. [48, 49] Furthermore, MDR *A. baumannii* is known to colonize different abiotic surfaces, such as catheters, respiratory lines, and glass [50-52]. The decreasing efficacy of current antibiotics to treat MDR *A. baumannii* infections has served as the impetus toward elucidating its mechanisms of pathogenesis in order to develop therapeutics that can target these mechanisms and render the microbes non-virulent.

Despite its recognized association with various infections of the human body, there is relatively little known about the mechanisms of *A. baumannii* pathogenesis. Nonetheless, several virulence factors, including proteins and small molecules, have been characterized in their roles in contributing to *A. baumannii* pathogenesis. Notably, *A. baumannii* has acquired a number of mechanisms to confer resistance to all of the major classes of antibiotics, thereby facilitating its

pathogenesis. [48, 53, 54] Genome sequencing of MDR *A. baumannii* clinical strains has revealed that a major proportion of the genes that encode antibiotic resistance are situated on a resistance island. [55, 56] This genetic island is absent in antibiotic-susceptible strains and appears to be readily transferred due to the presence of transposase genes encoded within this locus. Transposases can facilitate transfer of these genes onto a plasmid that can then be distributed to other isolates.

Several protein virulence factors from *A. baumannii* have also been identified. One of the most extensively characterized includes members of a family of cell surface proteins, known as outer membrane proteins (OMPs). OMPs have been demonstrated to contribute to pathogenesis through the induction of host cell apoptosis [57] and cytotoxicity [58], mediation of antibiotic resistance [59-61], cell invasion [62], as well as biofilm formation on abiotic surfaces [63]. Another significant protein virulence factor of *A. baumannii* includes the type IV pili system, which comprise a complex of proteins that project from the microbial cell and commonly work to enable the microbe to attach to surfaces. These proteins have been identified as central components in biofilm formation [64] as well as in motility [65].

Other important targets for the development of anti-virulence antibiotics include small molecule secondary metabolites that enable the survival of pathogens. Two common classes of microbial secondary metabolites include 1) siderophores in the acquisition of iron and 2) quorum-sensing (QS) molecules that are responsible for population-dependent coordinated behaviors. Two such secondary metabolites from *A. baumannii* have been characterized: the siderophore acinetobactin [66] and quorum sensing molecule *N*-(3-hydroxydodecanoyl)-L-

homoserine lactone (3-OH-C12-HSL) [67]. Acinetobactin was recently demonstrated to be important in host infection and animal death against mice and the wax worm *Galleria mellonella*. [68] QS molecule 3-OH-C12-HSL has been linked to both biofilm formation [69] and surface motility [65]. QS antagonists against the *A. baumannii* QS receptor have been demonstrated to reduce both biofilm formation and motility [70].

The genome of *A. baumannii* strain ATCC 17978 was sequenced in 2007 [71]. We utilized the genome mining approach on this strain in order to identify genes for enzymes in the nonribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) families. Genome mining of this *A. baumannii* strain revealed the presence of a conserved locus (A1S_0112 to A1S_0119), containing NRPS and PKS enzymes, which is hypothesized to be involved in the biosynthesis of a lipopeptide. This locus was then discovered to not only be present in all sequenced strains of *A. baumannii* but also to be unique to this organism. The evolutionarily conserved and unique nature of this locus in *A. baumannii* has led us to hypothesize that its gene products and secondary metabolite contribute to the virulence of this pathogen.

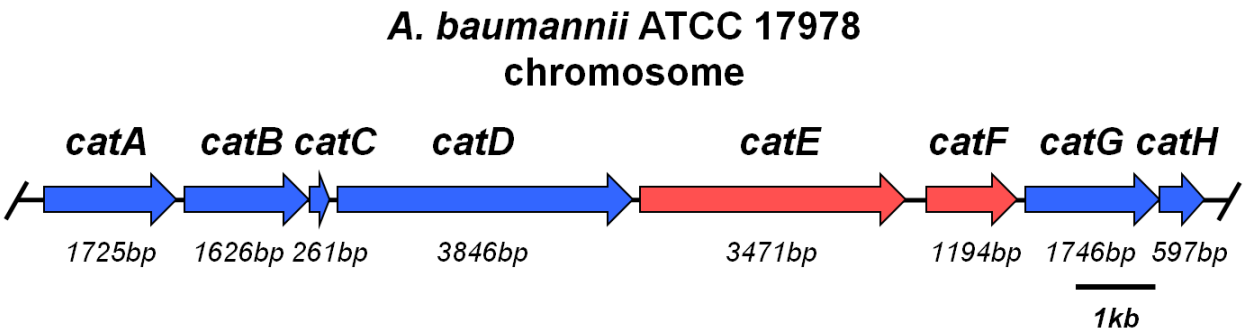
This locus contains a putative efflux pump of the RND superfamily (A1S_0116) and an OMP porin protein (A1S_0117). A previous investigation had demonstrated that this gene cluster played a role in surface motility and was also regulated by an upstream LuxR-type QS regulator, AbaR [65]. Furthermore, others have shown that an acyl carrier protein of this locus (A1S_0114) was transcriptionally upregulated in biofilms [72]. In this study, we generated an insertional mutation in the A1S_0112 gene that conferred altered phenotypes for virulence traits, including surface motility, hemolysis, desiccation resistance, virulence against *G. mellonella*, and heat

shock resistance. Transcriptional analysis of this mutant and comparative phenotypic analysis with a markerless, in-frame deletion mutant revealed a polar effect, resulting in the transcriptional downregulation of the entire A1S_0112-9 gene cluster. Its pleiotropic roles in the pathogenesis of *A. baumannii* prompted us to designate this gene cluster the "cryptic *Acinetobacter* toxin" (CAT).

2.1 RESULTS AND DISCUSSION

2.1.1 Cryptic *Acinetobacter* Toxin Gene Cluster

Genome mining of the *A. baumannii* ATCC 17978 strain revealed a gene cluster A1S_0112-9 predicted to encode for enzymes in the biosynthesis of a hybrid polyketide-nonribosomal peptide (PKS-NRPS) small molecule (**Figure 2.1**). Sequence homology analysis via Basic Local Alignment Search Tool (BLAST) revealed this gene cluster to be specific to *A. baumannii* and absent in other *Acinetobacter* sp. The CAT and acinetobactin gene clusters were also discovered to be the only NRPS and/or PKS genes predicted to be encoded within the *A. baumannii* genome.



Gene	Proposed Function
catA	Acyl-CoA synthetase (Fatty acid ligase-thiolation domains)
catB	Acyl-CoA dehydrogenase
catC	Acyl carrier protein
catD	NRPS (Adenylation-condensation-thioesterase domains)
catE	Efflux pump (RND superfamily)
catF	Outer membrane protein (porin)
catG	Esterase-lipase
catH	Phosphopantetheinyl transferase

Figure 2.1 *cat* gene cluster organization from strain ATCC17978 *A. baumannii* with proposed functions for each open reading frame. Genes colored blue are proposed to be directly involved in the biosynthesis of the putative lipopeptide, while red colored genes are proposed to be involved roles that regulate transport of the putative lipopeptide.

The CAT operon consists of eight genes (*catA-H*). Although the functions of these genes have not be demonstrated, they can be predicted base on sequence homology. Aside from *catE* and *catF* that are predicted to function in substrate transport across the outer membrane, all of the other six genes are predicted to encode proteins related to the PKS and NRPS superfamily and involved in the biosynthesis of the putative lipopeptide. CatA is predicted to be an acyl-CoA synthetase composed of fatty acid ligase-thiolation domains, responsible for the incorporation of a fatty acid substrate. CatB is predicted to be an acyl-CoA dehydrogenase, responsible for dehydration. CatC is predicted to be an acyl carrier protein, which shuttles intermediates from one catalytic domain to the next. CatD is predicted to be an NRPS enzyme composed of adenylation, condensation and thioesterase domains. CatG is predicted to be an esterase-lipase enzyme and involved in the hydrolysis of an ester bond. CatH is predicted to be the phosphopantetheinyl transferase, responsible for posttranslational modification of the acyl carrier protein with 4'-phosphopantetheine prosthetic moiety.

2.1.2 Generation of CAT Pathway Mutants and Transcriptional Analysis to Assess Polar Effects from the Δ*cat*A::Km Mutant

To understand the role of the CAT pathway, we used an approach reported previously [73] to generate an insertional mutant of *A. baumannii* that replaced the *cat*A gene with a kanamycin resistance cassette to yield mutant Δ*cat*A::*Km*. Briefly, a linear, chimeric DNA fragment was generated by overlap-extension polymerase chain reaction (OE-PCR). This fragment contained 5' and 3' DNA sequences, approximately 500bp in size, homologous to the flanking regions of

the *cat*A gene with internal sequences for the kanamycin resistance cassette (**Figure A1**). Kanamycin-resistant clones recombined with the OE-PCR product were subsequently verified through DNA sequencing. In order to assess polar effects conferred by the insertional mutant, a markerless, in-frame deletion mutant of the *cat*A gene (Δ*cat*A) was generated using a yeast recombineering-based suicide vector pMQ30, described previously (**Figure A2**). [74] Markerless, in-frame deletion mutants were also verified through DNA sequencing. Prior to sequence verification, candidate mutants were screened via colony PCR (**Figure A3**).

Insertional mutants generated via gene replacement with antibiotic resistance cassettes have been reported previously to exert polar effects on genes downstream of the replaced gene [75-77]. To investigate the possible polar effects of the Δ*cat*A::*Km* mutant, reverse transcriptase-polymerase chain reaction (RT-PCR) was performed to assess the transcription of downstream genes (**Figure A4**). The inability to generate amplicons from the cDNA library from the Δ*cat*A::*Km* mutant demonstrates a strong polar effect on the transcription of genes within the CAT pathway. This observation also suggests that these genes are co-regulated as an operon. Interestingly, transcription for *cat*H was undetectable under the conditions used. CatH is predicted to be a phosphopantetheinyl transferase required for posttranslational modification of acyl and peptidyl carrier proteins with the 4' phosphopantetheine prosthetic group.

2.1.3 Phenotypic Analysis of CAT Pathway Mutants

Through comparative phenotypic analysis of the wildtype, Δ*cat*A::*Km* mutant and Δ*cat*A mutant, we attribute a number of virulence traits of *A. baumannii* to the CAT pathway. These phenotypes

include desiccation resistance, motility, hemolysis, colony morphology, cytotoxicity against the wax worm *Galleria mellonella*, and heat shock viability.

2.1.3.1 Desiccation

Desiccation entails the removal of intracellular water. Desiccation resistance is characteristic of *A. baumannii* and contributes to its pathogenecity and ability to thrive in hospital settings. [78] Although details into the molecular mechanisms of desiccation resistance are largely unknown, a number of factors related to the preservation of cell structure and function have been attributed to this phenotype in bacteria. In particular, desiccation is known to cause extensive damage to the cell membrane, nucleic acids, and proteins. [79-81] Viability after 24hr of desiccation was assessed for the wildtype, Δ*catA* mutant, Δ*catA*::Km mutant, and the Δ*catA*::Km mutant containing the complementation plasmid (**Figure 2.2**).

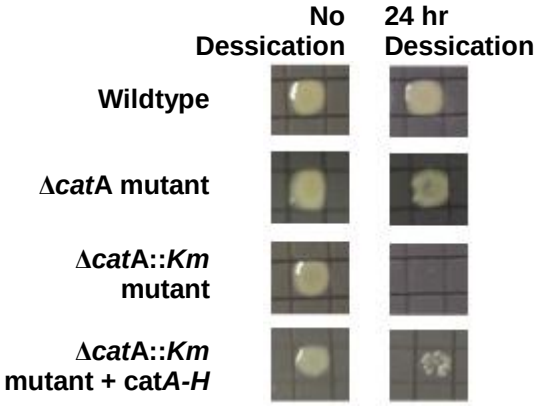


Figure 2.2 Desiccation resistance of wildtype *A. baumannii* ATCC 17978, the Δ*catA* mutant, the Δ*catA*::*Km* mutant, and the Δ*catA*::*Km* mutant containing a complementation plasmid consisting of the *sfp* gene and the *cat*A-H gene cluster. Bacteria were desiccated for 24hr at 37°C on filter paper prior to recovery on tryptic soy agar plates.

Compared to the wild type, a marked decrease in desiccation resistance was observed in the Δ*catA*::*Km* mutant, as demonstrated by an inability to grow following desiccation. Furthermore, introduction of a complementation plasmid partially restored viability of the desiccated Δ*catA*::*Km* mutant. On the contrary, desiccation resistance was comparable between the wildtype and Δ*catA* mutant, suggesting that this phenotype is not significantly attributable to CatA but likely due to other genes in the operon, such as CatE or CatF.

The complementation plasmid introduced into the Δ*catA*::*Km* mutant contained the entire *cat* operon as well as the *sfp* gene. The *sfp* gene encodes a phosphopantetheinyl transferase from *B. subtilis* that posttranslationally modifies apoenzymes of acyl and peptidyl carrier proteins with phosphopantetheinyl prosthetic groups to afford the holoenzyme. [82] Due to the absence of *catH* transcription (**Figure A1.4**), the *sfp* gene was introduced to ensure posttranslational modification of CatA and CatD.

2.1.3.2 Motility

The ability of pathogens to move across surfaces is a critical aspect of its pathogenesis against its hosts and its ability to colonize new environments. This behavior is known to be mediated via complex processes that involve proteinaceous surface appendages, such as flagella and pili [83, 84], as well as biosurfactants [85, 86]. Both Δ*cat*A::*Km* and Δ*cat*A mutants exhibited complete impairment of motility, compared to the wildtype after incubation at 37°C for 24 to 48hr (**Figure 2.3**).

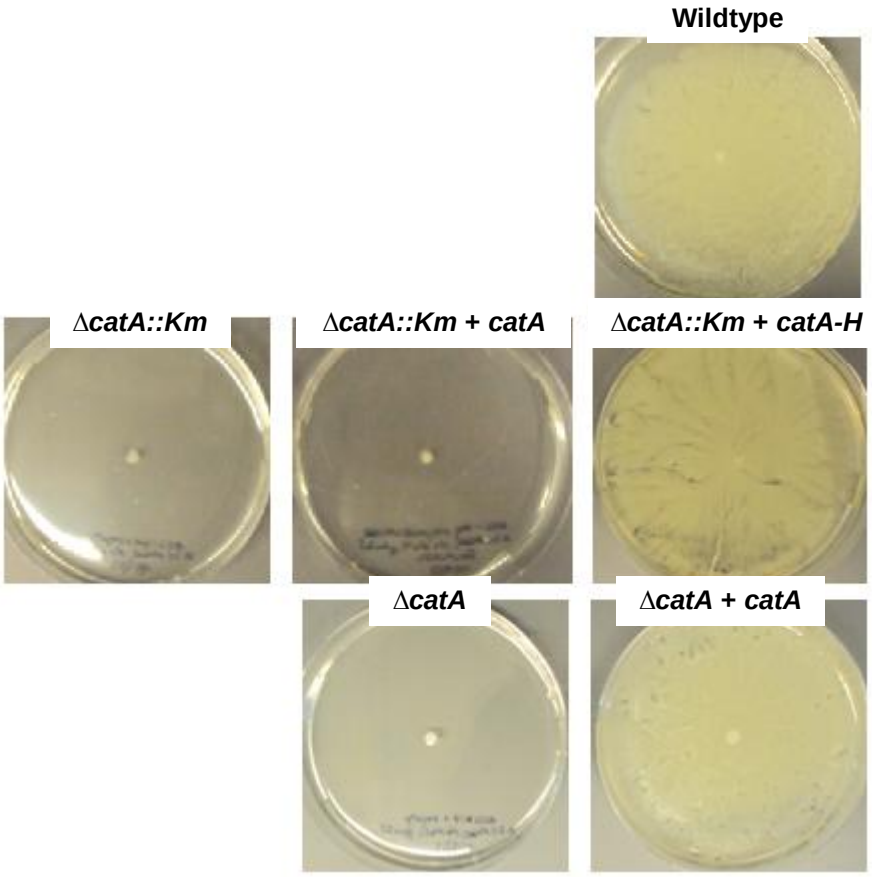


Figure 2.3 Motility phenotypes of the wildtype, Δ*cat*A::*Km* mutant, and Δ*cat*A mutant were assessed on low percentage agar plates after 24 to 48hr incubation at 37°C. The Δ*cat*A::*Km* mutant introduced with plasmids containing either the *cat*A gene alone or the entire operon (sfp and *cat*A-H) was also assessed. The Δ*cat*A mutant introduced with a plasmid containing the *cat*A gene alone was also assessed.

Restoration of the wild-type motility phenotype in the Δ*cat*A::*Km* mutant with the introduction of a complementation plasmid containing the entire operon (*cat*A-H) but not with a plasmid containing the *cat*A gene alone is consistent with the polar effect introduced by this insertional mutation. Restoration of the wild-type phenotype in the Δ*cat*A mutant with the introduction of a complementation plasmid containing the *cat*A gene alone supports a previous finding that suggests a lipopeptide is involved in the motility of *A. baumannii* [65].

2.1.3.3 Hemolysis

Hemolysis involves the destruction of red blood cells and is an important virulence trait that enables pathogens to acquire nutrients from the infected host. Increased hemolytic activity of the wild type compared to the Δ*cat*A::*Km* mutant was observed after incubation for 48 to 72hr at 37°C (**Figure 2.4**).

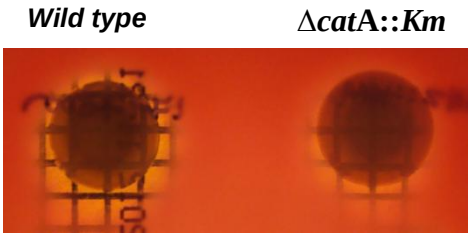


Figure 2.4 Hemolytic activity of the wildtype and Δ*cat*A::*Km* mutant was assessed following 48 to 72hr at 37°C post-inoculation on horse blood agar. Hemolysis was assessed by the presence of a transparent zone around each colony. A gridded background was used to enhance the contrast of the hemolysed region.

A possible mechanism of enhanced hemolytic activity from the wild type could be attributed to the putative CAT lipopeptide and/or CatF, a proposed OMP surface protein.

Microbial lipopeptides have been demonstrated to be hemolytic [87]. CatF, a predicted outer membrane protein, could facilitate the transport of the putative CAT lipopeptide or another hemolytic compound. To test this possibility, a *catF* mutant could be assessed for the hemolytic phenotype.

2.1.3.4 Colony Morphology

Bacterial colony morphology has been shown to be an important indicator of environmental stress and virulence. Previous investigations in *Pseudomonas fluorescens* have demonstrated that the undulate margin bacterial morphology is associated with microbial fitness. [88]

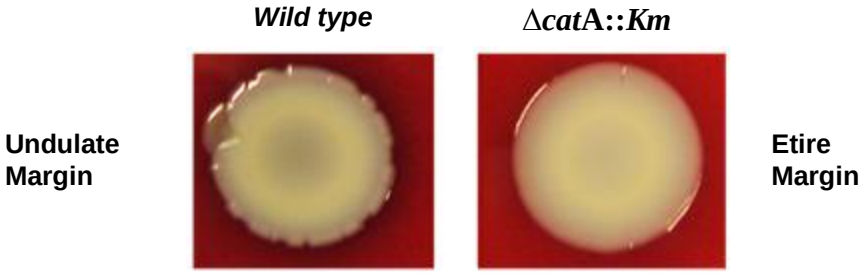


Figure 2.5 The colony morphologies of the wildtype and Δ*catA*::*Km* mutant were observed.

The wild type strain exhibits an undulate morphology along its margin; while on the other hand, that of the Δ*catA*::*Km* mutant is etire. Differences in colony morphology of the *A. baumannii* wild type and Δ*catA*::*Km* mutant (**Figure 2.5**) suggest that the CAT operon is associated with microbial fitness. This observation suggests that the CAT operon confers phenotypic plasticity to *A. baumannii*.

2.1.3.5 Virulence Against *Galleria mellonella*

Galleria mellonella, also known as the greater wax moth, has recently become an important invertebrate model host to investigate pathogenesis of *A. baumannii*. [89] To assess the virulence of the *cat* gene cluster of *A. baumannii*, the wildtype and Δ*catA*::Km mutant were introduced into wax moth larvae and the number of surviving larvae were measured periodically over a span of four days (**Figure 2.6**).

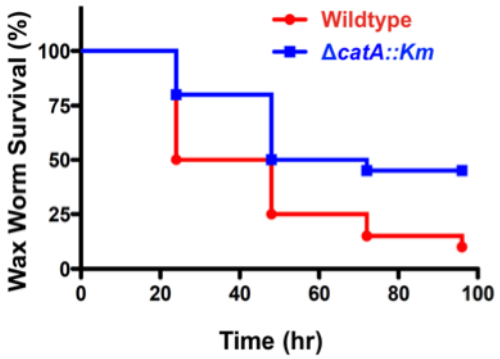


Figure 2.6 The wax worm *Galleria mellonella* was infected with either the wild type or Δ*catA*::Km mutant and the survival rate of *G. mellonella* was monitored over four days.

The survival rate of *G. mellonella* challenged with wild-type *A. baumannii* ATCC 17978 decreased from 48.2% to 22.8% to 13.2% to 8.8% after 24, 48, 72, to 96hr of incubation, respectively. The survival rate of *G. mellonella* challenged with the Δ*catA*::Km mutant decreased from 79.8% to 48.2% to 43.0% to 43.0% after 24, 48, 72, to 96hr of incubation, respectively. After each time point, the survival of *G. mellonella* were markedly decreased with the greatest difference in survivability observed after 24hr with a difference of 31.6% in survival rate between the *G. mellonella* challenged with the Δ*catA*::Km mutant versus the wild type.

2.1.3.6 Heat Shock Resistance

Viability of bacteria following heat shock is an important virulence trait, as microorganisms that invade a host are often exposed to elevated temperature, among other environmental changes. Heat shock induces the upregulation of heat shock regulators that serve to maintain cellular protein structure, primarily through the expression of chaperone proteins. [90] In addition, the heat shock response is known to upregulate the expression of virulence factors. A prominent example of this response is in the increased expression of listeriolysin, a central toxin of the human pathogen *Listeria monocytogenes*. [91]

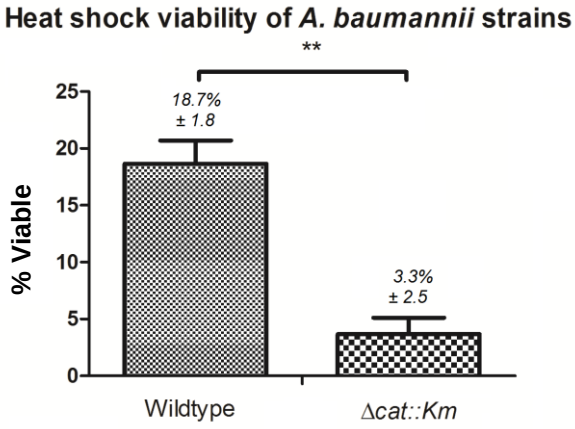


Figure 2.7 Viability following heat shock treatment was assessed for the wild type and Δ*cat*A*::*Km mutant after heating at 55°C for 10 minutes.

The viability of *A. baumannii* was assessed following heat shock at 55°C for 10 minutes (**Figure 2.7**). Quantification of colony forming units (CFUs) of the wildtype and the Δ*cat*A*::*Km mutant revealed a significant difference ($P < 0.01$) in the viability between the two.

2.1.3.7 Regulation of the Motility Phenotype

A Lux-type QS system was recently identified and characterized in *A. baumannii* and its activation has been attributed to virulence-associated phenotypes, including biofilm formation [67] and motility [92]. In addition, disruption of the autoinducer synthase *abaI* resulted in the motility deficiency phenotype, which was restored upon exogenous addition of *N*-(3-hydroxy)-dodecanoylhomoserine lactone (3-OH C12-HSL). [65] In this same study, it was reported that AbaR regulates the transcription of *cat*A and presumably, the remainder of *cat* operon.

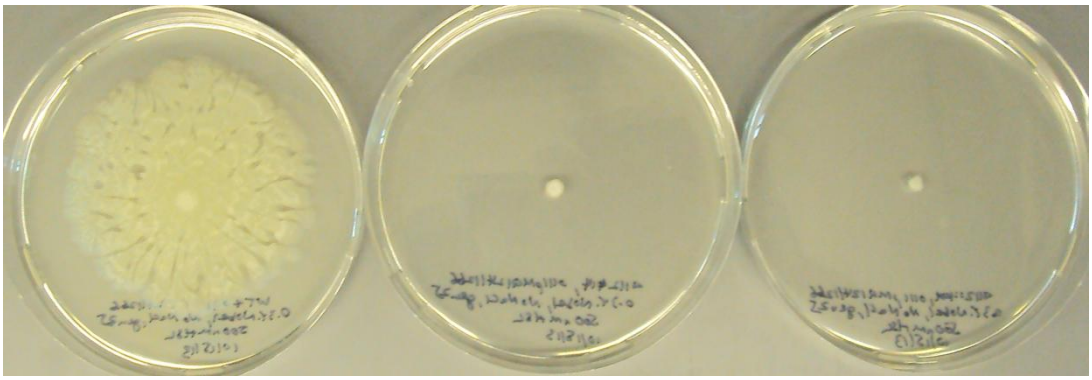


Figure 2.8 Motility deficiency of the Δ*catA* and Δ*catA*::Km mutants was not restored when a neighboring transcriptional regulator *abaR* was overexpressed in the presence of 3-OH C_{12}-HSL.

To ascertain whether expression of *abaR* or the *cat* operon presides over the motility phenotype of *A. baumannii*, we attempted to restore the mutant phenotype from the Δ*cat*A::Km and Δ*cat*A mutants via the overexpression of *abaR* in the presence of 3-OH C12-HSL (**Figure 2.8**). The motility phenotype was not restored in either of these mutants, suggesting that AbaR regulation of swarming requireds the CAT operon.

2.2 DISCUSSION

This study has characterized the roles in pathogenesis of a cryptic pathway of *A. baumannii* that is involved in the biosynthesis of a putative lipopeptide. Phenotypic analysis of an insertional mutant (Δ*cat*A::Km) of *A. baumannii* ATCC 17978 for the *cat*A gene (A1S_0112) revealed that the roles of the *cat*A-H locus (A1S_0112-9) are pleiotropic and include desiccation resistance, motility, hemolysis, colony morphology, heat shock viability, and wax worm virulence. Transcriptional (**Figure A1.4**) and comparative phenotypic (desiccation, motility) analysis with an in-frame deletion mutant of the *cat*A gene (Δ*cat*A) support the finding of a strong polar effect in the insertional mutant Δ*cat*A::Km that significantly diminished expression of the *cat* operon. In addition to *cat* genes proposed to be directly involved in the biosynthesis of the CAT lipopeptide toxin (*cat*A-D and *cat*G-H), there is a predicted efflux pump (CatE) and porin protein (CatF) situated medially within the gene cluster. The observed virulence phenotypes of the CAT pathway are proposed to be attributed to the functions of the CAT lipopeptide, CatE, and/or CatF.

Microbial lipid secondary metabolites commonly exhibit biosurfactant activity. For example, the cyclodepsi-lipopeptide serratamolide of *Serratia* sp. [87], the lipopeptide surfactin of *B. subtilus* [85, 93], and the rhamnolipids of *P. aeruginosa* [86, 94] are all involved in both the hemolysis response as well as in bacterial motility. The biological properties of these microbial lipid-based metabolites supports the view that the motility and hemolysis phenotypes observed from the *cat*A mutants may be attributable to the production of the CAT lipopeptide directly.

Motility was significantly impaired in both the insertional mutant Δ*cat*A::Km and the deletion mutant Δ*cat*A. Motility impairment in the Δ*cat*A mutant suggests that the CAT lipopeptide is required for motility, an observation that is consistent with that of Clemmer *et. al*, who showed that transposon mutants of *catB* (A1S_0113) and *catD* (A1S_0115) also exhibit deficient motility [65]. CatA, CatB, and CatD correspond to an acyl-CoA synthetase, acyl-CoA dehydrogenase, and NRPS, respectively, and are predicted to be directly involved in the biosynthesis of the CAT lipopeptide. Further, restoration of the motility phenotype in the Δ*cat*A::Km mutant only after introduction of a complementation plasmid containing the entire *cat* operon and not one with only the *cat*A gene is consistent with the polar effect of the insertional mutation.

Desiccation resistance was comparable between the wildtype and Δ*cat*A mutant, but this phenotype was absent in the Δ*cat*A::Km mutant (**Figure 2.2**). Because CatA is the acyl-CoA synthetase presumed to be involved in the biosynthesis of the lipopeptide, this observation suggests that the lipopeptide does not significantly contribute to the desiccation resistance phenotype. It follows that either CatE or CatF may be responsible for this virulence trait. *Salmonella* sp. are known for their ability to persist under desiccation conditions for extended periods of time. Characterization of these mechanisms has proceeded through investigations of the osmoregulation response of these bacteria. [95] In one investigation, it was shown that a porin protein (OmpC) of *S. enterica* was upregulated under osmotic stress, presumably as a response to maintain cytoplasmic water. [96] A plausible mechanism for the observed mutant desiccation resistance phenotype of the Δ*cat*A::Km mutant could be explained by the

significantly diminished expression of the porin protein CatF and mal-regulation of osmotic stress.

Consistent with the motile, hemolytic, and desiccation resistance phenotypes characterized from the CAT gene cluster, other general virulence phenotypes, including heat shock viability, colony morphology, and wax moth larvae virulence, were also associated with the CAT gene cluster. Heat shock viability and the rugose colony morphology have been associated with regulatory systems that function to enhance the virulence of the pathogen. [91, 97] Furthermore, our attempts to isolate the CAT lipopeptide from liquid culture broth and bacteria cells via heterologous expression in *E. coli* as well as in *A. baumannii* were unsuccessful (data not shown). Collectively, these findings demonstrate that the CAT gene cluster is involved in the pathogenesis of *A. baumannii*.

2.3 CONCLUSIONS

We have characterized the virulence-associated phenotypes attributable to a cryptic operon (CAT) involved in the biosynthesis of a secondary metabolite lipopeptide of the human pathogen *A. baumannii*. The roles of the CAT locus in the pathogenesis of *A. baumannii* are pleiotropic and include desiccation resistance, motility, hemolysis, colony morphology, heat shock viability, and wax worm virulence. In addition to the lipopeptide toxin, some of these virulence traits are hypothesized to be attributed to the function of a predicted efflux pump and/or porin protein that are co-regulated with the biosynthesis of the lipopeptide.

There exist several future directions that may be pursued to further characterize the physiological significance of the CAT toxin. Notably, the exact genes responsible for virulence phenotypes may be scrutinized by overexpression of individual genes in the wild type and also through complementation studies of the insertional mutant containing the CAT pathway modified with systematic mutations of individual genes of the operon. Additionally, additional secondary metabolite isolation experiments would need to be employed in order to characterize the toxin of interest. Lastly, the contribution of the CAT operon to the virulence of *A. baumannii* can be assessed through invertebrate animal models, e.g., silkworm, fruity fly, roundworm. Nonetheless, these findings demonstrate that the CAT pathway is a desirable drug target for the development of next generation anti-virulence antibiotics.

3.0 DISCOVERY AND CHARACTERIZATION OF A GENE CLUSTER ENCODING A CRYPTIC *PSEUDOMONAS* VIRULENCE FACTOR

Pseudomonas aeruginosa is a gram-negative opportunistic human pathogen, most notably responsible for chronic infections of the airways of patients who are immunocompromised and/or afflicted with cystic fibrosis. [98] Numerous protein and small molecule virulence factors of *P. aeruginosa* enable it to persist and exert pathogenesis in biotic and abiotic environments. Furthermore, a number of protein toxins (exotoxin A, exoenzyme) as well as proteases (elastase, alkaline protease) are known to contribute to its pathogenesis. The success of this pathogen in diverse environments has been attributed to its ability to utilize an array of different virulence factors.

A prominent virulence factor of *P. aeruginosa* includes the formation of biofilms. Biofilms are complex microbial communities housed in polymeric matrices consisting of polysaccharides, proteins, and nucleic acids. The formation of these community structures contributes to the pathogenesis of *P. aeruginosa* and has been shown to develop on both biotic—including tissues of the cystic fibrosis lung [98]—as well as abiotic surfaces, such as catheter lines [99, 100]. Biofilms are especially recalcitrant due to their roles in host colonization and antibiotic resistance that these structures confer to their microbial denizens [101].

In addition to protein virulence factors, a number of secondary metabolites of *P. aeruginosa* also contribute to its pathogenesis. The most prevalent include the rhamnolipids, pyocyanin, quorum sensing molecules (specifically, homoserine lactones), quinolones, lipopolysaccharides, as well as siderophores (**Figure 3.1**). The rhamnolipids are biosurfactants that exert pleiotropic pathogenic consequences against the host, including interference with the immune system [102], inhibition with normal tracheal ciliary function [103], and lysis of blood cells [94]. Furthermore, rhamnolipids have been shown to facilitate motility [104] and the development of biofilm structures [25]. Due to their direct roles in *P. aeruginosa* pathogenesis, the biosynthetic and regulatory pathways of secondary metabolites have been the targets in the development of anti-virulence antibiotics.

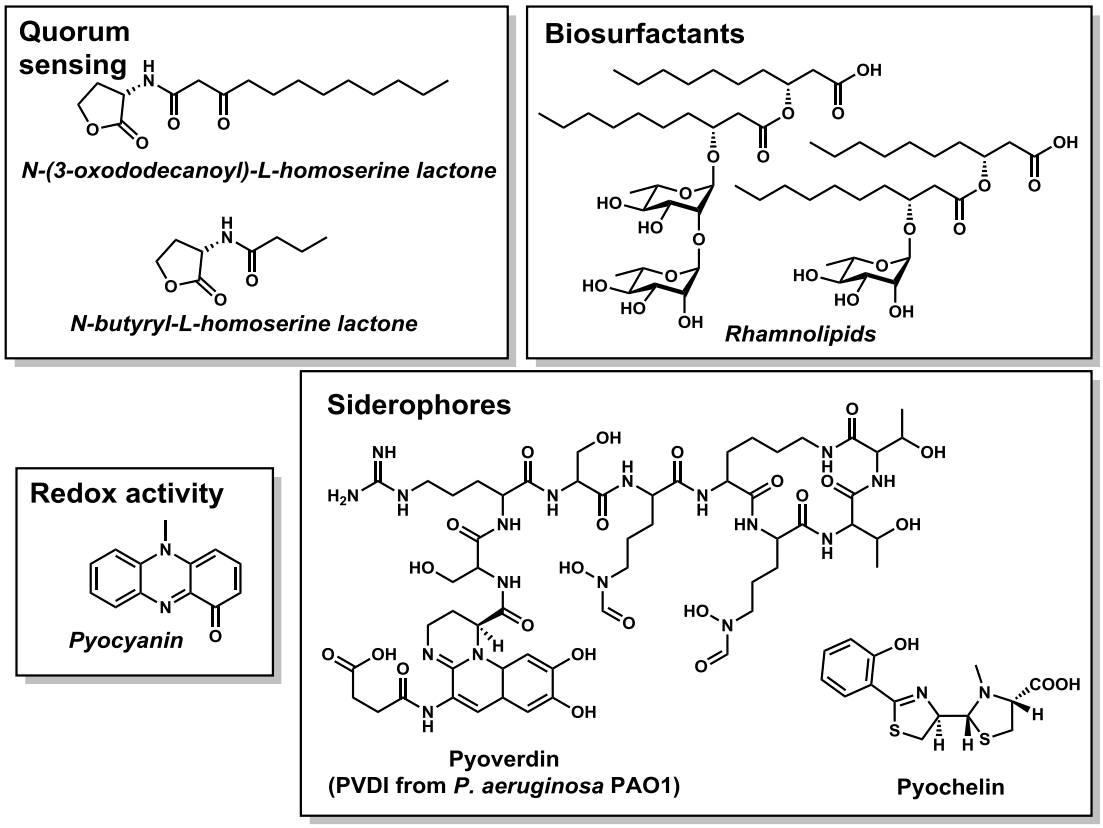


Figure 3.1 Secondary metabolite virulence factors of *P. aeruginosa.*

Pyocyanin is a redox-active metabolite and is the compound that is frequently attributed to the blue-green color of *P. aeruginosa* [105]. The exact role of pyocyanin in the pathogenesis of *P. aeruginosa* is still under investigation, but it has, most notably, been demonstrated to be crucial toward infections of the lung [106]. Furthermore, *in vitro* studies have shown that pyocyanin modulates glutathione recycling within lung epithelial and endothelial cells [107] and can also disrupt catalase activity [108].

Lipopolysaccharides (LPS) are membrane-bound structures embedded on the outer membrane of gram-negative bacteria and elicit strong immune responses from the host. LPS is primarily comprised of a membrane-bound lipid, lipid A, as well as a variable polysaccharide moiety, known as the O-antigen side chain, that is largely responsible for its immunological response in the body [109]. Specifically, lipid A is known to elicit a strong immune response via activation of the Toll-like receptors (TLR), specifically TLR-4. [110] Severe bacterial infections can lead to systemic shock induced by the recognition of high levels of LPS by the host immune system [111]. Due to its roles as a virulence factor, vaccines targeted toward *P. aeruginosa* LPS have been pursued, although with limited success. For example, extracts derived from *P. aeruginosa* growth cultures representing a mixture of LPS showed initial promise as a vaccine candidate by conferring protection yet were also shown to be toxic. [112] A recent example of a promising vaccine candidate is based on a *P. aeruginosa* mutant deficient in aromatic amino acid biosynthesis [113]. It was discovered that while this strain exhibited no adverse immunological effects, it was still able to generate antibodies against LPS and provide protection against cytotoxic *P. aeruginosa* strains [113, 114].

Quorum sensing (QS) in *P. aeruginosa* is regulated by three QS systems, two of which are independent systems that utilize homoserine lactone co-factors to direct the concerted gene expression of virulence factors in a population-dependent manner. Notable virulence traits regulated by QS in *P. aeruginosa* include biofilm formation [115] and motility [116]. These include the *las* and *rhl* systems, which rely upon *N*-(3-oxododecanoyl)-L-homoserine lactone and *N*-butyryl-L-homoserine lactone, respectively, for autoinduction. Due to its hierarchical regulatory roles, QS pathways are attractive targets in the development of anti-virulence therapeutics. [117]

Quinolones exhibit antibiotic activity but have also been shown to be involved in *P. aeruginosa* cell-to-cell signaling pathways that interact with the *las* and *rhl* QS systems of *P. aeruginosa*. The most prevalent quinolone of *P. aeruginosa* includes 2-heptyl-3-hydroxy-4-quinolone, which, due to its roles in regulating virulence gene expression, has been known as the *Pseudomonas* quinolone signal (PQS) [118]. Major virulence traits that have been upregulated by the presence of PQS include the LasB elastase [118], vesicles that transport antimicrobials and toxins [119], and DNA transport [120].

Siderophores are a class of compounds produced by microbes to assist in the acquisition of iron from the environment. As a micronutrient found predominantly as the Fe^{3+} ion, iron is a co-factor for numerous essential enzymatic processes in the cell. Siderophores function by forming soluble complexes with Fe^{3+} and that are transported into the cell via cell surface receptor proteins. There are two major siderophores from *P. aeruginosa*: the pyoverdines and pyochelin. The pyoverdines comprise a structurally diverse group of related siderophores

produced from *Pseudomonas* sp. [121] and have been demonstrated to be the major siderophore produced in clinical isolates of *P. aeruginosa* [122, 123]. In addition to its direct function as a siderophore, pyoverdine production has been shown to be essential in *P. aeruginosa* virulence against mice [124]. Furthermore, pyoverdine has been demonstrated to regulate the expression of other virulence factors, including exotoxin A and endoprotease [125]. The roles that siderophores play in *P. aeruginosa* cellular processes have also made them an attractive target for anti-virulence antibiotic development [126, 127].

A similar approach to identify the CAT locus of *A. baumannii* was utilized to mine the genome of *P. aeruginosa* for species-specific NRPS and PKS loci **(Refer to *A. baumannii*, Chapter 2**). Genome mining of the *P. aeruginosa* PAO1 strain revealed the presence of a cryptic biosynthetic gene cluster consisting of genes PA4078-80 that are conserved within the *P. aeruginosa* species and predicted to biosynthesize a nonribosomal peptide secondary metabolite. The highly conserved nature of this cryptic biosynthetic gene cluster is suggestive to its roles in the survival and persistence of *P. aeruginosa*. In this study, the PA4078-80 locus is demonstrated to be involved in pathogenesis of this organism, notably in its roles in biofilm formation, motility, and colony morphology. Due to its influence on virulence traits of *P. aeruginosa* and its cryptic nature, this locus has been named CPT for its roles in the biosynthesis of a Cryptic *Pseudomonas* Toxin. Its roles in *P. aeruginosa* pathogenesis suggest that the CPT pathway is an attractive target for the development of anti-virulence antibiotics.

3.1 RESULTS AND DISCUSSION

3.1.1 Genome Mining of the Cryptic Pseudomonas Toxin (CPT)

PA4078, PA4079, and PA4080 consist of predicted nonribosomal peptide synthetase (NRPS), dehydrogenase, and LuxR-type transcriptional regulator, respectively (**Figure 3.2**).

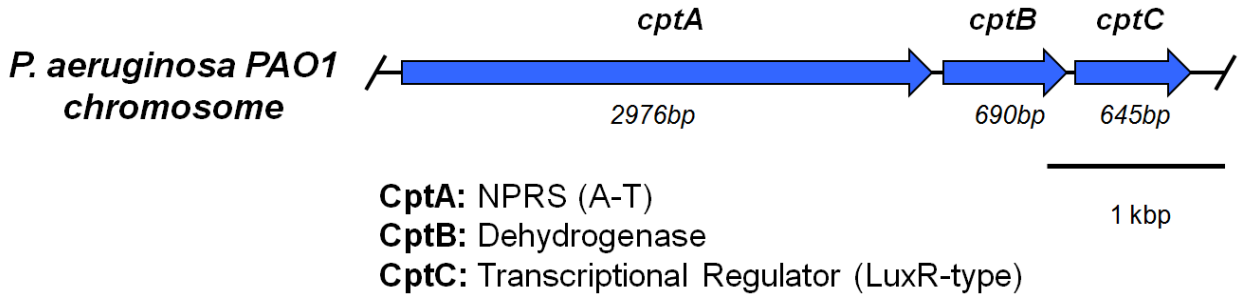


Figure 3.2 Organization of the CPT gene cluster with proposed functions in the biosynthesis and regulation of the CPT toxin.

The NRPS enzyme is responsible for the synthesis of a peptide secondary metabolite that is presumably oxidized by the CptB, a dehydrogenase. The CptC a quorum-sensing modulated, LuxR-type transcriptional regulator is presumably involved in the regulation of CptAB. NRPS enzymes require a posttranslational modification that involves the covalent attachment of the 4'-phosphopantetheine prosthetic moiety, which is facilitated by a phosphopantetheinyl (ppan) transferase. Interestingly, a ppan transferase is not encoded within this locus and is presumable located elsewhere in the chromosome.

3.1.2 CPT Pathway Mutant Generation and Phenotype Complementation

To assess phenotypes attributed to the CPT pathway, an in-frame markerless deletion mutant of the *cpt*AB locus was generated. This mutant was generated by construction of a suicide plasmid based on yeast recombineering [74]. Complementation studies were executed by construction of a plasmid containing the *cpt*AB genes. The complementation plasmid was assembled based on the *Pseudomonas* replicable yeast recombineering shuttle plasmid, pMQ124 [74].

3.1.3 CPT Regulates Biofilm Formation on Abiotic Surfaces

Biofilm represents a bacterial community encased in a matrix composed of polysaccharides, nucleic acids, and proteins. The ability for *P. aeruginosa* to subsist on abiotic surfaces, such as catheter lines [99, 100], has enabled this pathogen to subsist in the hospital setting. Biofilm formation is a significant virulence factor due to its ability to confer enhanced antibiotic tolerance to its microbes [101]. To assess *P. aeruginosa* biofilm formation on an abiotic surface, cultures were grown in a polyvinyl chloride 96-well plate at 30°C for 18hr. The biofilm was assessed via crystal violet (CV) staining and quantified by measuring the absorbance of solubilized CV at a wavelength of 540 nm and normalizing to the optical density (OD) of the bacteria culture at 600 nm.

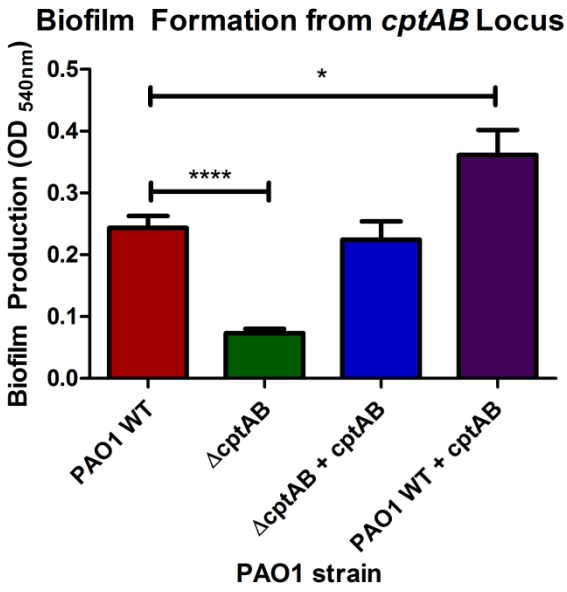


Figure 3.3 Biofilm formation of *P. aeruginosa* strains after 18hr at 30°C growth under static conditions (N = 6). The complementation plasmid includes *cptAB* in pMQ124.

Biofilm formation of the wildtype and mutant *P. aeruginosa* strains were measured (**Figure 3.3**). The *P. aeruginosa* PAO1 wildtype exhibited greater biofilm formation (OD540nm = 0.243 ± 0.046) compared to the Δ*cpt*AB mutant (OD540nm = 0.073 ± 0.017) ($p < 0.0001$). The wildtype biofilm formation phenotype was restored in the Δ*cpt*AB mutant (OD540nm = 0.224 ± 0.066) by introduction of a complementation plasmid containing the *cpt*AB genes (Δ*cpt*AB + *cpt*AB), indicating that the *cpt*AB genes contribute to abiotic biofilm formation in *P. aeruginosa*. Furthermore, overexpression of *cpt*AB in the wild type (OD540nm = 0.361 ± 0.090) increased biofilm formation over the wild type alone ($p < 0.05$). Collectively, these data indicate that biofilm formation is in part positively regulated by the CPT peptide.

3.1.4 Twitching Motility is Impaired in the CPT Mutant

Common bacterial surface structures, such as pili and fimbriae, are known to be involved in the ability of the microbe to adhere to surfaces. The type IV pili are distinct in their ability to promote flagellum-independent motility [128]. It was later discovered that type IV pili enable bacterial translocation via a mechanism where the pili adhere to and retract from its adjacent surface, enabling propulsion of the cell [129]. Motility is an important phenotype that contributes to the pathogenesis of *P. aeruginosa*, as it enables the bacteria to spread throughout the host, acquire nutrients and infect its host.

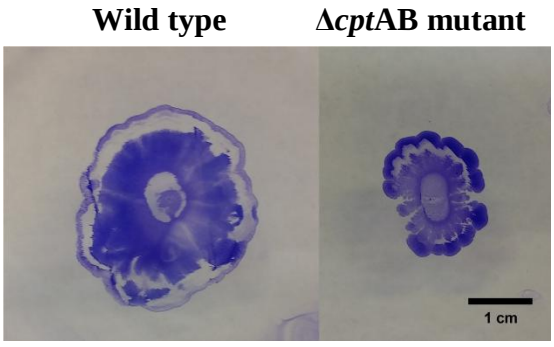


Figure 3.4 Twitching of the PAO1 wildtype and Δ*cpt*AB mutant after incubation at 37°C for 48hr and 25°C for 24hr. Twitching zones were visualized via crystal violet staining.

To assess the twitching motility phenotype in the PAO1 wildtype and Δ*cpt*AB mutant, each were stab inoculated into agar plates and incubated at 37°C for 48hr and then at 25°C for 24hr. Zones of twitching motility were visualized by crystal violet staining (**Figure 3.4**). The twitching motility phenotype was observed to be reduced in the Δ*cpt*AB mutant (twitching zone

diameter = 1.95 cm ± 0.15, N = 3) compared to that in the *P. aeruginosa* PAO1 wild type (twitching zone diameter = 2.83 cm ± 0.18, N = 3) ($p < 0.05$).

3.1.5 Colony Morphology and Autolysis

The roles and mechanisms of bacterial autolysis are unclear but it has been suggested to be a response to environmental stress [130-132]. Specifically, it has been implicated as a mechanism to eliminate unfit members of the population, which reduces nutrient consumption by the unfit, leaving more for the fit and enhances their probability of survival [133]. In gram-positive bacteria, the lack of the autolytic phenotype has been correlated with antibiotic tolerance [134].

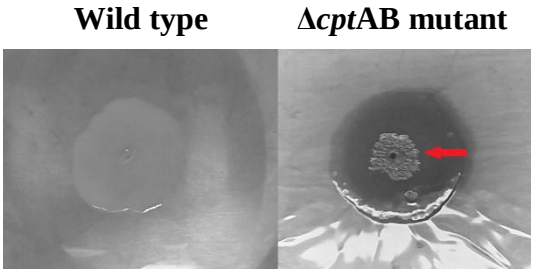


Figure 3.5 Autolysis colony morphology of PAO1 wildtype and the Δ*cpt*AB mutant after incubation at 37°C for 48hr on LB agar.

Colony morphologies of the *P. aeruginosa* PAO1 wildtype and the Δ*cpt*AB mutant were assessed by stab inoculating LB agar plates and allowing the bacteria to grow at 37°C for 48hr and then at 25°C for 24hr. Following the incubation period, the Δ*cpt*AB mutant exhibited marked autolysis, exhibited by plaque-like formations in the center of the growing colony, which

is indicated by the red arrow in **Figure 3.5**. In contrast, the PAO1 wildtype strain did not display any autolytic phenotype and instead appeared smooth.

3.2 DISCUSSION

In this study, a cryptic biosynthetic pathway that is evolutionarily conserved in *P. aeruginosa* was discovered via genome mining. To investigate the role of this NRP secondary peptide, an in-frame markerless deletion mutant (Δ*cpt*AB) of *P. aeruginosa* strain PAO1 was generated. Comparative phenotypic analysis revealed that the deletion of these genes results in reduced virulence traits of this human pathogen. Presumably, the secondary metabolite biosynthesized by CptA and CptB contribute to the pathogenesis of *P. aeruginosa*. Notably, the mutant resulted in reduced abiotic biofilm formation, diminished twitching motility, and the appearance of a colony morphology exhibiting autolysis.

Biofilm formation is an especially prominent virulence trait of *P. aeruginosa* that contributes to its pathogenesis against immunocompromised patients as well as patients afflicted with cystic fibrosis [98]. Abiotic biofilm formation is an important virulence trait, as it enables *P. aeruginosa* to persist in the hospital setting and because PVC is a common material used for medical equipment, including intravenous bags, catheters, collection bags, enteral feeding sets, gloves, and tubing [135]. Compared to wild-type PAO1, the Δ*cpt*AB mutant exhibited significantly reduced abiotic biofilm formation (**Figure 3.3**) on polyvinyl chloride (PVC) plastic. Furthermore, *in trans* complementation of the Δ*cpt*AB mutant phenotype to restore wildtype biofilm formation is evidence that a mutation in *cpt*AB, and not elsewhere in the chromosome, is

responsible for attenuated biofilm formation. Overexpression of the *cpt*AB genes in the wild type also resulted in biofilm formation that exceeded that observed in the wild type without overexpression of these genes. Decreased biofilm formation from the Δ*cpt*AB mutant and enhanced biofilm formation when the *cpt*AB genes are overexpressed suggest that this locus is responsible for abiotic biofilm formation.

The ability for *P. aeruginosa* to twitch across surfaces is dependent on surface proteins of the type IV pili system and involves the attachment to and retraction from its adjacent surface. Compared to the PAO1 wild type, twitching motility was observed to be impaired in the Δ*cpt*AB mutant (**Figure 3.4**). Both impaired twitching motility and biofilm formation from the Δ*cpt*AB mutant could be explained by reduced functionality in the type IV pili. Specifically, it has been shown that mutants of the *pil* genes that encode for the biogenesis and function of *P. aeruginosa* pili proteins were twitching-deficient and unable to form complete biofilms on abiotic surfaces [136]. The role of type IV pili in *P. aeruginosa* in biofilm formation was hypothesized to be involved in either surface attachment stabilization and/or cell-cell interactions in the growing biofilm. In one role, the CPT toxin could directly augment surface attachment, thereby promoting twitching and biofilm formation. The NRP secondary metabolite indigoidine of the marine bacteria *Phaeobacter* sp. have been shown to be important in surface attachment [137]. Alternatively, the CPT peptide could behave as a signaling molecule that regulates biofilm formation, such as the case in QS-dependent cell-cell communication in biofilm formation [115]. Cell signaling roles of CPT could also explain the increased autolysis observed in the Δ*cpt*AB mutant. The *Pseudomonas* quinolone signal (PQS) has been implicated as a signaling molecule that, in addition to several other phenotypes, regulates the autolytic behavior of bacterial

populations. [133] Collectively, comparative phenotypic analysis of the Δ*cpt*AB mutant and PAO1 wild type strains suggest that the CPT secondary metabolite contributes to the pathogenesis of *P. aeruginosa* and may be an attractive target in the development of antivirulence therapeutics.

3.3 CONCLUSIONS

In the present study, a conserved locus of *P. aeruginosa* was identified through the genome mining of *P. aeruginosa* strain PAO1. This locus encodes for the biosynthesis of a nonribosomal peptide and it has been implicated in pleiotropic roles in the pathogenesis of *P. aeruginosa*. Specifically, it has been shown to be an important contributor to virulence phenotypes, including biofilm formation, twitching motility, as well as autolysis. Its significance toward the pathogenesis of *P. aeruginosa* suggests that this biosynthetic pathway can be the target for the development of next generation antibiotics that attenuate virulence traits of human pathogens.

3.4 PART I: OUTLOOK AND FUTURE DIRECTIONS

The increasing emergence of new antibiotic-resistant human microbial pathogens is a serious public health threat. Treatment strategies will require innovative approaches to the discovery and development of novel, next generation antibiotics that circumvent the major drawback of traditional antibiotics—the ability for microbial pathogens to develop drug resistance over time.

A promising approach is to target virulence phenotypes that are nonessential to the cell, thereby minimizing evolutionary selective pressures that give rise to pan-resistant pathogenic strains. Secondary metabolite small molecules have been shown to be attractive targets in the development of next generation antibiotics, some of which contribute toward the pathogenesis of human hosts. The developments from Chapter 2 and 3 extend upon the current understanding of mechanisms in the pathogenesis of human bacterial pathogens *Acinetobacter baumannii* and *Pseudomonas aeruginosa*. Notably, evolutionarily conserved and cryptic gene clusters from these bacterial species were discovered via genome mining approaches. Such conserved gene sequences are proposed to confer survival advantages to these microbes, and indeed phenotypic analysis revealed that these silent gene clusters contribute to human pathogenesis. An important direction forward will entail the isolation and chemical characterization of the secondary metabolites biosynthesized from these gene clusters.

Initial efforts to characterize the metabolites of interest via heterologous expression were unsuccessful. Several major challenges exist in the isolation and elucidation of novel compounds. First, native or heterologous expression of the biosynthetic pathway may require particular chemical inducers or environmental conditions that may require exhaustive optimization in order to achieve. Pathways may be regulated at the transcriptional, translational, and/or posttranslational levels. Second, laboratory culture conditions may lack sufficient quantities of a precursor necessary for the biosynthesis of the secondary metabolite of interest. Lastly, production of the desired compound may be successful but inadequate quantities preclude their detection or pragmatic isolation. Nonetheless, established roles in the phenotypes conferred by these metabolites could aid in the elucidation of their structures. For example, collectively,

current evidence suggests that the CAT virulence factor could be a lipopeptide, based on both bioinformatic predictions as well as phenotypic analysis (e.g., hemolysis and surface motility).

The findings from both of these chapters have demonstrated the essential roles of these cryptic operons toward the virulence of these pathogens, i.e., have revealed a number of genes that may be attractive targets in the development of novel antibiotics.

4.0 PART II INTRODUCTION: BIOSYNTHESIS OF CYANOBACTERIAL NATURAL PRODUCTS

Historically, natural products have been a pivotal source of drug leads, as nearly half of all new chemical entities introduced in the last three decades were natural products, semi-synthetic natural product analogues or natural products-based synthetic compounds.[138] Microorganisms, in particular, have been a prolific reservoir of antibiotics, immunosuppressants, statins, anthelmintics as well as antiviral, anticancer and antiparasitic drugs.[139, 140] These therapeutic compounds originate as secondary metabolites, molecules produced by its host for auxillary and non-essential functions that confer competitive advantages for its producer, but have also realized their potential as therapeutic agents in human medicine.

Before the availability of genomic data, most natural products were discovered through the laboratory cultivation of the producing organism, extraction of the growth media and/or microbes, and then bioassay-guided purification of active compounds—a compound-centric approach. This was highly successful at the time, however, by the 1980s, the number of newly discovered natural products declined rapidly.[141] This decline can be attributed to the re-discovery of natural products commonly produced by a wide range of species as well as the inability to grow and/or produce secondary metabolites under laboratory conditions. An estimated 99% of all bacterial species cannot be cultured.[142] As a result of this decline,

pharmaceutical companies largely abandoned natural products research and development programs.

4.1 NATURAL PRODUCTS DISCOVERY IN THE GENOMICS ERA

A renaissance in natural products discovery has been realized in recent years. This emerging era leverages genomic data to afford novel and compelling approaches toward the characterization of secondary metabolites that were inaccessible by traditional methods. Specifically, bioinformatic analyses have enabled rationally-driven mining of secondary metabolic pathways in which the genes of putative regulatory proteins and biosynthetic enzymes can be identified *in silico* based on sequence homology. These genes can then be explored experimentally through expression studies in a heterologous host or through comparative analyses with defined mutants lacking essential biosynthetic genes. Heterologous expression is advantageous if the compounds are structurally complex and difficult to obtain synthetically or if the native producer is difficult to culture in the laboratory and yields are low.

Genome mining is especially effective in prokaryotes, because biosynthetic genes involved in the same pathway are physically clustered together on the chromosome. A genomics-centric approach is valuable, because it enables the discovery of products from silent biosynthetic gene clusters (i.e., those not expressed under standard laboratory conditions) and also circumvents the common problem of natural products re-discovery. These strategies were employed in the successful characterization of secondary metabolites from *Streptomyces* species,

but they were not without their own challenges, many of which are common in natural products discovery from various microbial sources.

4.2 THE SUCCESS OF STREPTOMYCES AS A SOURCE OF NATURAL PRODUCTS

A hallmark characteristic of microorganisms of the *Actinomycetales* order is their ability to produce numerous secondary metabolites, and members of the *Streptomyces* genus exemplify this trait through their prominent production of antibiotics, over two dozen of which are used clinically.[140] Recent genome sequencing of several *Streptomyces* strains revealed their remarkable capacity to produce over twenty secondary metabolites, most of which are unknown, i.e., cryptic. Notably, investigations into the model strain, *S. coelicolor* A3(2), genome revealed the presence of eighteen silent biosynthetic gene clusters to which their products are cryptic—by comparison, this strain was only known to produce three secondary metabolites prior to the elucidation of its genome sequence.[143] The rich biosynthetic faculties of *Streptomyces* served as an impetus to access these new molecules, but major technical challenges impeded early investigations.

Streptomyces are gram-positive bacteria that undergo complex lifecycles, from germinating spores to multinucleoid substrate mycelium and aerial hyphae.[144, 145] Furthermore, notable features of the *Streptomyces* genome include their relatively high GC content[143] as well as linear chromosomes[146]. Their genetic instability has been known for decades[147], but only recently has its molecular basis been elucidated. Specifically, it was

discovered that numerous tandem inverted repeat sequences located at the ends of these linear chromosomes were responsible for recombination events that caused circularization and re-linearization of the chromosome, resulting in frequent and spontaneous acquisition as well as loss of large segments of DNA.[148-151] Therefore, early investigations that involved the direct genetic manipulation of many *Streptomyces* members were fundamentally challenging. However, because of the compelling need to access these molecules, numerous genetic tools have been developed in order to isolate and characterize novel secondary metabolites from these organisms.

4.2.1 Heterologous Expression Systems for *Streptomyces*

As a consequence of the significance of *Streptomyces* secondary metabolites, many expression plasmids and heterologous hosts have been developed in order to elucidate the biosyntheses of these compounds. Due to the ease of their acquisition of foreign DNA molecules, *S. coelicolor*, *S. lividans*, and *S. albus* were early suitable candidates for the cloning and heterologous expression of *Streptomyces* genes.[152] Later, the sequencing of the *S. coelicolor* A3(2) genome established the framework for further systematic genetic investigations. These *Streptomyces* species were further optimized for heterologous expression by deleting endogenous secondary metabolite biosynthetic pathways that can compete for substrates [153, 154] and also by minimizing the genome to only essential genes, thereby reducing the probability of undesirable genome shuffling [155]. The success of *Streptomyces* natural products discovery has set an example for the investigation of elusive secondary metabolites from new microbial sources, notably other *Actinomycetes*, cyanobacteria, and unculturable bacteria. The strategies employed

in accessing valuable natural products from *Streptomyces* highlight the principle methods necessary to unlock cryptic compounds from other elusive sources.

4.3 ENZYMOLOGY OF NATURAL PRODUCTS BIOSYNTHESIS

Despite the remarkable diversity of natural products, their core assemblies arise from a relatively few number of enzyme families and biosynthetic pathways, which can be further diversified via various auxiliary modifications and utilization of atypical substrates. The widespread occurrence of these biosynthetic pathways is evidence to their physiological significance and underscores the importance of these molecules in drug discovery and development. The following subsections will highlight several major classes of natural products and general features of their biosyntheses.

4.3.1 Microbial Nonribosomal Peptide and Polyketide Secondary Metabolite Biosynthesis

Of all the microbial natural products, which are represented from several major classes of molecules, including terpenoids, oligosaccharides, alkaloids, and ribosomally synthesized peptides, nonribosomal peptides (NRPs) and polyketides (PKs) have been especially fascinating due to their vast structural diversity from rather simple building blocks. Further, NRP synthetase (NRPS) and PK synthase (PKS) enzymes proceed by predictable enzymatic logic in an assembly line-like fashion to generate their product(s) of interest. In this assembly line, the substrates are tethered via a thioester linkage to a carrier protein (acyl or peptidyl carrier protein, ACP or PCP,

for PKS and NRPS enzymes, respectively) and transferred to downstream enzymes for subsequent chemical transformations. An important feature of ACPs and PCPs includes the prosthetic group 4'-phosphopantetheine, which is covalently attached in a conserved serine residue of these proteins. This moiety results from posttranslational modification by a phosphopantetheinyl transferase and serves as the "swinging arm", shuttling intermediates from one catalytic domain of the PKS or NRPS to the next.

The backbone of these NRP and PK compounds comprise of amides and C–C linkages via condensation reactions. Other transformations include oxidations and halogenations as well as decorations with acyl, glycosyl, and prenyl moieties among others. NRPS substrates can also vary widely from all proteinogenic as well as non-proteinogenic amino acids, including D-isomers. Likewise, PKS substrates can often differ by carbon number and subsequent enzymatic modifications can yield saturated and unsaturated acyl moieties.

Decarboxylation and Claisen condensation yield the backbone linkages between PKS substrates, but other common enzymatic reactions exist to generate diversity (**Figure 4.1**). To begin PK assembly, a ketide unit is tethered to the ACP domain via thioester bond with support from an acyl transferase (AT) domain and is subsequently transferred to the first ketosynthase (KS) domain. Decarboxylation of a downstream ketide unit affords a nucleophilic α-carbon that attacks the thioester linkage of an upstream ketide-*S*-ACP intermediate to generate the diketide. The ketide can also undergo reduction and dehydration processes to yield saturated or unsaturated products via ketoreduction, dehydration, and enoyl reduction by ketoreductase,

dehydratase, and enoyl reductase domains, respectively. In this way, PKS pathways strongly resemble fatty acid biosynthetic pathways.[156]

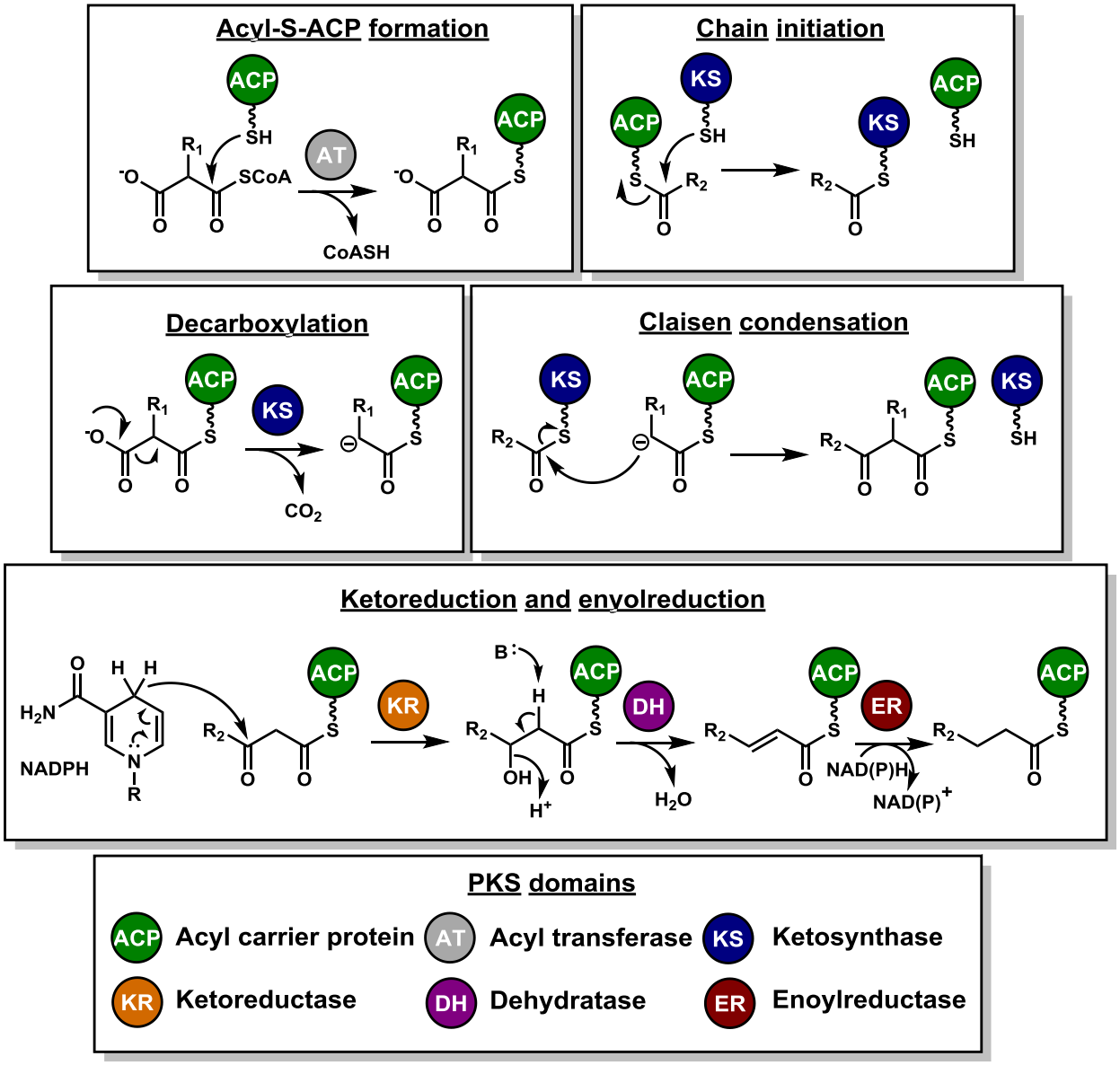


Figure 4.1 The key steps in the assembly of polyketides is shown. Note that the biosynthesis of polyketides is analogous to fatty acid biosynthesis. Figure adapted from [156].

The 6-deoxyerythronolide B synthase (DEBS) of *Saccharopolyspora erythraea* is the hallmark type I PKS assembly line and illustrates the elegance of PK assembly. DEBS is responsible for the biosynthesis of the aglycon component of antibiotic erythromycin A (**Figure 4.2**).[157, 158] A thioesterase (TE) domain at the C-terminus of DEBS facilitates cyclization and chain release, a common feature of both PKS and NRPS assemblies.

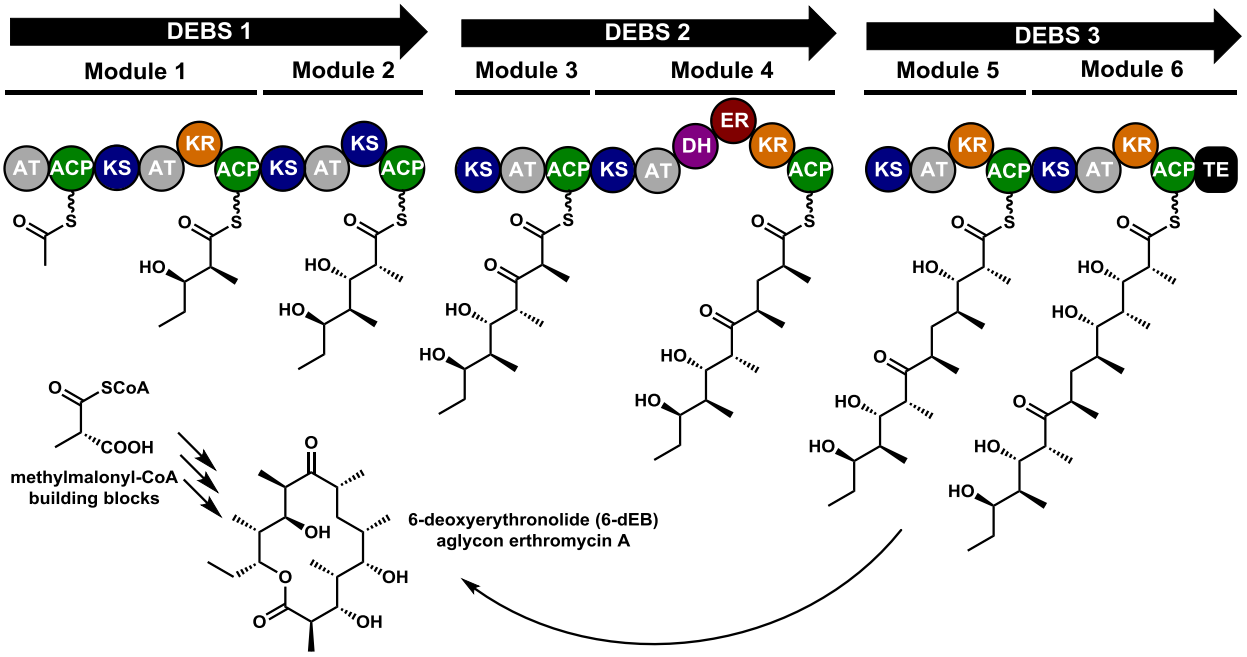
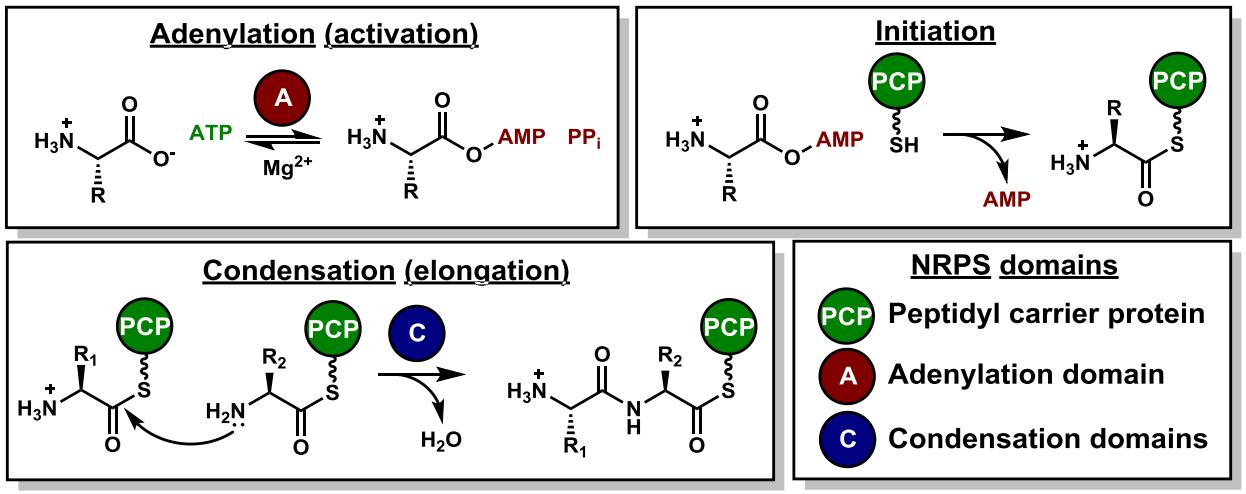


Figure 4.2 The biosynthesis of the aglycon structure of erythromycin A is facilitated by PKS enzymes in an assembly line-like fashion. Figures adapted from [157, 158].

NRP assembly is similar to ribosomal peptide biosynthesis in which amino acid monomers are activated as mixed esters and proceeds condensation to form amide bonds between individual units (**Figure 4.3a**).[159] Specifically, amino acid monomers are activated to form the adenylyl-amino acid mixed anhydride via the adenylation domain, which is responsible

for amino acid specificity. Adenosine triphosphate (ATP) is required for this process, yielding the adenosine monophosphate (AMP) product. The activated monomer is subsequently tethered to the PCP domain via a thioester linkage and is ready for nucleophilic attack by an upstream amino acid with the aid of the condensation domain. The biosynthesis of the cyclic lipopeptide surfactin A illustrates the assembly of various amino acids to produce a final cyclic lipopeptide (**Figure 4.3b**).[160, 161]

(a)



(b)

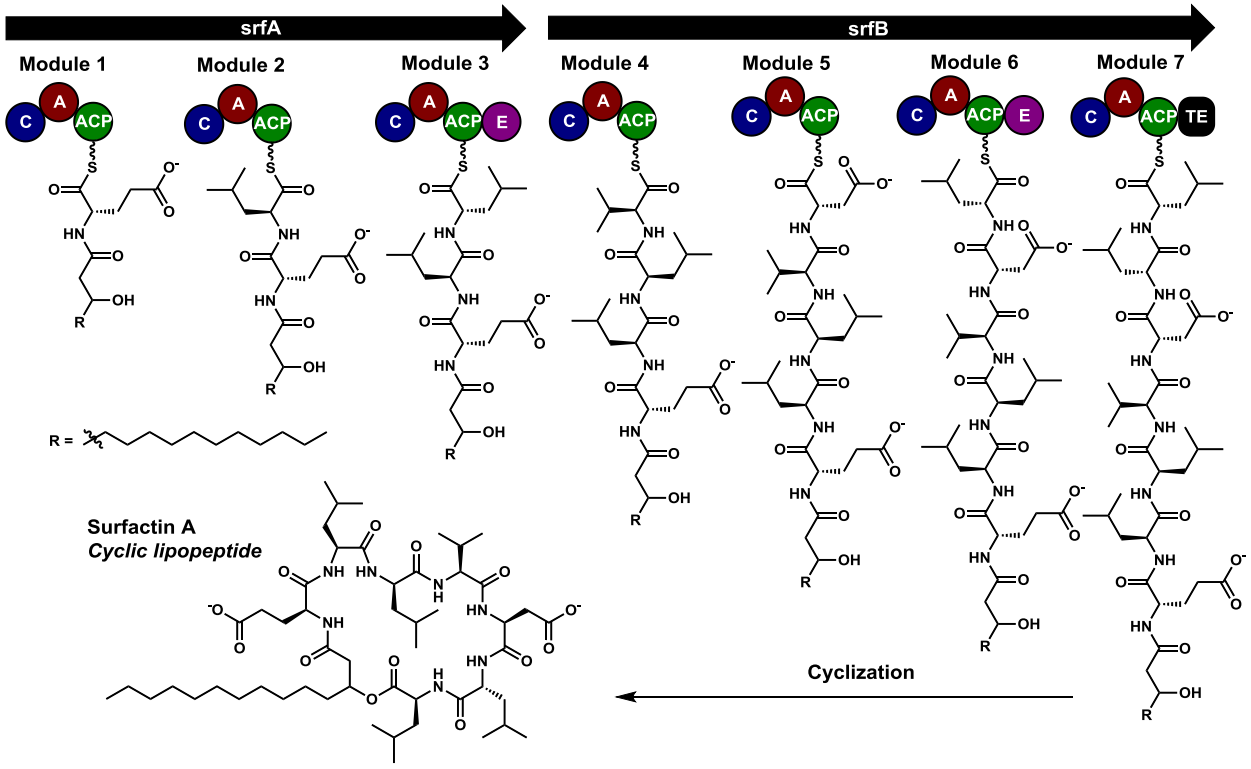


Figure 4.3 (a) The key steps in the assembly of NRPs. (b) The biosynthesis of the cyclic nonribosomal peptide surfactin from the *srf operon*. Figures adapted from [159-161].

In addition to the domains responsible for substrate incorporation, a vital component of many PKS and NRPS assembly lines is the C-terminal TE domain. Not only is this important component responsible for chain release of the mature PK or NRP product but, when necessary, it can also cyclize the product to its active form. In both of the previous examples of aglycon erythromycin A and surfactin A biosynthesis, cyclization and chain release occur upon nucleophilic attack from a side chain hydroxyl moiety to the thioester tether, both mediated by the TE domain.

4.3.2 Ribosomal peptides

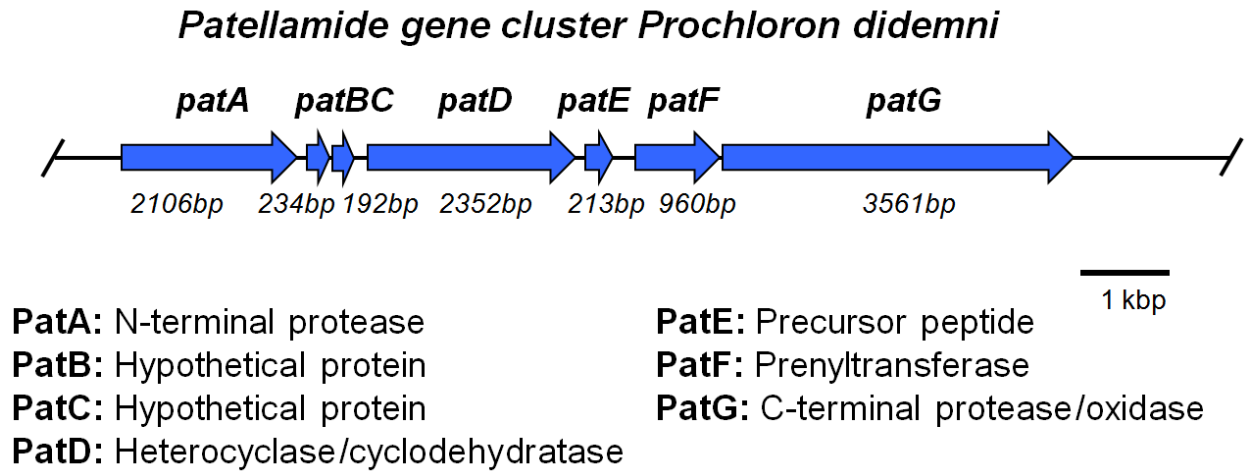
Similar to nonribosomal peptides (NRPs), ribosomally synthesized and posttranslationally modified peptides (RiPPs) exhibit diverse biological roles (cofactors, antibiotics, cytotoxins, quorum sensing[162]) and originate from diverse sources, such as prokaryotes (lantipeptides[163], linaridins[164], thiopeptides[165], microcins[166]), fungi (amatoxins, phallotoxins[167]), plants (cyclotides[168]), and animals (conopeptides[169]). Their extensive distribution throughout many species testifies to their physiological significance. Unlike NRPs, RiPPs are composed exclusively of proteinogenic and D- amino acids that are assembled by the ribosome to afford a peptide scaffold. They can sometimes undergo posttranslational modification, including heterocyclization, oxidation, and prenylation, that can yield products that may resemble their NRP counterparts.

With over one hundred structures elucidated, cyanobactins are RiPPS that belong to one of the largest families of bioactive small molecules from cyanobacteria, having been identified in up to 30% of all cyanobacterial species.[170, 171] Cyanobactins are especially important

because not only can they be explored as therapeutics, such as antibiotics as well as antiviral, antimalarial, and anticancer lead compounds [170, 172, 173] but also because they are known to be toxic to animals [174]. The initial characterization of the patellamides and its biosynthetic gene cluster from *Prochloron didemni* a cyanobacterial symbiont of the ascidian *Lissoclinum patella*, established the groundwork for the subsequent bioinformatics-guided discovery of numerous other cyanobactin gene clusters (**Figure 4.4**). [175] Hallmark features elucidated from the patellamide biosynthetic pathway include a precursor peptide that matures to the final cyanobactin structures and two proteases, one of which is also responsible for N-to-C macrocyclization.[175, 176] The characterization of other cyanobactin gene clusters revealed that these three features were conserved, but others exist, such as enzymes responsible for prenylation and methylation [177] as well as for heterocyclization [178] to afford thiazol(in)e and oxazol(in)es from cysteine and serine or threonine residues, respectively. Recently, it was discovered that some cyanobactin pathways can also afford highly modified linear peptides. [177]

Each cyanobactin biosynthetic gene cluster has been known to produce multiple related structures. This early observation was followed by intense interest in understanding how structural diversity is generated from a single precursor peptide gene encoded in the pathway. Notably, the tailoring enzymes are conserved amongst various cyanobactin gene clusters, yet the substrate precursor peptides are highly diverse within a gene cluster and amongst different gene clusters. This suggested that the precursor peptide gene itself contained the information required for structure diversity.

(a)



(b)

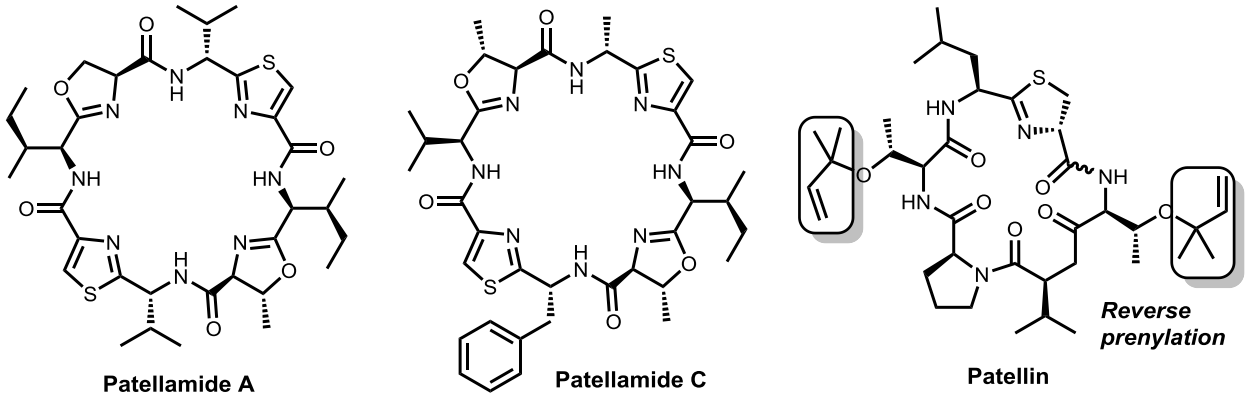


Figure 4.4 (a) Organization of the patellamide gene cluster and established gene functions. **(b)** Structures of patellamides A and C from the *pat* gene cluster and structure of patellin from a *pat*-like gene cluster *tru*, which is responsible for the biosynthesis of the cyanobactins trunkamides and patellins. Note that the presence of a prenyltransferase (patF) does not result in prenylation of the final patellamide structures. In contrast, the prenyltransferase of the *tru* gene cluster is responsible for reverse prenylation of patellin. Figure adapted from [179].

Sequence analysis of various precursor peptides from different cyanobactin gene clusters revealed highly conserved consensus sequences, later established to be leader and recognition sequences [180-183], and hypervariable regions, corresponding to the core peptides present in the final structures [184, 185] (**Figure 4.5**). Although the N-terminal leader sequence has no effect on product formation [186], the recognition sequences proved to be essential for proteolytic cleavage of the precursor as well as macro- and heterocyclization modifications [180-182]. Notable divergence in the core peptide sequence and the translation of multiple precursor peptide sequences serve as the basis for marked cyanobactin structural diversity [184, 185].

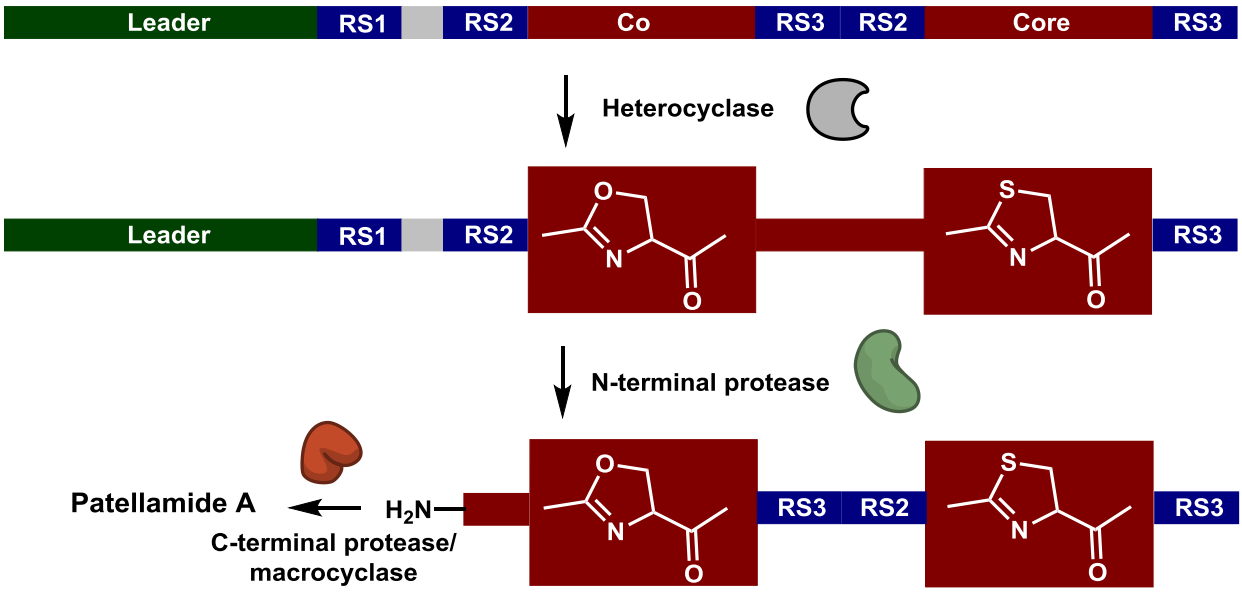


Figure 4.5 The precursor peptide sequence contains instructions for subsequent posttranslational modification to yield the final cyanobactin structure. These include three recognition sequences (RS1-3), a leader peptide, and the core peptide. The heterocyclase encoded in the cyanobactin gene cluster is recognizes RS1 to cyclize threonine, serine, and cysteines to their corresponding azolines. Subsequently, the N-terminal protease binds to RS2 and cleaves the leader peptide. Lastly, the C-terminal protease binds to RS3 and is also responsible for macrocyclization. Adapted from [183].

Therefore, once the precursor peptides are translated, the PatA homolog (protease of the patellamide gene cluster) cleaves the cyanobactin precursors at the N-terminus. [187] Following cleavage, the PatG protease homolog recognizes the C terminus of the precursors and subsequently executes macrocyclization [187, 188]. In some cyanobactin biosynthetic pathways, PatG is also responsible for oxidation of the heterocycle to afford the azole. [187, 188] Collectively, these findings are significant because the tailoring enzymes are highly conserved yet the precursor peptide sequences exhibit substantial divergence. This substrate promiscuity can enable the rational engineering of the cyanobactin biosynthetic platform to generate defined libraries of macrocyclic peptides with therapeutic applications. In one investigation, unnatural amino acids were introduced into the backbone of the cyanobactin trunkamide to afford a library of novel macrocyclic structures that could theoretically number in the tens of thousands.[189]

4.4 HETEROLOGOUS EXPRESSION OF CYANOBACTERIAL BIOSYNTHETIC PATHWAYS

It has long been observed that cyanobacteria are prolific producers of bioactive natural products. Given the inhospitable and often extreme environments in which cyanobacteria species subsist, it is believed that a diverse repertoire of secondary metabolites confer survival advantages against their competitors in what can be considered “chemical warfare”.[190] Furthermore, recent genome sequencing of several dozen strains revealed that cyanobacteria are, in fact, capable of producing significantly more secondary metabolites than have been characterized, with some strains dedicating at least 5% of their genomes exclusively to secondary metabolite biosynthesis.[171]

Despite the immense potential for natural products discovery from cyanobacteria, access to these molecules has been challenging, due to their cryptic nature and/or scarce quantities obtained through laboratory cultivation. There are several notable features of cyanobacteria that have hindered their use in genetic investigations. They can be unicellular or multicellular and can possess multiple copies of their chromosomes per cell, complicating genetic manipulations with high fidelity.[191] Additionally, they present extracellular mucilaginous sheaths that inhibit the introduction of exogenous molecules, including DNA. Any exogenous DNA that can surpass the physical barriers of the cyanobacterial cell would then be challenged by restriction nucleases that can degrade them.[192] Furthermore, cyanobacteria grow slowly, some strains with doubling times approaching six days[193] (in comparison, *E. coli* can double in twenty minutes under ideal growth conditions). Lastly, few regulatory genetic elements, such as promoters, terminators, transcription factors, and ribosome binding sites have been characterized in cyanobacteria, making genetic engineering difficult to obtain in these organisms.

Given the incentive to characterize these potentially valuable natural products but also the major challenges to their direct access through genetic manipulation, the development of cloning methods and heterologous hosts for the production of these compounds has become an increasing attractive approach. However, there are few reported examples of cloning and expression systems proven to successfully assembly and express entire cyanobacterial biosynthetic pathways containing complex enzymes, such as NRPS or PKS. In one investigation, the gene cluster for the molluscicidal agent barbamide from cyanobacterium *Moorea producens* was successfully produced in a *Streptomyces* host.[194] In this study, the 26 kilobase (kb) barbamide gene cluster was assembled via a series of sub-cloning steps. Expression yielded a

novel derivative of the hybrid NRP/PK natural product, 4-*O*-demethylbarbamide, that was shown to exhibit more potent activity than the parent molecule. In another study, the gene cluster for the NRP-terpene hybrid and protein kinase C activator lyngbyatoxin from *M. producens* was expressed in *E. coli* [195]. Promoter engineering of the fosmid containing the lyngbyatoxin biosynthetic pathway (*ltxABCD*) afforded high titers (25.6 mg L^{-1}) of the final product. This finding was especially significant because a previous investigation of the heterologous expression of the *ltxABCD* pathway in a *Streptomyces* host resulted in premature transcriptional termination of *ltxA*, resulting in no production of lyngbyatoxin.[193] This example underscores the significance in the choice of heterologous host as well as some of the regulatory complexities that can be encountered in establishing a cloning and expression system for cyanobacterial natural products.

Unlike for the complex mechanisms of NRP and PK assembly, the cloning and heterologous expression of cyanobacterial RiPPs, specifically ribosomally synthesized cyanobactins, have proven more successful, particularly due to their comparatively simpler biosynthesis. The successful expression and characterization of the patellamide biosynthetic pathway from *Prochloron didemnii,* a cyanobacterial symbiont of the ascidian *Lissoclinum patella*, revealed the canonical gene cluster organization that enabled the facile identification of other cyanobactin gene clusters [175]. Other cyanobactin biosynthetic pathways that have been cloned and heterologously expressed in *E. coli* include those for the piricyclamides[196], microcyclamides [197], microviridins [198], trunkamides [184], and anacyclamides [199]. Given the highly conserved tailoring enzymes amongst the various cyanobactin biosynthetic pathways and the common building blocks upon which they are assembled (i.e., proteinogenic amino

acids), the expression of one cyanobactin gene clusters highly suggested that subsequent clusters could also be expressed successfully.

5.0 A PLASMID TOOLKIT FOR THE RAPID CHARACTERIZATION OF CYANOBACTERIAL NATURAL PRODUCTS BIOSYNTHESIS

Cyanobacteria are prolific producers of bioactive natural products, including antibiotics, anticancer, and anthelmintics. Recent genome sequencing of dozens of cyanobacterial strains revealed they are capable of producing significantly more secondary metabolites than have been characterized. Some strains dedicate at least 5% of their genomes exclusively to secondary metabolite biosynthesis, notably through nonribosomal peptide synthetases (NRPS), polyketide synthases (PKS), and cyanobactin biosynthetic pathways.[171] However, despite this vast reservoir, biochemical investigation into many of these cyanobacterial products has been elusive, primarily due to their cryptic nature and the scarce quantities that can be obtained through laboratory cultivation. Furthermore, the physiological characteristics of cyanobacteria have proven these organisms to be difficult to manipulate genetically, further complicating investigations into their secondary metabolism. Notably, they can be unicellular or multicellular and can possess multiple copies of their chromosomes per cell.[191] They also exhibit extracellular mucilaginous sheaths that inhibit the introduction of exogenous molecules, including DNA. DNA that can surpass the physical barriers of the cyanobacterial cell can be challenged by restriction nucleases.[192] Furthermore, cyanobacteria grow slowly, some strains with doubling times approaching six days[193]. Lastly, few regulatory genetic elements

(promoters, terminators, transcription factors, ribosome binding sites) have been characterized in cyanobacteria.

These major challenges have prompted the development of new cloning methods and the exploration of suitable heterologous expression hosts. There are few proven examples of cloning and expression systems that can successfully assembly and express entire cyanobacterial biosynthetic pathways. In one investigation, the gene cluster for the molluscicidal agent barbamide from cyanobacterium *Moorea producens* was successfully produced in a *Streptomyces* host.[194] The large barbamide gene cluster (26 kbp) was assembled via a series of sub-cloning steps. Expression yielded a novel derivative of the hybrid NRP/PK natural product, 4-*O*-demethylbarbamide, that was shown to exhibit more potent activity than the parent molecule. In another study, the gene cluster for the protein kinase C activator lyngbyatoxin (an NRP-terpene hybrid) from *M. producens* was expressed in *E. coli* [195]. Promoter engineering of the fosmid containing the lyngbyatoxin biosynthetic pathway (*ltxABCD*) afforded high titers (25.6 mg L^{-1}) of the final product. A number of cyanobactins of ribosomal peptide pathways from cyanobacteria have also been successfully characterized via cloning and heterologous expression in *E. coli*. The successful expression and characterization of the patellamide biosynthetic pathway from *Prochloron didemnii*, a cyanobacterial symbiont of the ascidian *Lissoclinum patella*, revealed the canonical gene cluster organization that enabled the facile identification of other cyanobactin gene clusters [175]. Other cyanobactin biosynthetic pathways that have been cloned and expressed include those for the piricyclamides[196], microcyclamides [197], microviridins [198], trunkamides [184], and anacyclamides [199].

The characterization of these cyanobacterial biosynthetic pathways has predominantly relied on traditional (ligation-based) cloning vectors. Two major drawbacks to ligation-based cloning strategies include (1) the limited selection of restriction enzymes suitable for cloning and (2) the multiple rounds of subcloning steps required to co-assembly multiple genes together. The latter is an especially important consideration in the cloning of biosynthetic pathways that are large (e.g., NRPS or PKS-containing pathways) or that contain several genes (e.g., the cyanobactin pathways typically contain at least five or more). In contrast to ligation-based cloning, yeast recombineering has proven to be a promising alternative in the assembly of complete biosynthetic gene clusters. Recombineering relies on yeast's natural mechanism in the repair (recombination) of homologous sequences of DNA [200]. Previous yeast recombineering methods utilize PCR primer designs to yield linear amplicons and a linearized plasmid with homologous end sequences (as few as 30bp in length) that seamlessly recombine to produce the final plasmid construct [200-203]. To address these major limitations of currently-utilized plasmid vectors in the investigations of cyanobacterial biosynthetic pathways, a new designer plasmid based on yeast-recombineering is proposed here. This designer plasmid can be used to selectively assemble native biosynthetic pathways from diverse cyanobacterial species in a single cloning step. In addition, this designer plasmid can be utilized to re-assemble biochemical pathways, to engineer genetic regulatory sequences, and to generate unnatural analogs in heterologous hosts. This designer plasmid is shown to be utilized in the systematic investigation of various biosynthetic pathways (cyanobactin, NRPS, PKS) of natural products from diverse cyanobacterial species and can be used as a tool to facilitate the discovery of novel natural products from these microorganisms.

5.1 DESIGNER PLASMIDS ARE USED TO STUDY DIVERSE BIOSYNTHETIC PATHWAYS FROM VARIOUS CYANOBACTERIAL SPECIES

Hallmark features of an expression platform for natural products biosynthesis consists of a cloning plasmid that can be utilized to rapidly assemble large fragments of DNA and that can be introduced into a laboratory amenable heterologous host for expression. This expression system utilizes homologous end sequences introduced via PCR primers to assemble complete biosynthetic pathways. Homologous end sequences can also be tailored to introduce synthetic regulatory elements, such as ribosome binding sites (RBS), and also enable site-directed mutagenesis (e.g., for codon optimization). The proposed cyanobacterial expression plasmid described herein has been utilized to characterize natural products from a diversity of biosynthetic enzyme families, notably, NRPS, PKS, and cyanobactin enzymes (**Figure 5.1**).

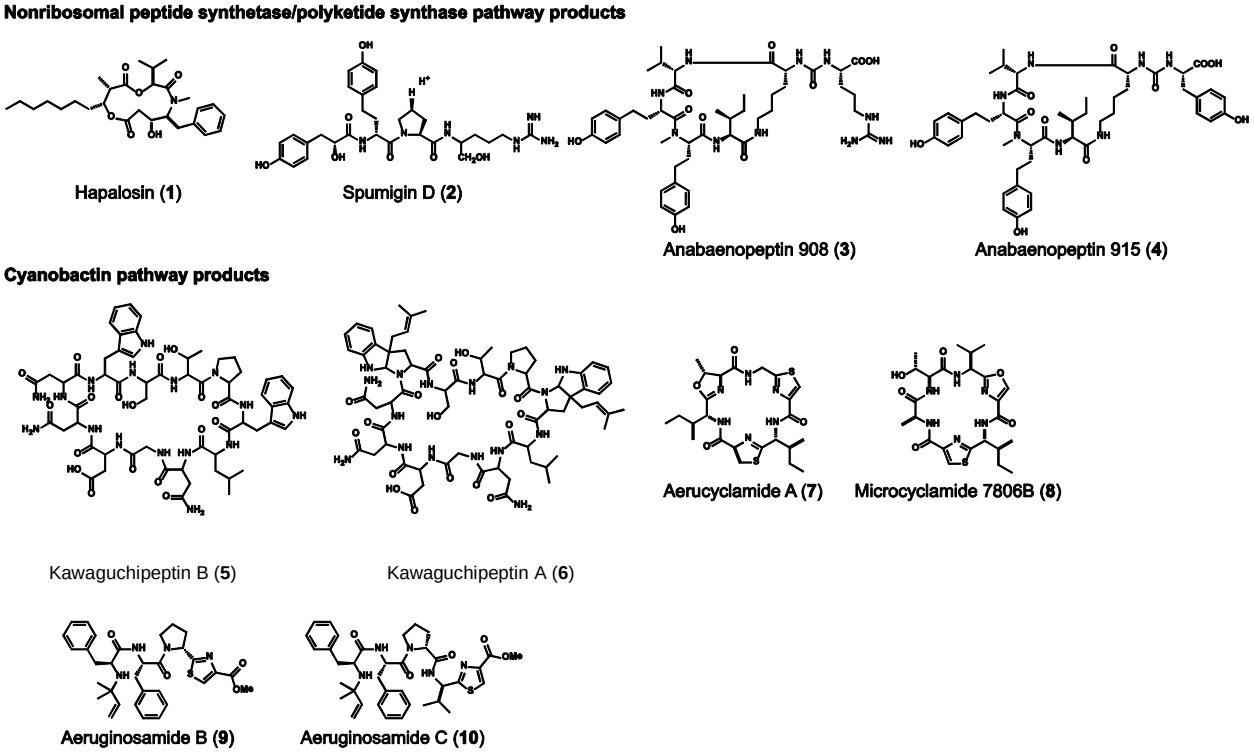


Figure 5.1 Cyanobacterial natural products and their biosynthetic machinery investigated in this study.

To illustrate the utility of these designer plasmids, the assembly of the hapalosin (*hal*) biosynthetic gene cluster is discussed here. The construction of the complete *hal* gene cluster incorporated both synthetic RBS sequences and a synthetic, high frequency, ATG start codon in place of the native, low frequency TTG start codon. The *hal* pathway was assembled in two sequential cloning steps to afford the *halDE* and *halBCDE* expression plasmids (**Figure 5.2a**). PCR primer designs include synthetic RBS or mutant codon sequences to yield amplicons with end sequences that were homologous with either an adjacent DNA fragment or with the backbone of the cloning plasmid. In **Figure 5.2b**, *halD* and *halE* PCR amplicons include synthetic end sequences that are homologous with each other or with the plasmid (indicated by corresponding color-coded regions of the gene inserts or the linear plasmid). The resulting plasmid simultaneously introduces a new RBS sequence upstream *halD* and mutates the native TTG start codon of *halD* to the ATG start codon.

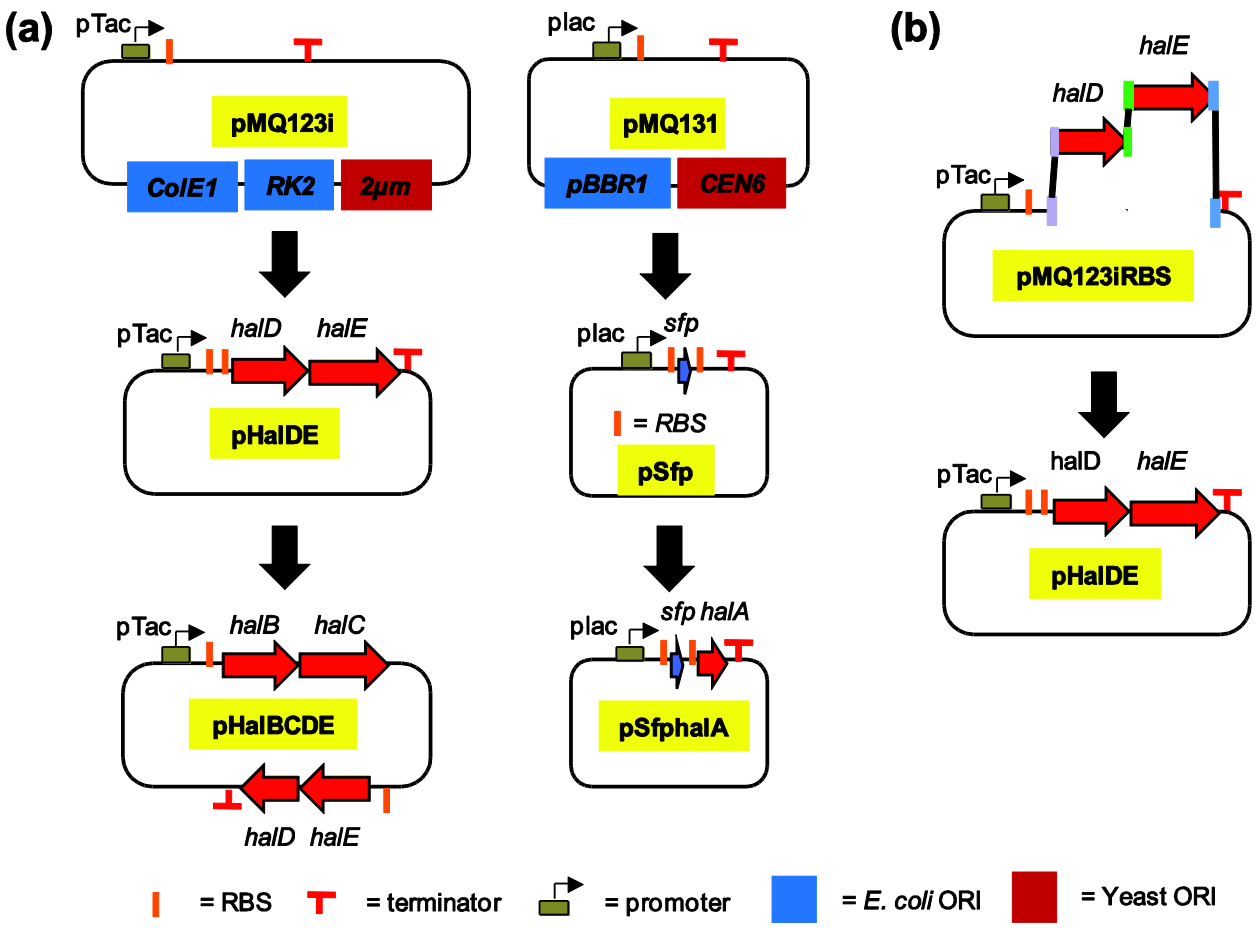


Figure 5.2 (a) Stepwise assembly of the entire hapalosin biosynthetic pathway was reconstituted using two separate yeast-*E. coli* binary plasmid systems. **(b)** Homologous sequences (colored in purple, green, and blue) were introduced via selective PCR primer design to facilitate yeast-assisted homologous recombination (recombineering) to afford the desired expression plasmid.

Cyanobacteria exhibit highly diverse morphological features, which contribute to their genetic intractability. Cyanobacteria are taxonomically classified into five subsection (I-V), based on similar bacteriological features [204, 205]. For example, species of subsections I and II are similar because all are unicellular, yet species of subsection I reproduce via binary fission or budding, while those of subsection II reproduce via multiple fission. Members of subsection III are filamentous, non-heterocystous cyanobacteria. Subsection IV and V species are all filamentous, heterocystous cyanobacteria. Heterocystous species utilize specialized cells to fix nitrogen in the absence of ammonia or nitrate [206]. Subsection IV species branch in one plane (false branching), while members of subsection V exhibit branching in multiple planes (true branching).

Table 5.1 Cyanobacterial biosynthetic pathways investigated in this study

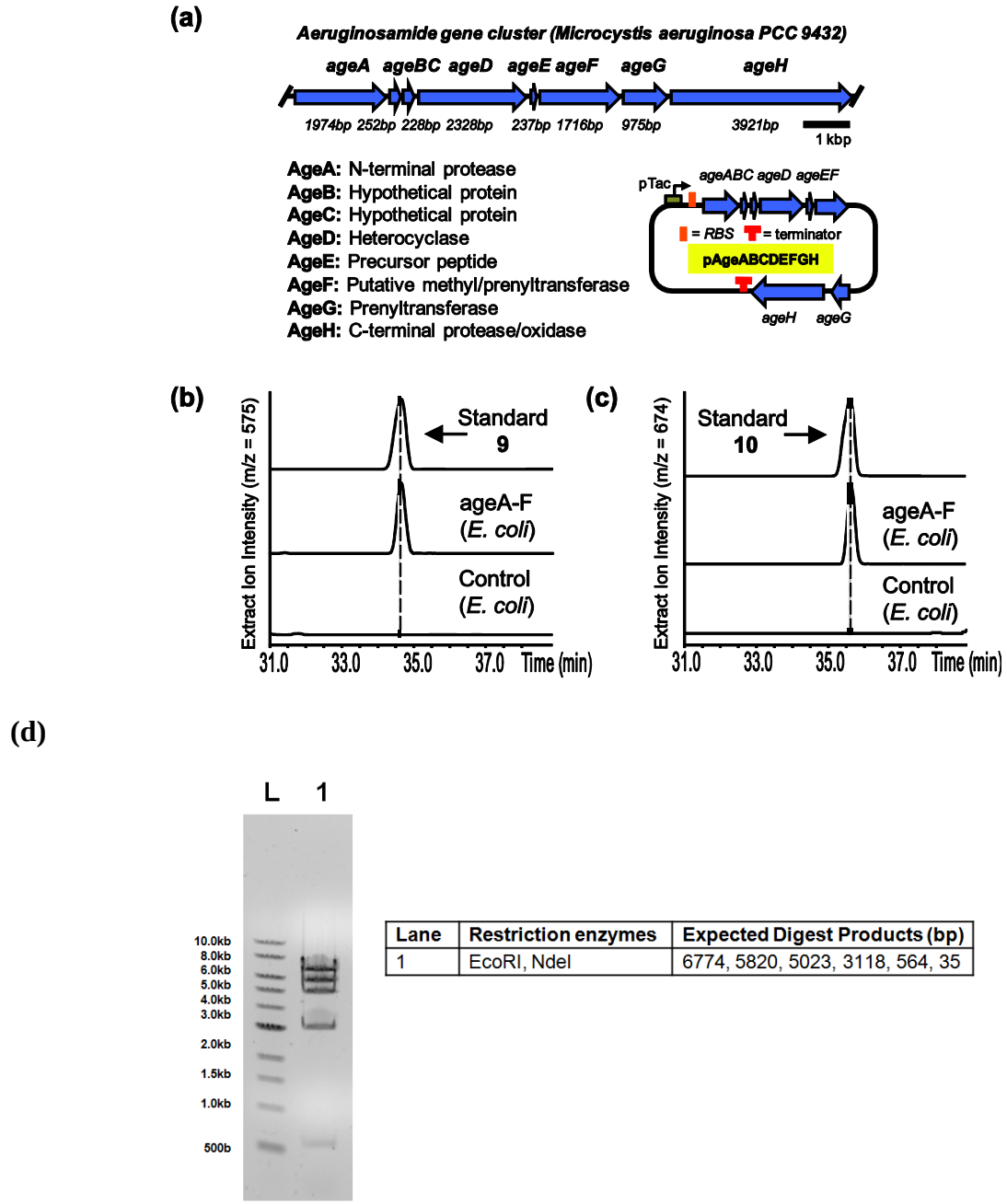
Cyanobacterial natural product	**Pathway Machinery**	**Cyanobacterial Species**	**Subsection**	**References**
Aeruginosamide	Cyanobactin	*Microcystis aeruginosa* strain PCC 9432	I	[177, 207]
Kawaguchipeptin	Cyanobactin	*Microcystis aeruginosa* NIES-88	I	[173, 208]
Microcyclamides	Cyanobactin	*Microcystis aeruginosa* strain PCC 7608	I	[197, 209]
Anabaenopeptin	NRPS	*Planktothrix agardhii* CYA 126/8	III	[210]
Spumigin	NRPS	*Nodularia spumigena* AV1	IV	[211]
Hapalosin	NRPS/PKS	*Hapalosiphon welwitschii* UTEX B1830	V	[212]

Given the diversity amongst the cyanobacteria subsections, an important consideration in the utility of a universal cyanobacterial expression platform is the scope of this system to clone and express biosynthetic pathways from diverse cyanobacterial sources. Therefore, biosynthetic pathways of members from all but one (subsection II) were investigated in this study (**Table**

5.1). In addition, pathways containing the major enzymes involved in cyanobacterial secondary metabolite biosynthesis—nonribosomal peptide synthetase (NRPS), polyketide synthase (PKS), and cyanobactin enzymes—were explored. The ability to clone and express diverse biosynthetic pathways from cyanobacterial members of various subsections demonstrates the widespread utility of this expression platform in the investigation of cyanobacterial natural products.

5.2 RAPID ASSEMBLY OF COMPLETE CYANOBACTERIAL GENE CLUSTERS

The elucidation of novel cyanobacterial natural products and/or the investigation of their biosynthetic pathways necessitate the rapid assembly of the entire gene cluster into a plasmid suitable for heterologous expression. To demonstrate the utility of the proposed cloning plasmid system, the aeruginosamide (*age*) gene cluster from *Microcystis aeruginosa* PCC7432 was assembled. Aeruginosamides are natural products with potent protease inhibitor activity [207], and the recent identification of the *age* gene cluster illuminated the aeruginosamides as the products of a cyanobactin assembly line [177]. This is an unexpected discovery, as most cyanobactin products are cyclic, with the aeruginosamides as being only one of two (the other being the viridisamides) known families of linear cyanobactins.



Lane	Restriction enzymes	Expected Digest Products (bp)
1	EcoRI, NdeI	6774, 5820, 5023, 3118, 564, 35

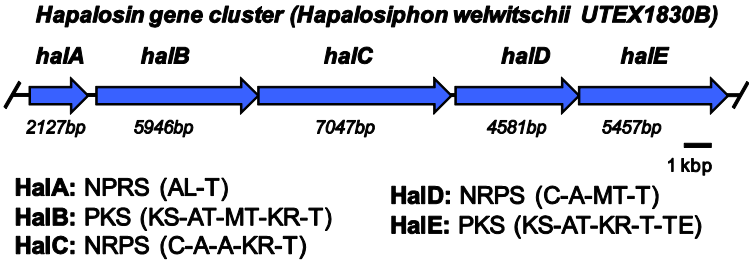
Figure 5.3 (a) Aeruginosamide biosynthetic pathway organization from *M. aeruginosa* PCC9432. **(b)** Heterologous expression of the complete *age* pathway afforded the production of aeruginosamide B. **(c)** Heterologous expression of the complete *age* pathway afforded the production of aeruginosamide C. **(d)** Restriction enzyme digest analysis of the complete *age* pathway to confirm plasmid assembly.

The entire *age* gene cluster was assembled in a single cloning step consisting of four ~3kb inserts to afford the complete age pathway (**Figure 5.3a**). The successful assembly of the *age* pathway expression plasmid was verified by restriction enzyme digest analysis (**Figure 5.3d**). The aeruginosamides **9** and are both highly modified, including heterocyclization of the cysteine residue to afford a thiazole in addition to N-terminus prenylation and C-terminus methylation. Prenylation requires the substrate dimethylallyl pyrophosphate (DMAPP); however, *E. coli* cannot produce DMAPP endogenously. Martin and colleagues constructed a mevalonate (MEV)-dependent pathway (plasmid pMBI) that can synthesize DMAPP in the presence of the precursor MEV for heterologous expression in *E. coli* [213]. Co-expression of the *age* gene cluster with pMBI afforded the production of both **9** and **10**, which were exhibited accurate mass values and retention times consistent with the aeruginosamides observed from a crude extract of the native producer *M. aeruginosa* PCC9432 (**Figure 5.3b-c**).

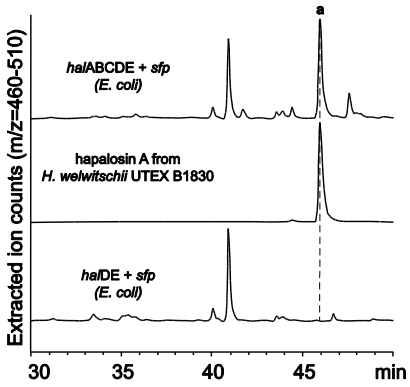
5.2.1 Characterization of the Hapalosin Biosynthetic Gene Cluster

Rapidly correlating uncharacterized gene clusters to their biosynthetic products is a central objective of expression systems. To demonstrate the ability of this platform to successfully construct and express biosynthetic pathways, a large NRPS biosynthetic pathway was investigated: the hapalosin (*hal*) pathway, which is approximately 25.2 kbp in length. Hapalosin is produced by *H. welwitschii* UTEX B1830 and is a hybrid NRP-PK lipo-depsipeptide that exerts multidrug resistance (MDR) reversing activity [212]. Five genes (*halABCDE*) are involved in the assembly of hapalosin (**Figure 5.4a**).

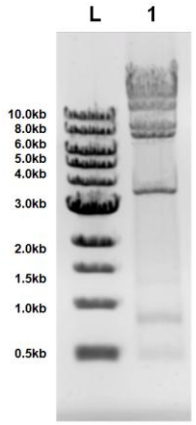
(a)



(b)



(c)



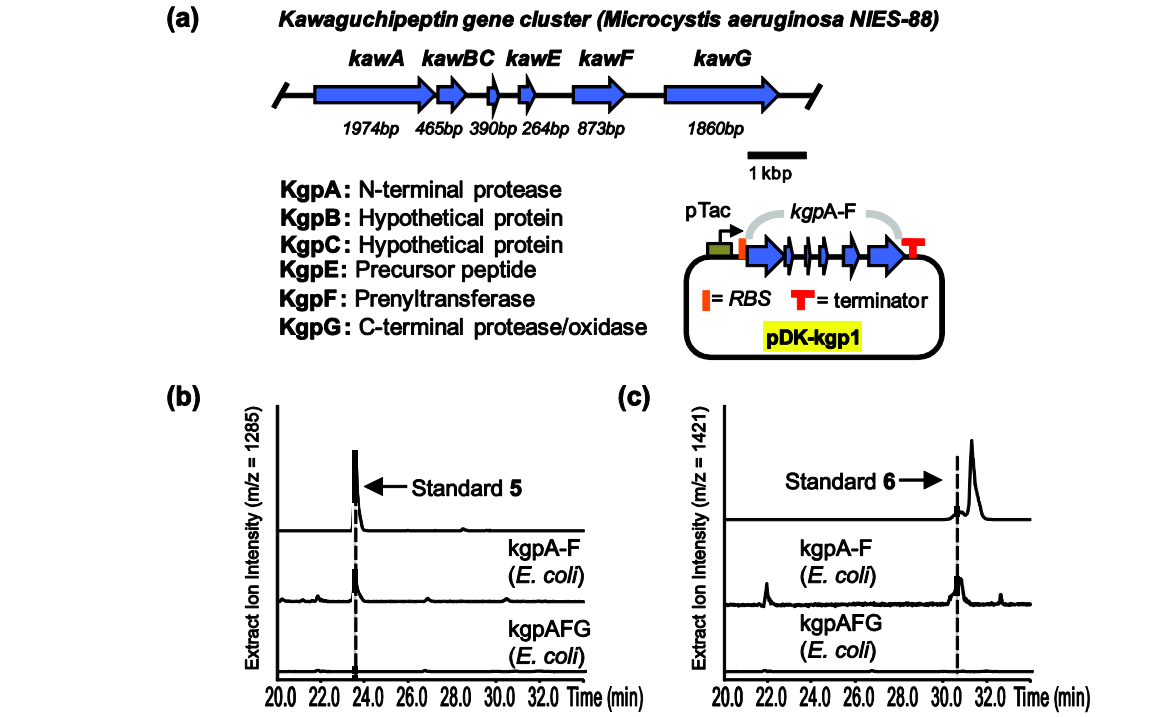
Lane	Restriction enzymes	Expected Digest Products (bp)
1	BglII	11426, 8299, 7286, 3671, 837, 479

Figure 5.4 (a) Hapalosin biosynthetic pathway organization from *H. welwitschii* UTEX1830B. **(b)** Heterologous expression of the complete pathway (*hal*ABCDE) afforded the production of hapalosin (peak a). **(c)** Restriction enzyme digest analysis of the complete *hal* pathway to confirm plasmid assembly.

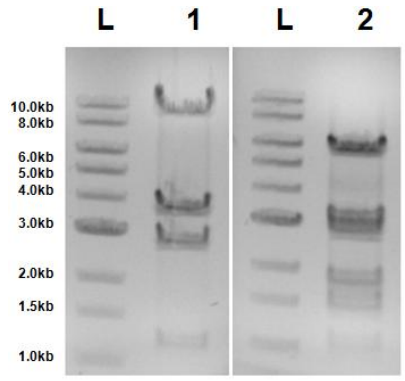
The complete *hal* pathway was reconstituted on two separate plasmids: pMQ131 contains the *halA* and *sfp* genes, and pMQ123i contains the *halBCDE* genes (**Figure 5.2**). The *sfp* gene is a phosphopantetheinyl transferase that is responsible for the essential posttranslational modification of acyl and peptidyl carrier proteins. The successful construction of the *hal* pathway expression plasmid was verified by restriction enzyme digest analysis (**Figure 5.4c**). Expression of the *halABCDE* pathway in *E. coli* afforded the production of hapalosin, which was verified by retention time and accurate mass with hapalosin produced from the native producer *H. welwitschii* UTEX B1830 (**Figure 5.4b**). Expectedly, the expression of the partial pathway construct (*halDE*) afforded no observable hapalosin.

5.2.2 Characterization of the Kawaguchipeptin Biosynthetic Gene Cluster

The kawaguchipeptins are cyclic undecapeptides that are produced via cyanobactin machinery (encoded by the *kgpABCDEGF* gene cluster, **Figure 5.5a**). They were first isolated from *M. aeruginosa* NIES-88, and while the non-prenylated kawaguchipeptin (kawaguchipeptin B) exhibited antibacterial activity against *Staphylococcus aureus*, interestingly, the diprenylated analog (kawaguchipeptin A) does not exhibit such bioactivity [173, 208]. Assembly of the *kgp* pathway was verified by restriction enzyme digest analysis (**Figure 5.5d**).



(d)



Lane	Restriction enzymes	Expected Digest Products (bp)
1	EcoRI, BglII	8873, 3621, 2673, 1095
2	EcoRV, PvuII	5858, 3104, 2819, 1785, 1442, 1012, 149, 93

Figure 5.5 (a) Kawaguchipeptin biosynthetic pathway organization from *M. aeruginosa* NIES-88. **(b)** Heterologous expression of the complete *kgp* operon afforded the production of kawaguchipeptin A. **(c)** Heterologous expression of the complete kgp operon afforded the production of kawaguchipeptin B. Plasmid pMBI was co-expressed with both the complete and partial *kgp* pathways. Standards **5** and **6** correspond to kawaguchipeptins B and A, respectively, obtained from a crude extract of *M. aeruginosa* NIES-88. **(d)** Restriction enzyme digest analysis of the complete *kgp* pathway to confirm plasmid assembly.

Co-expression of the *kgp* operon with pMBI yielded both kawaguchipeptins A and B (**Figure 5.5b and 5.5c**). As expected, the expression of the partial *kgp* operon (*kgpAGF*) did not produce either kawaguchipeptin. Interestingly, the *M. aeruginosa* NIES-88 native producer of the kawaguchipeptins revealed the production of two putative isomers of kawaguchipeptin A (**Figure 5.5c**). However, heterologous expression of the complete *kgp* operon afforded the production of a single peak corresponding to kawaguchipeptin B. It is possible that the heterologous expression of the *kgp* operon preferentially produces the minor conformational isomer of the macrocycle.

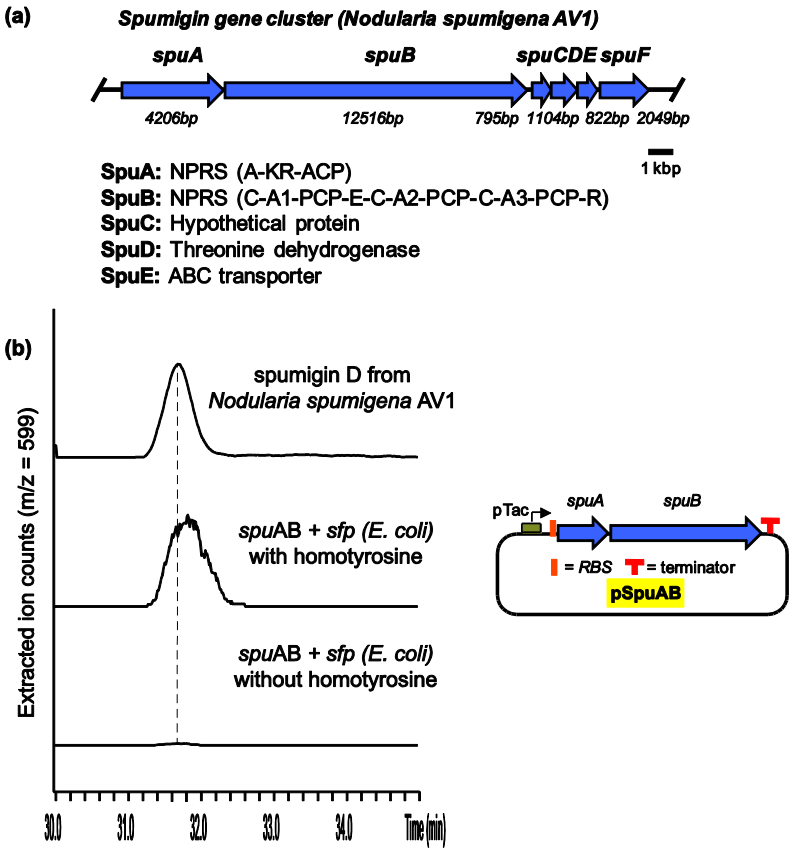
5.3 ESTABLISHING GENES ESSENTIAL TO BIOSYNTHETIC PATHWAYS

Delineating the pathway for natural products biosynthesis enables insight into the enzymatic function of the genes involved and affords opportunities to rationally engineer genetic elements to enhance the production of these molecules. Traditional approaches rely on the generation of biosynthetic pathway mutants and the analysis of product or intermediate accumulation. This approach is currently precluded by the morphological complexity of cyanobacteria and the lack of available genetic tools for manipulation. The designer plasmids, on the other hand, can circumvent these hurdles to understanding the essential genes required for the biosynthesis of cyanobacterial products. The utility of this expression system in defining minimal biosynthetic pathways is demonstrated in the following two examples. First, the essential genes of the spumigin (*spu*) pathway were explored through selective cloning of the *spu* pathway. Second, the essential genes of the microcyclamide (*mca*) biosynthetic pathway were explored through the

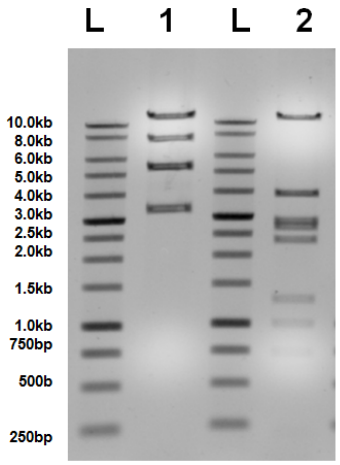
systematic generation of mutant pathway plasmid constructs and their heterologous expression in *E. coli*.

5.3.1 Selective cloning of the Spumigin pathway provides insight into essential biosynthetic genes

Spumigins are protease inhibitors produced from *N. spumigena*, which are filamentous nitrogen-fixing cyanobacterium that form toxic blooms in brackish waters worldwide [211]. The spumigin gene cluster from *N. spumigena* AV1 consists of *spuABCDE* (**Figure 5.6a**), and it is predicted that only *spuAB* (NRPS modules) are essential in the biosynthesis of spumigins. The designer plasmid platform was utilized to construct *spuAB* (plasmid pSpuAB in **Figure 5.6b**), which was subsequently co-expressed with *sfp*. The construction of the spuAB plasmid was verified by restriction enzyme digest analysis (**Figure 5.6c**). Only in the presence of homo-L-tyrosine was spumigin D produced via heterologous expression of *spuAB* in *E. coli*, as determined by retention time and accurate mass comparison to an extract of the native producer *N. spumigena* AV1 (**Figure 5.6b**). This finding suggests that *spuAB* are essential in the biosynthesis of the spumigins and that SpuC (hypothetical protein), and SpuD (threonine dehydrogenase), and SpuE (reductase) serve auxiliary and non-essential roles in spumigin biosynthesis.



(c)

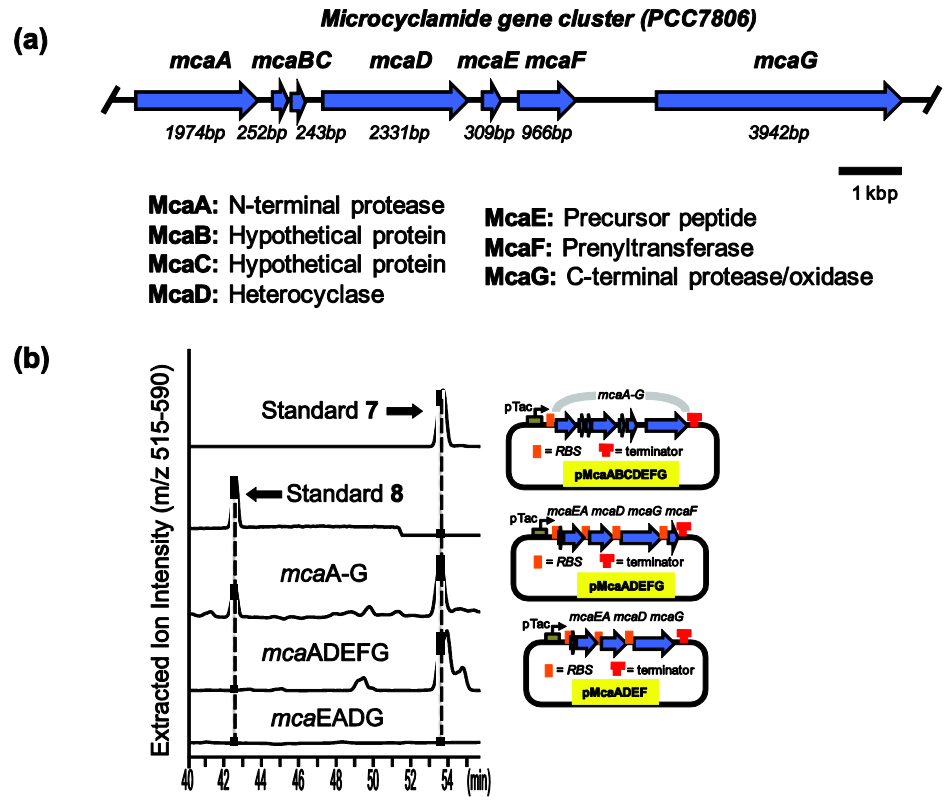


Lane	Restriction enzymes	Expected Digest Products (bp)
1	NdeI, SpeI	10365, 7094, 5058, 3118
2	EcoRV	10938, 3833, 2819, 2631, 2290, 1254, 970, 705, 195

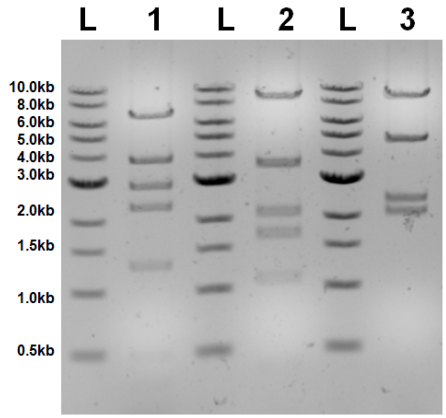
Figure 5.6 (a) Spumigin biosynthetic pathway organization from *N. spumigena* AV1. **(b)** Heterologous expression of *spu*AB and *sfp* afforded the production of spumigin D only in the presence of L-homotyrosine. **(c)** Restriction enzyme digest analysis of the *spu*AB pathway to confirm plasmid assembly.

5.3.2 The McaF of the Microcyclamide Biosynthetic Pathway is Essential

The microcyclamides are a group of related cyclic hexapeptides from the cyanobacterium *M. aeruginosa* and have shown to be cytotoxic [209]. The biosynthetic pathway for the microcyclamide (*mca*) pathway was characterized previously. [197] The native *mca* operon from *Microcystis aeruginosa* strain PCC 7806 was cloned and assembled to afford *mcaABCDEFG*. The proposed functions of the various genes of the *mca* pathway are described in **Figure 5.7a**. Additionally, the *mca* operon was re-organized to yield reduced pathways *mcaADEFG* and *mcaADEG*. The *mcaBC* genes are predicted to encode hypothetical proteins but have been previously determined to be non-essential to the biosynthesis of cyanobactins [197]. The assembly of the complete *mca* biosynthetic pathway was verified by restriction enzyme digest analysis (**Figure 5.7c**). Expression of *mcaADEG* resulted in no detection via MS of any microcyclamides production, compared to the *mcaABCDEFG* pathway (**Figure 5.7b**). However, re-introduction of *mcaF* to *mcaADEG* to yield *mcaADEFG* restored microcyclamide 7806B production. McaF is a predicted prenyltransferase, and although the microcyclamides are not modified with prenyl groups, this observation suggests that McaF is still essential to the biosynthesis of microcyclamides, an observation noted previously by Schmidt *et al.* in the biosynthesis of patellamide [175].



(c)

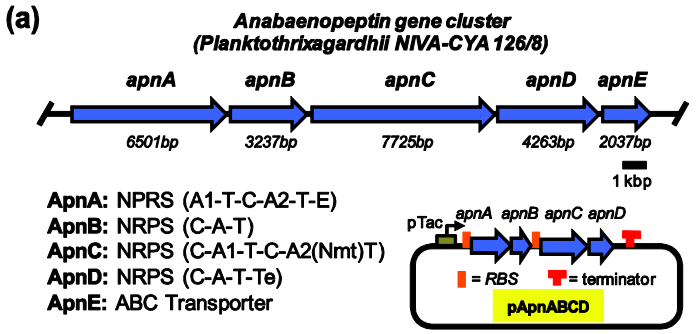


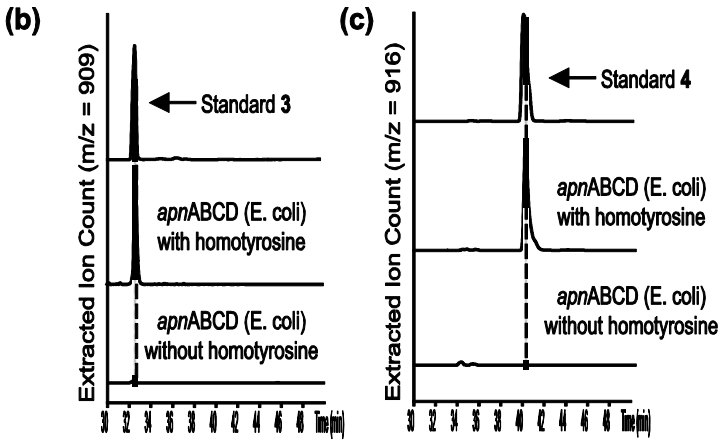
Lane	Restriction enzymes	Expected Digest Products (bp)
1	EcoRV	6647, 3724, 2819, 2246, 467, 306
2	PvuII, NdeI	8653, 3558, 2058, 1659, 1060, 382, 93
3	XbaI, HindIII	8517, 4599, 2311, 2036

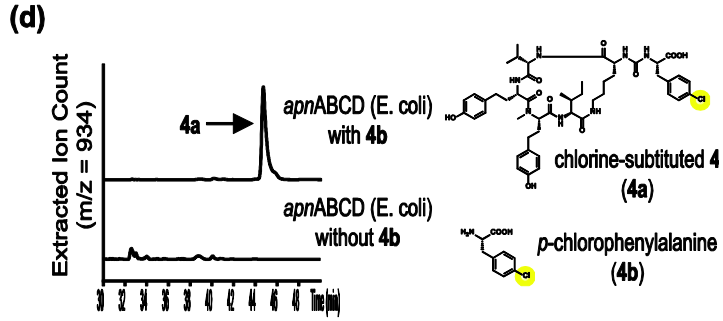
Figure 5.7 (a) Microcyclamide biosynthetic pathway organization from *M. aeruginosa* PCC7806. **(b)** Heterologous expression of either the complete or partial (*mca*ADEFG) operon afforded the production of microcyclamides. Standards 7 and 8 correspond aerucyclamide and microcyclamide compounds observed from a crude extract of *M. aeruginosa* PCC7806. **(c)** Restriction enzyme digest analysis of the complete *mca* pathway to confirm plasmid assembly.

5.3.3 Designer plasmids and their heterologous expression can be used to generate analogs

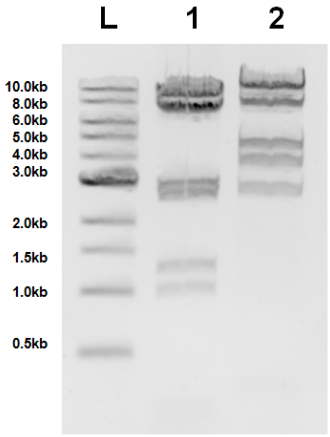
A desirable feature of heterologous expression systems is the ability to generate novel analogs with enhanced bioactivity or with functional groups that can serve as chemical handles for subsequent modification. Anabaenopeptins are a group of diverse cyclic peptides that have demonstrate protease inhibition and are produced from several different cyanobacterial species, including those from *Anabaena*, *Planktothrix*, *Microcystis* and *Nodularia* [210, 214]. The anabaenopeptin (*apn*) gene cluster from *P. agardhii* CYA 126/8 was reconstituted as *apnABCD* (**Figure 5.8a**). Because *apnE* encodes for an ABC transporter and is not directly involved in the biosynthesis of the anabaenopeptins, these were excluded from the biosynthetic gene cluster construct. The assembly of the *apnABCD* pathway was verified by restriction enzyme digest analysis (**Figure 5.8e**). As expected, the co-expression of *apn*ABCD and *sfp* in the presence of homo-L-tyrosine afforded the production of **3** and **4**, as verified by both accurate mass and retention time, compared to the anabaenopeptins observed from a crude extract of *P. agardhii* NIVA-CYA 126/8 (**Figure 5.8b-c**).







(e)



Lane	Restriction enzymes	Expected Digest Products (bp)
1	BglII, EcoRV	8612, 7180, 6870, 2901, 2623, 1254, 998, 196
2	NcoI, NdeI	10911, 8033, 4697, 3875, 2824, 294

Figure 5.8 (a) Anabaenopeptin biosynthetic pathway organization from *P. agardhii* NIVA-CYA 126/8. **(b)** Only heterologous expression of the complete *apn* operon in the presence of homo-L-tyrosine afforded the production of 3. **(c)** Only heterologous expression of the complete *apn* operon in the presence of L-homotyrosine afforded the production of 4. **(d)** Supplementation of the *E. coli* culture with 4b afforded the production of the chlorinated analog of **3** (**4a**). Standards 3 and 4 correspond to anabaenopeptins from crude extract of *P. agardhii* NIVA-CYA 126/8. **(e)** Restriction enzyme digest analysis of the complete *apn* pathway to confirm plasmid assembly.

The incorporation of halogenated substrates in the biosynthesis of **4** from *P. agardhii* CYA 126/8 was investigated. Compound **4** contains an exocyclic tyrosine residue that can potentially be substituted with a halogenated phenylalanine residue. The incorporation of unnatural substrates into the **4** structure was investigated by supplementing the culture media with *p*-chlorophenylalanine or *p*-fluorophenylalanine in addition to homo-L-tyrosine, which is a natural substrate in the biosynthesis of native anabaenopeptins. LCMS analysis of the supernatant extract revealed that compounds corresponding to chlorinated and fluorinated analogs of **4** were observed from the expression culture, only when supplemented with their respective halogenated phenylalanine substrate. The observation of a chlorinated **4** (**4a**) only in the presence of *p*-chlorophenylalanine is shown in **Figure 5.8d**. The ratios (based on ion count) of halogenated **4** to native **4** was determined to be 1:100 and 1:10 for the fluorinated and chlorinated analogs, respectively. This indicates that tyrosine is the preferred substrate for the ApnA module and that chlorinated phenylalanine, compared to fluorinated phenylalanine, is more readily incorporated into the pathway. The electronegativity of the atom in the *para* position of the phenyl ring of tyrosine or phenylalanine can explain the ease of incorporation of these various substrates, whereby the least electronegative (oxygen) is most preferred, followed by a moderately electronegative atom (chlorine), and then the most electronegative (fluorine).

5.4 DISCUSSIONS AND CONCLUSIONS

Cyanobacteria have been known to be a promising source of natural products with diverse biological activities, and recent genome sequencing has continued to suggest that these

microorganisms are prolific producers of yet characterized secondary metabolites that may be useful in the development of new medicines [200]. Despite the prevalence of biosynthetic gene clusters for compounds yet to be discovered, few methods exist to access these compounds or to investigate their biosynthesis. This study establishes a new designer plasmid platform that utilizes yeast recombineering technology to successfully clone and then heterologously express diverse types of biosynthetic pathways from cyanobacterial members from four of the five subsections of this genera (**Table 5.1**).

This platform was successfully applied to: 1) rapidly assemble entire biosynthetic pathways in a single cloning step (aeruginosamide); 2) correlate cryptic biosynthetic pathways to their product (hapalosin, kawaguchipeptin); 3) establish the genes essential for biosynthesis (spumigin, microcyclamides); and 4) generate unnatural analogs that may serve as precursors for further synthetic modifications or that may exhibit enhanced bioactivity (anabaenopeptins).

In one example from this study, the systematic expression of genes from the microcyclamide (*mca*) pathway defined the essential (minimal) pathway required for the biosynthesis of microcyclamides. Deletion of the PatF-like prenyltransferase of the *mca* operon (McaF) resulted in no detectable production of microcyclamide, while re-introduction of McaF to the pathway restored microcyclamide production (**Figure 5.7b**). This is consistent with previous studies that observed the essential role of PatF in the complete maturation of the patellamides [175]. Rapid access to complete and partial pathways of the *mca* biosynthetic pathway for heterologous expression investigations is expected to facilitate understanding of the

role a predicted prenyltransferase in the maturation of natural products lacking prenyl modifications.

Another notable feature of the proposed cyanobacterial biosynthetic pathway cloning and expression system is the ability to generate unnatural and novel analogs. This was demonstrated by the biosynthesis of halogenated analogs of anabaenopeptin via substrate-guided biosynthesis (i.e., the introduction of defined unnatural substrates to direct the biosynthesis of novel natural products.) The facile access to unnatural analogs of cyanobacterial natural products will enable the investigation and development of compounds with enhanced and novel biological activities.

There exists a lack of genetic tools to access the promising reservoir of cyanobacterial natural products. This study has established a new designer plasmid founded on a previously reported yeast-recombineering based plasmid [203] that can be used to rapidly assemble diverse biosynthetic gene clusters from cyanobacteria from members of several subsections of cyanobacterial genera. A limitation of this platform is the ability to produce appreciable quantities of product that can be useful for characterization, particularly of novel natural products. Possible solutions to this challenge include optimization strategies that alter transcriptional and translation regulatory elements in addition to methods to enhance substrate availabilities and heterologous host suitability. Nonetheless, this platform represents a tool that can accelerate the discovery and development of cyanobacterial natural products.

6.0 HAPALOSIN BIOSYNTHESIS IN HAPALOSIPHON WELWITSCHII AND ITS COMBINATORIAL GENERATION IN ESCHERICHIA COLI

Secondary metabolites from cyanobacteria of the order *Stigonematales* have demonstrated promising anticancer, anti-inflammatory, antibiotic, and anti-fungal activities. Prominent examples of natural products from *Stigonematales* include the hapalindoles [215], fischerellins [216], hapalosin [212], and ambigols [217] (**Figure 6.1**). Although their biological activities are well-recognized, little is known about their biosyntheses. Investigating their biosynthetic pathways can enable access to these structurally complex compounds and also provide insight into new chemical transformations [218]. Recent characterizations of sequenced genomes from species of the order Stigonematales have shown that a significant portion of their genomes are dedicated to polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) genes. In one study, it was shown that approximately 5% of the *Fischerella* sp. genome was found to consist of PKS and NRPS genes.[171] Therefore, there has been a growing interest in sequencing the genomes of cyanobacteria to uncover the gene clusters responsible for the production of cyanobacterial natural products. Our research group has recently sequenced the genome of three cyanobacterial strains in the order of Stigonematales, including *Hapalosiphon welwitschii* UTEX1830, *Fischerella muscicola* UTEX1829, and *Fischerella ambigua* UTEX1903, in an effort to investigate the biosyntheses of their natural products. Genome mining of these species has elucidated the biosynthetic machinery of valuable natural products, including those of the

hapalindole- and ambiguine-type indole alkaloids [219, 220], the multidrug resistance (MDR) reversing agent hapalosin [219], as well as a number of other cryptic NRPS and/or PKS biosynthetic pathways.

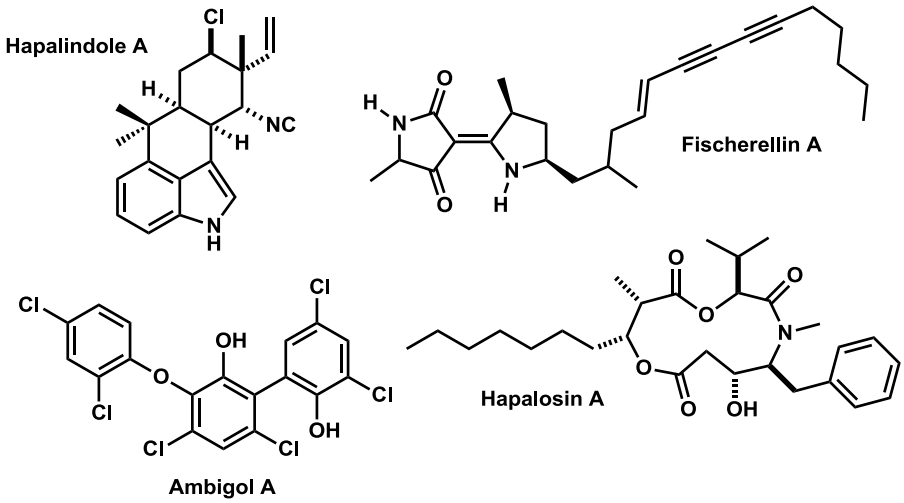


Figure 6.1 Natural products discovered from species of the cyanobacterial order *Stigonematales*.

A central challenge in the study of cyanobacterial natural products biosynthesis arises from the lack of tools to genetically manipulate these organisms. The extracellular mucilaginous sheaths that encase cyanobacterial cells prohibit the introduction of exogenous DNA [192], which makes the study of defined cyanobacterial mutants of biosynthetic pathways difficult to accomplish. Furthermore, genetic regulatory elements, such as promoters, ribosome binding sites, and terminators, have not been well-characterized in cyanobacteria. Other research groups have been able to investigate cyanobacterial biosynthetic pathways by cloning and heterologously expressing them, such as in the case for the cyanobactins patellin A [185] and

trunkamide [221]. However, such approaches rely on ligation-based cloning plasmids and can require multiple cloning steps to assemble complete biosynthetic pathways.

To address the shortcoming of current approaches in understanding the enzymatic machinery involved in the biosynthesis of cyanobacterial natural products, we developed a designer plasmid system that enables the rapid assembly of entire biosynthetic pathways and their expression in the heterologous host *E. coli*. The designer plasmid is a yeast-*E. coli* shuttle vector that enables the concomitant cloning and genetic engineering of large sequences of DNA via yeast-assisted homologous recombination. We utilized this system to discover the biosynthetic pathway for the MDR-reversing agent hapalosin from *H. welwitschii*. Expression of the hapalosin biosynthetic pathway in *E. coli* revealed that this hybrid NRPS-PKS assembly line exhibits high substrate promiscuity, which affords the generation of a combinatorial library of hapalosin analogs.

6.1 IDENTIFICATION OF PUTATIVE HAPALOSIN BIOSYNTHETIC GENE CLUSTER IN HAPALOSIPHON WELWITSCHII BY DE NOVO GENOME SEQUENCING

Hapalosin is a 12-membered cyclic depsi-lipopeptide that was originally discovered from *H. welwitschii* UTEX1830B by Moore and coworkers[212]. The unique 12-membered ring, mixed lactone and lactam scaffold suggests hapalosin is most likely assembled by a mixed enzymatic machinery operated by both PKS and NRPS. However, the sequence of monomer assembly to

form the final structure of hapalosin is unclear. Genome mining of *H. welwitschii* UTEX1830, revealed a 26 kbp-long gene cluster that contains five PKS-NRPS genes (**Figure 6.2**). Domain predictions of the NRPS and PKS modules of this gene cluster suggested that it is responsible for the assembly of hapalosin and has therefore been denoted the *hal* operon.

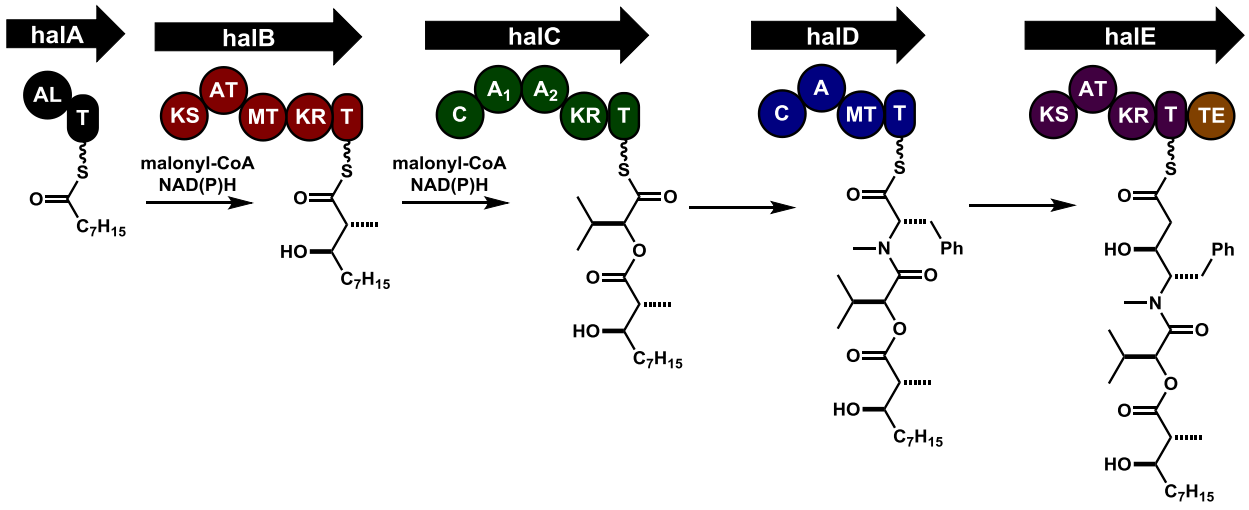


Figure 6.2 Organization of the hapalosin gene cluster and proposed gene functions and PKS/NRPS domain predications. Genes and their respective substrates of the full structure are correlated by color. Abbreviations: AL: fatty acid ligase; T: thiolation domain; KS: ketosynthase domain; AT: acyltransferase domain; MT: methyltransferase domain; KR: ketoreductase domain; T: thiolation domain; C: condensation domain; A: adenylation domain; TE: thioesterase domain.

The first gene *halA* is a didomain NRPS that contains a fatty acid ligase and was predicted to incorporate a C8-fatty acid into the structure of the final product. Next, *halB*, a pentadomain PKS, was predicted to include the ability to incorporate a C2 ketide unit via Claisen

condensation, followed by either C- or O-methylation by a methyltransferase. *halC* also consisted of a pentadomain NRPS whereby the second adenylation domain was predicted to have the capacity to incorporate a ketoacid [222], while the ketoreductase may subsequently reduce the ketone to an α-hydroxy acid, analogous to that which has been discovered in the biosynthesis of cereulide [223]. *halD* is a tetradomain NRPS where the A domain is predicated to be able to select phenylalanine as well as carry out methylation via a methyltransferase domain. Lastly, *halE* was predicted to be a pentadomain PKS with a similar capacity as *halB* to incorporate a C2 ketide unit via Claisen condensation. The C-terminus thioesterase domain of *halE* was presumed to catalyze the final macrolactonization event.

6.2 ASSEMBLY OF THE HAPALOSIN GENE CLUSTER USING A DESIGNER PLASMID

To establish the biosynthetic function of the hapalosin gene cluster from *Hapalosiphon welwitschii* UTEX1830, the DNA sequences of interest from the hapalosin locus were cloned and used to construct the expression plasmids via yeast-assisted homologous recombination (yeast recombineering). [74],[224] Both plasmids were constructed to incorporate synthetic translational and transcriptional regulatory sequences, including ribosome binding sites (RBS) and the isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible Lac (plasmids pHal3 and 4) or Tac (plasmid pHal2) promoters (**Figure 6.3**). Additionally, the low frequency TTG start codon of *halD* was mutated to the high frequency ATG start codon of *E.* coli to further enhance translational efficiency. The ATG start codon was introduced via site-directed mutagenesis of the cloning primer for this gene. The sequential plasmid construction to assemble the *halBCDE*

locus enabled the incorporation of the synthetic RBS as well as the ATG codon. The construction of expression plasmids pHal2 and pHal4 were verified by restriction enzyme DNA digestion analyses. The observed DNA fragments were consistent with the expected DNA products for each respective restriction enzyme digestion analysis.

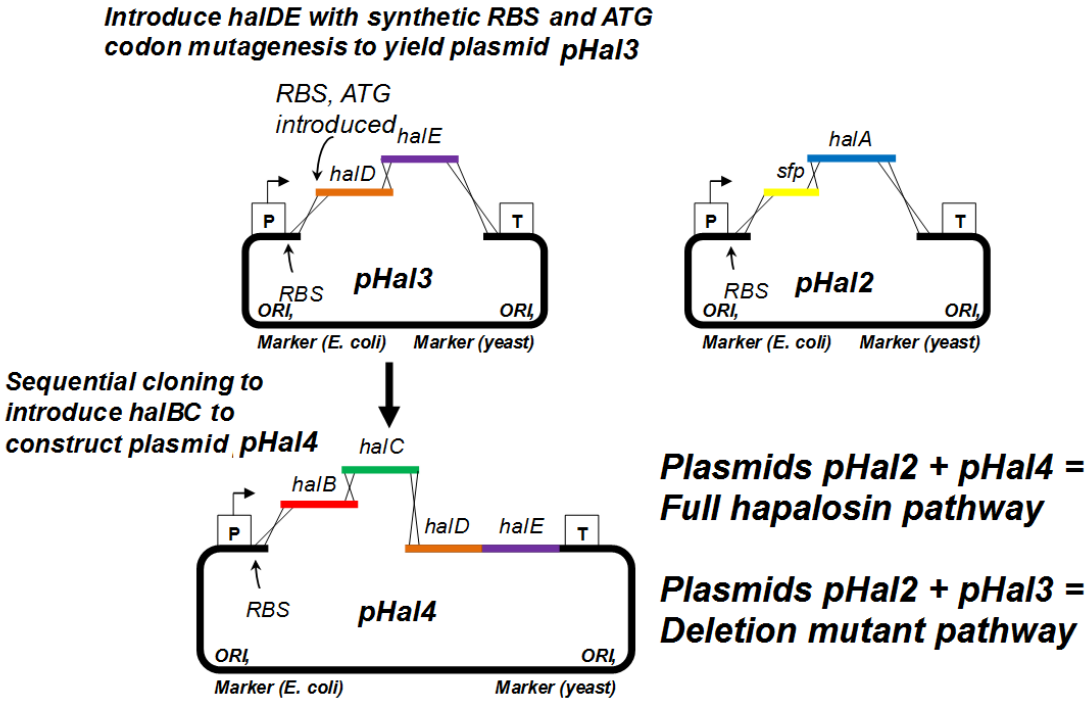


Figure 6.3 Construction of hapalosin expression plasmids pHal2 – 4. Sequential cloning enabled the introduction of synthetic translational regulatory element RBS sequences and the site-directed mutagenesis of the native low frequency TTG codon to the high frequency ATG codon. P: Lac (pHal3 and 4) or Tac (pHal2) promoter sequences; T: terminator; ORI: origin of replication; Marker: selectable marker.

6.3 HETEROLOGOUS EXPRESSION OF HAPALOSIN GENE CLUSTER AND ITS DELETION MUTANT IN HETEROLOGOUS EXPRESSION IN E. COLI

The heterologous expression of the *hal* gene cluster was assessed in *E. coli*, and the culture supernatant was analyzed for metabolite production. The biosynthesis of hapalosin A from *E. coli* was compared to an extract from the native producer *Hapalosiphon welwitschii* UTEX1830 (**Figure 6.4**). A molecule from the heterologous expression of the entire *hal* pathway in *E. coli* eluted at the same retention time and nominal mass as that of hapalosin A obtained from *H. welwitschii*. The production of hapalosin A was verified by isolating the compound and by comparative ^{1}H-NMR and high resolution MS analysis. Furthermore, as expected, the *halDE* partial pathway expression culture did not yield production of hapalosin A. Peaks corresponding to hapalosin A from *E. coli* and *Hapalosiphon welwitschii* are labeled **A** in **Figure 6.4**.

Three other prominent peaks were also discovered in the E. coli expression culture of the full *hal* pathway. These compounds are labeled **B**,**C** and **D** in **Figure 6.4** and their nominal masses were determined to be 504, 462, and 486 Da, respectively, whereas the nominal mass of hapalosin A is 490 Da. The differences in nominal masses suggested the possibility of hapalosin analogs varying only in $–CH_2–$ (14 Da) and $–C_2H_4–$ (28 Da) groups. HR-MS^2 analysis verified the presence of hapalosin analogs corresponding to hapalosins B, C and D. The HR-MS^2 fingerprint analysis for hapalosin A is shown in **Figure 6.5**. Hapalosin B had only previously been reported to be produced naturally from the cyanobacteria strain *Nostoc sp*. 4A [225]; therefore, this finding represents the first observation of hapalosin B production from *Hapalosiphon welwitschii* UTEX1830. The structures of hapalosins B – D are shown in **Figure 6.6**.

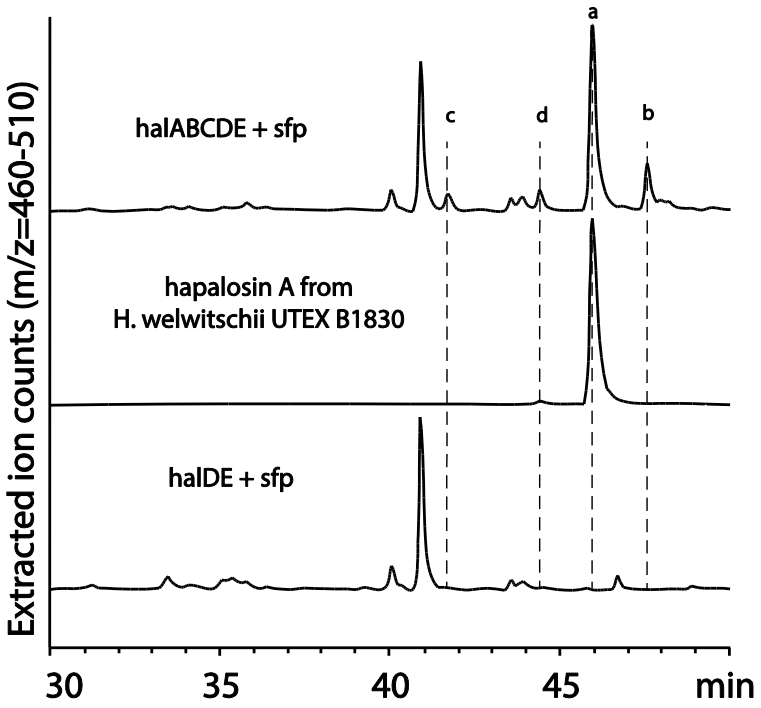
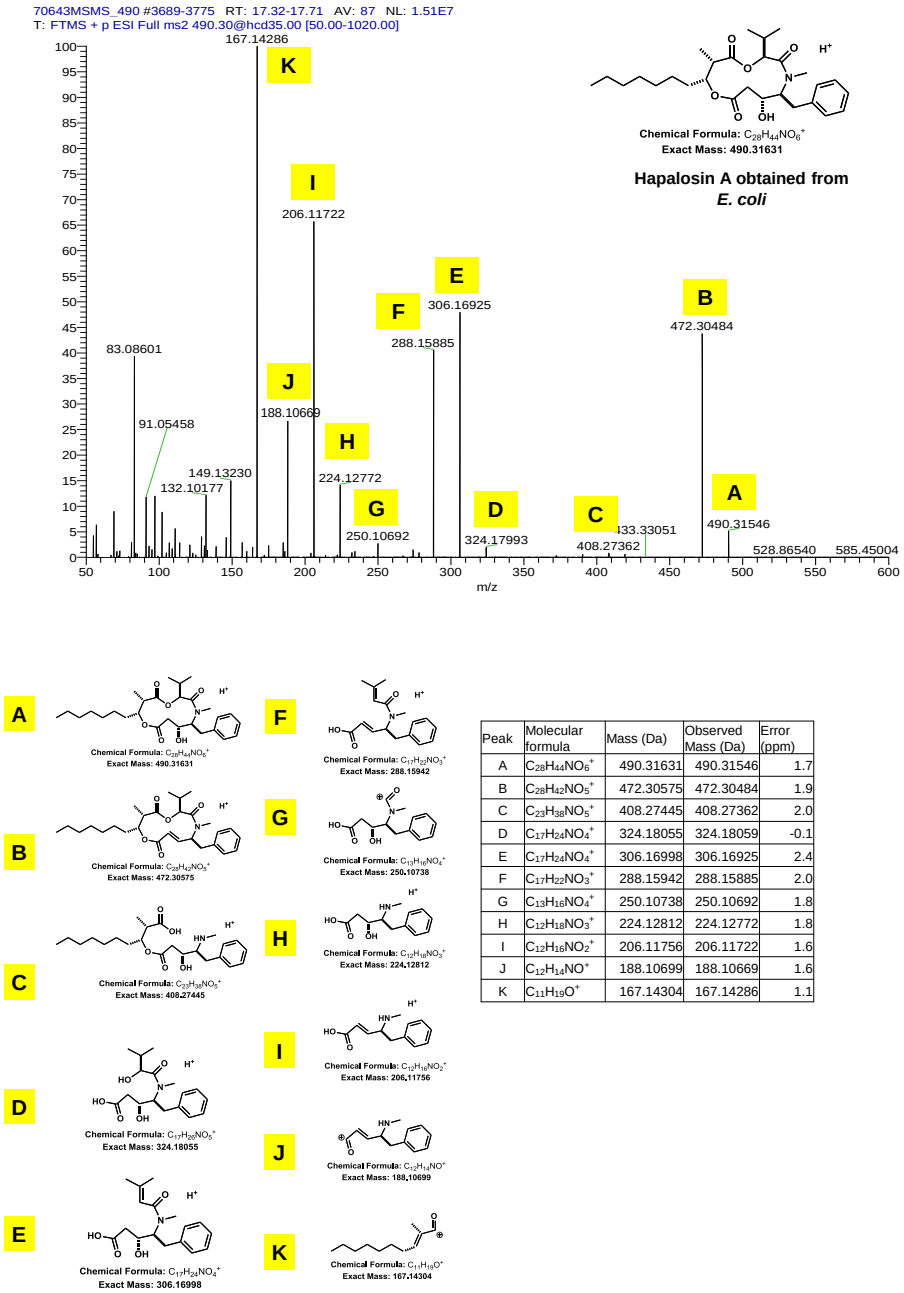


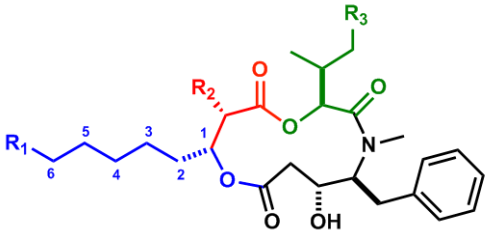
Figure 6.4 Comparative LC-MS chromatograms of hapalosin biosynthetic pathway production from *E. coli* and cyanobacterial *Hapalosiphon welwitschii* strain UTEX1830. Hapalosin A is denoted in the LC-MS traces as A for both *E. coli* and cyanobacterial extracts. Three other prominent compounds were discovered in the *E. coli* expression strain containing the hapalosin pathway, which are labeled B, C, and D. These will be discussed later in this document.



Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{28}H_{44}NO_6^+$	490.31631	490.31546	1.7
B	$C_{28}H_{42}NO_5^+$	472.30575	472.30484	1.9
C	$C_{23}H_{38}NO_5^+$	408.27445	408.27362	2.0
D	$C_{17}H_{24}NO_4^+$	324.18055	324.18059	-0.1
E	$C_{17}H_{24}NO_4^+$	306.16998	306.16925	2.4
F	$C_{17}H_{22}NO_3^+$	288.15942	288.15885	2.0
G	$C_{13}H_{16}NO_4^+$	250.10738	250.10692	1.8
H	$C_{12}H_{18}NO_3^+$	224.12812	224.12772	1.8
I	$C_{12}H_{16}NO_2^+$	206.11756	206.11722	1.6
J	$C_{12}H_{14}NO^+$	188.10699	188.10669	1.6
K	$C_{11}H_{19}O^+$	167.14304	167.14286	1.1

Figure 6.5 The HR-MS^2 fragmentation of hapalosin A obtained from heterologous expression of the biosynthetic pathway in *E. coli*. Key fragments and structures are labeled from A – J.

The structural elucidation of hapalosins C and D and consideration of the enzymes responsible for each chemical transformation in the biosynthesis of hapalosin suggested substrate promiscuity of the fatty acid ligase domain of HalA and differential reactivity at the *C*-methyltransferase domain of HalB. Therefore, in addition to substrate promiscuity of the adenylation domain of HalC to yield hapalosin B, the production of the three hapalosins B, C, and D serves as evidence that at least three points within the biosynthetic pathway result in the divergent synthesis of hapalosin compounds.



Hapalosin	R_1	R_2	R_3
A	$-CH_2CH_3$	$-CH_3$	-H
B	$-CH_2CH_3$	$-CH_3$	$-CH_3$
C	-H	$-CH_3$	-H
D	$-CH_2CH_3$	-H	-H

Figure 6.6 Structures of hapalosins A-D provide insight into the differential biosynthesis of hapalosin products within the pathway. Colored atoms are derived from substrates of different Hal modules. Green atoms correspond to substrates of HalC, blue to HalA, and red to HalB.

6.4 DEFINING THE HAL MINIMAL PATHWAY

To determine the genes essential in the production of hapalosin, *E. coli* expression strains harboring deletion mutant variants of the pathway were assembled and expressed. As discussed previously, it was determined that expression of *halDE* and *sfp* were insufficient to biosynthesize hapalosin. However, expression of the *halBCDE* partial pathway was observed to produce

hapalosin, although yields were comparably modest to that of the full *hal* pathway (**Figure 6.7**). This unexpected observation suggested that the function of HalA could be provided by an enzyme produced in *E. coli*.

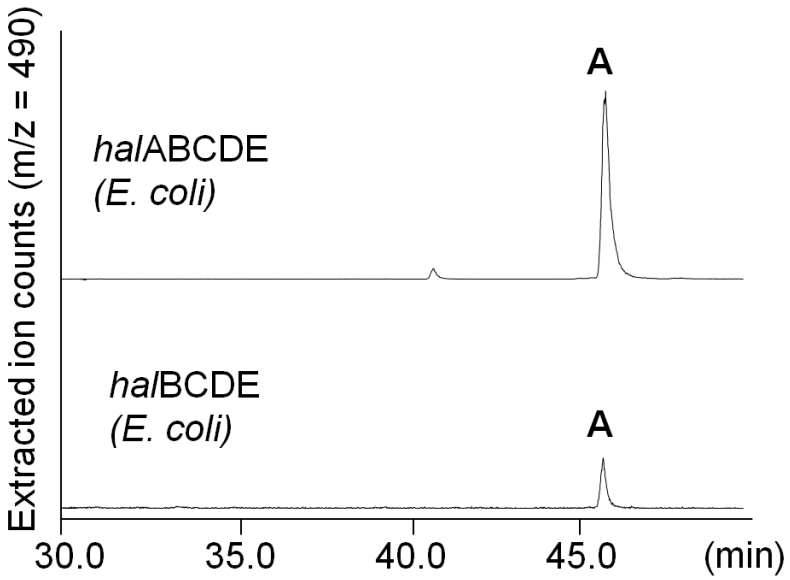


Figure 6.7 LC-MS chromatogram comparison of *E. coli* expression strains harboring the full pathway *halABCDE* or the pathway without the fatty acid ligase, *halBCDE*. Hapalosin A (labeled A) was still produced from the partial pathway strain, although production was comparably modest.

Lipoic acid in *E. coli* serves an essential role as a prosthetic group in key enzymes involved in primary metabolism. [226] The biosynthesis of lipoic acid is accomplished by two enzymes, LipA and B, which function as the lipoate synthase and octanoyltransferase, respectively. Biosynthesis initiates with the interaction of LipB with octanoyl-ACP, an intermediate of the fatty acid biosynthetic pathway, as only marginal concentrations of freely

available octanoic acid are present within the *E. coli* cell. Next, the octanoyl-LipB complex presents the octanoyl substrate to LipA, which functions to introduce the terminal thiolane ring to yield lipoic acid. [227]

Given the essential role of lipoic acid within *E. coli*, it is hypothesized that the thiolation domain of LipB interacts with the adenylation domain of HalB to shuttle the octanoyl moiety, the first substrate used for hapalosin biosynthesis, to the hapalosin pathway. Therefore, hapalosin production from the *halBCDE* partial pathway expressed in *E. coli* suggests that a PKS module from cyanobacterial (HalB) can interact with a fatty acid ligase from a different species (*E. coli*). If this interaction were to take place, it would be a rare event, as interactions between domains are highly specific. [228] This unexpected result, establishes the capacity for enzyme-enzyme promiscuity from the Hal pathway.

6.5 GENERATION OF HAPALOSIN ANALOGS

Evidence of substrate promiscuity within the hapalosin biosynthetic pathway expressed in *E. coli* consequently prompted the systematic investigation of hapalosin analogs that could be generated from other *E. coli* substrates as well as from unnatural substrates.

6.5.1 Generation of Hapalosin Libraries in Escherichia coli

Evidence of biosynthetic divergence at points HalA, HalB, and HalC of the pathway prompted an exploration for other hapalosin analogs. Thus far, promiscuity of the adenylation domains of

HalA and HalC had been established with the structural identification of hapalosins C and B, respectively. Furthermore, variation of the *C*-methyltransferase domain of HalB was established with the identification of hapalosin D. Aside from hapalosin B, none of these hapalosin compounds were present from the native cyanobacteria producer *Hapalosiphon welwitschii*. To investigate the scope of substrate promiscuity and enzymatic divergence of the Hal pathway, hypothetical hapalosin products were proposed based on alternative substrates and variations of the enzymatic machinery for each member of the pathway. Proposed hapalosin analogs from *E. coli* were elucidated via HR-MS^2.

The production of hapalosin D had demonstrated the capacity of the hapalosin intermediate to elude methylation by the *C*-methyltransferase of HalB. It was logical to inquire whether the *N*-methyltransferase of HalD could also allow hapalosin intermediates to proceed through the biosynthetic pathway to yield the non-*N*-methylated products. Examination of the heterologous expression of *halABCDE* in *E.coli* revealed the production of the non-*N*-methylated hapalosin compound, confirming that both methyltransferases present within the pathway do not proceed with completion (**Figures D10** and **D11**).

Incomplete *C*-methylation underscores the ability of these domains to accommodate the absence of methylation but did not yield direct insight into the substrate promiscuity of the HalB adenylation domain itself. Malonyl-CoA serves as the native substrate for the HalB adenylation domain, although other malonyl-CoA substrates are also commonly utilized by acyltransferase (AT) domains, including methylmalonyl- and ethylmalonyl-CoA. To examine whether other malonyl-CoA substrates can be selected by the AT domain of HalB, *halABCDE* heterologous

expression extracts from *E. coli* were analyzed for the presence of hapalosin compounds substituted with methyl- and ethylmalonyl-CoA moieties. Indeed, trace amounts of an α-*gem*-dimethyl analog, the hapalosin product of methylmalonyl-CoA incorporation, was discovered, validating the promiscuity of HalB, a rare feature of AT domains. [229]

Although the production of C6 and C8 fatty acid analogs of hapalosin (hapalosins A and C, respectively) had been confirmed, a C16 fatty acid analog had also been envisioned, due to the intracellular abundance of C16 saturated fatty acids within *E. coli*. [230] Although a C16 fatty acid hapalosin had not been discovered, screening for analogs of shorter fatty acid chain lengths revealed the presence of trace amounts of a C10 fatty acid analog instead.

This discovery suggested a chain length limitation of the fatty acid ligase domain of HalA or the association of an *E. coli* acyltransferase with HalB. The basis of the former hypothesis relies on the fact that free C10 fatty acid concentrations are considerably lower than that of other long chain fatty acids, such as C16 fatty acid, and transfer of non-native fatty acid substrates would preferentially select those in abundance. Instead, the less abundant C10 fatty acid was incorporated, evidence that acyl chains between 6 to 10 carbon atoms is likely optimal for the substrate binding pocket of the HalA fatty acid ligase.

The final domain of the Hal gene cluster consists of the thioesterase (TE) domain, responsible for the cyclization reaction to yield the final product. In the hapalosin biosynthetic pathway, a nucleophilic hydroxyl group, generated from the Claisen condensation with malonyl-CoA from the HalB domain, attacks the electrophilic carbon of the substrate-enzyme thioester

tether. This macrocyclization yields the final cyclized hapalosin product. However, a nucleophilic water molecule can compete for the electrophilic carbon of the substrate-enzyme thioester tether. This results in premature release to yield linearized products. Examination of the *halABCDE* pathway expression from *E. coli* demonstrated that six different linear hapalosin products can be observed, suggesting that the hapalosin intermediate can be prematurely hydrolyzed from the TE domain of HalE. Alternatively, it is possible to explain the observation of these linear hapalosin products due to hydrolysis of the mature macrocycle after its complete assembly. This mechanism would also suggest that products would arise from the hydrolysis of the other lactone moiety. However, examination of the *halABCDE* pathway expression from *E. coli* did not yield the presence of such alternatively hydrolyzed hapalosin products, which supports the hypothesis that these linear hapalosin compounds arise due to the premature hydrolysis of the hapalosin intermediate from the TE domain of HalE.

6.5.2 The Generation of Unnatural Hapalosin Analogs

The unusual ability of the HalA fatty acid ligase to select longer and shorter acyl chain substrates motivated the investigation of unnatural substrate (azido-conjugated) selectivity of this fatty acid ligase. Furthermore, the specificity of the adenylation domain of HalD was examined through *E. coli* feeding studies to incorporate halogen-substituted phenylalanine substrates.

Fatty acid substrates containing a terminal azide were synthesized, but the choice of fatty acid chain length was carefully considered based on previous findings of fatty acid substrate selectivity. It was demonstrated that C8 fatty acids proved to be the preferred substrate, while C6 and C10 fatty acids appeared to serve as lower and upper limits in chain length. Based on these

findings, 6-azidohexanoic, 7-azidoheptanoic, and 8-azidooctanoic acids were synthesized and used as feeding substrates to *E. coli* expression cultures (**Figure 6.8**).

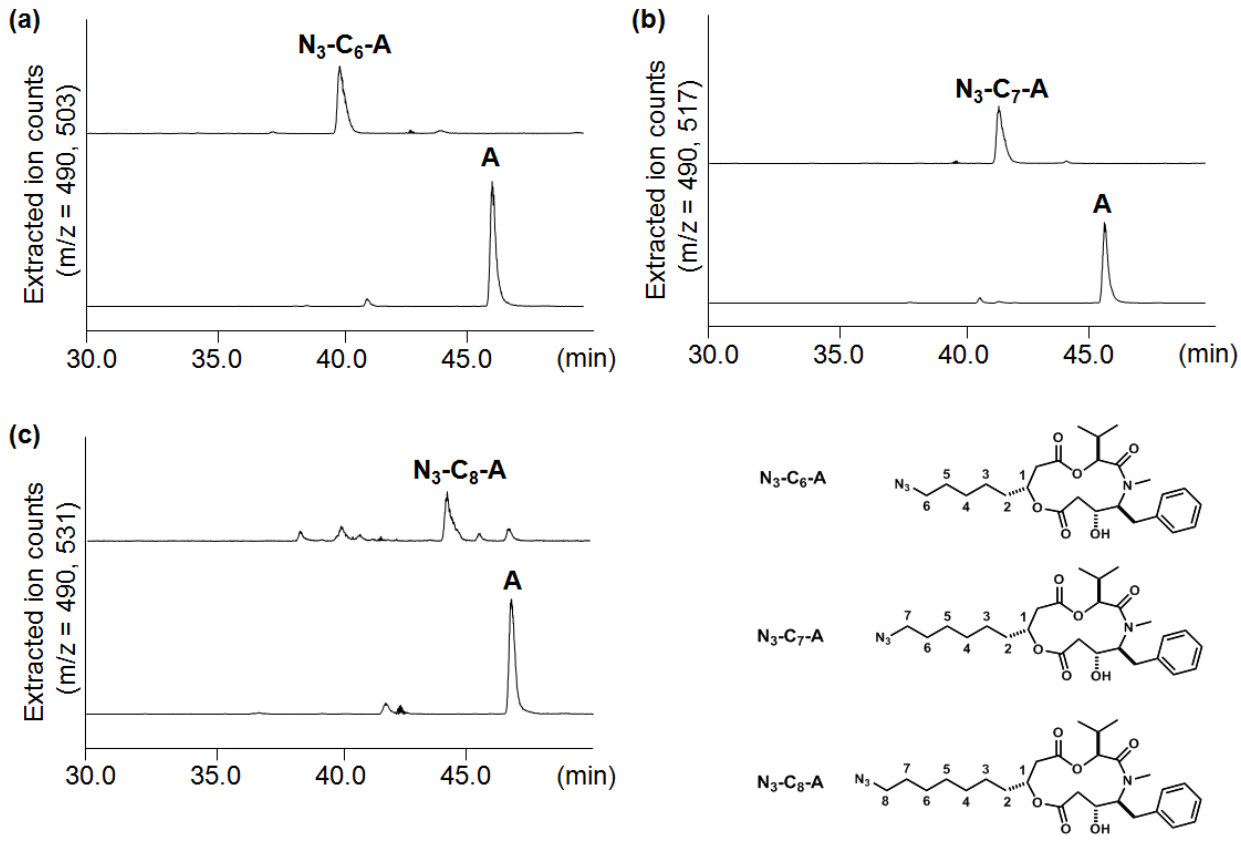


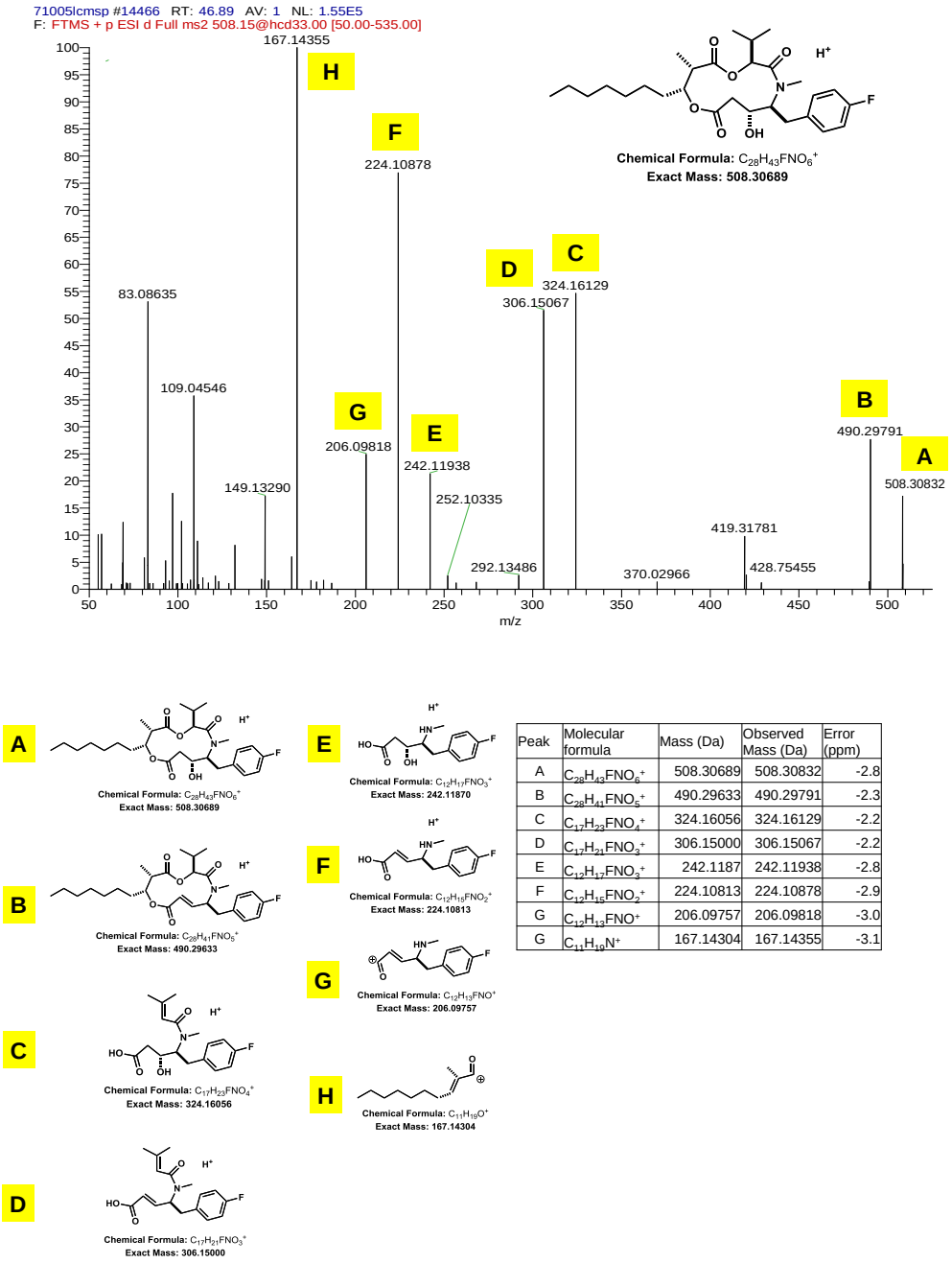
Figure 6.8 Azido-alkanoic acid feeding experiments in *E. coli* expression cultures. Hapalosin A production from each expression culture is shown for comparison (labeled A).

LC-MS and HR-MS^2 analyses indicated that each of the azido-alkanoic acids were incorporated into the Hal pathway to yield their corresponding azide-conjugated hapalosin products. While the yield of azido-alkanoic acid incorporation was comparable between the 6-azidohexanoic and 7-azidoheptanoic acids, the yield was markedly decreased for the 8-azidooctanoic acid incorporation. In fact, to facilitate incorporation of the 8-azidooctanoic acid

moiety, the final concentration of this substrate was doubled in the culture medium, compared to that of the other azido-alkanoic acid substrates, in order to observe modest incorporation.

This series of unnatural fatty acid feeding experiments serves as further evidence that a fatty acid chain length limitation exists for the incorporation of the substrate into the Hal pathway. An inspection of the natural and unnatural substrates optimal for incorporation suggests that chains of 8 to 9 atoms following the carboxylic acid (i.e., counting from the α carbon to the last C–H or N=N bond of the chain) is the optimal upper limit, since octanoic, 6-azidohexanoic, and 7-heptanoic acids were all incorporated with comparable efficiency, based on LC-MS analysis. On the other hand, a substrate with a chain of 10 atoms following the carboxylic acid displayed a negative bias toward incorporation. It would therefore be unsurprising if longer acyl chains could not be incorporated at all and also explains why C12, C14, and C16 hapalosin analogs were not detected, even by HR-MS^2 analysis.

Previous attempts to detect natural analogs of hapalosin bearing substitutions of the phenylalanine residue with tyrosine or tryptophan were unsuccessful and were attributed to high substrate selectivity of the adenylation domain of HalD. To test whether increased concentrations of an unnatural phenylalanine substrate could be incorporated instead, 4-fluoro-phenylalanine was introduced at concentrations up to 5mM in the growth medium. Trace amounts of the fluorinated hapalosin compound was subsequently discovered by HR-MS^2 in the expression culture supernatant (**Figure 6.9**).



Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{28}H_{43}FNO_6^+$	508.30689	508.30832	-2.8
B	$C_{28}H_{41}FNO_5^+$	490.29633	490.29791	-2.3
C	$C_{17}H_{23}FNO_4^+$	324.16056	324.16129	-2.2
D	$C_{17}H_{21}FNO_3^+$	306.15000	306.15067	-2.2
E	$C_{12}H_{17}FNO_3^+$	242.1187	242.11938	-2.8
F	$C_{12}H_{15}FNO_2^+$	224.10813	224.10878	-2.9
G	$C_{12}H_{13}FNO^+$	206.09757	206.09818	-3.0
G	$C_{11}H_{19}N^+$	167.14304	167.14355	-3.1

Figure 6.9 HR-MS^2 analysis and fragmentation assignments of a fluorinated hapalosin analog.

As another example of unnatural amino acid incorporation into the hapalosin structure, 4-chloro-phenylalanine was also introduced to *E. coli* expression cultures to generate the corresponding chlorinated hapalosin analog, which was also detected by HR-MS^2. Although halogenated amino acids had been successfully incorporated into the hapalosin structure, the yields of corresponding halogenated hapalosin analogs were considerably lower than that of the azide-modified hapalosin structures. The capacity of the adenylation domain of HalD to select for unnatural substrates demonstrates that in total, there exist six points of biosynthetic divergence in the Hal pathway. These points of biosynthetic divergence allowed the generation of a combinatorial library of hapalosin analogs with multiple substitutions, resulting in the generation of a total of 32 hapalosin analogs.

6.6 CONCLUSIONS

The present study demonstrates the utility of the demonstrated expression plasmid used to investigate biosynthetic pathways of cyanobacterial natural products. This study identified the gene cluster responsible for the production of the multidrug-resistance (MDR) reversing agent hapalosin from *H. welwitschii* UTEX1830B. Assembly and heterologous expression of this large (26-kbp) gene cluster in *E. coli* revealed the unexpected generation of hapalosin analogs, owing to the unusual plasticity of this pathway. The incorporation of unnatural substrates (azide-conjugated fatty acids and halogenated amino acids) underscores the potential of this cloning and expression system as a platform for generating cyanobacterial natural product analogs via precursor-directed biosynthesis. [231]

The incorporation of "clickable" functional groups (propargylation or azido-conjugation) to afford analogs of natural products is highly desired and has been demonstrated in polyketides [232] and nonribosomal peptides [233] through mutagenesis studies of acyltransferase and adenylation domains, respectively. The incorporation of click functional groups can enable the facile synthesis of novel structures with enhanced bioactivities. These analogs can also serve as small-molecule probes via bioorthogonal conjugation to investigate the biochemical mechanisms of these therapeutic compounds, accelerating drug development [234]. Likewise, the introduction of carbon-halogen bonds into natural products can expand the potential of native scaffolds. For example, the introduction of a carbon-fluorine bond confers bioactivity for some drugs (reviewed in [235]). The carbon-chlorine bond is similarly important because it can serve as a chemical handle to selectively introduce new functional groups. An approach to generate analogs of the peptide antibiotic pacidamycin was recently demonstrated through the biosynthesis of the chlorinated product, followed by selective cross-coupling various moieties [236].

Recent genome sequencing efforts of several cyanobacteria species have revealed that these organisms are capable of producing numerous cryptic secondary metabolites with potential anticancer, anti-inflammatory, antibiotic, and anti-fungal activities. However, due to the lack of genetic tools and to difficulties in their genetic manipulation, investigating their biosynthesis has been elusive. This study has established a designer plasmid that can rapidly assemble and introduce engineered regulatory elements to link cyanobacterial biosynthetic pathways to their natural products. This plasmid has been used to characterize the biosynthetic pathway for the MDR reversing agent hapalosin and the generation of a library of unnatural (azide-conjugated and halogenated) analogs via heterologous expression in host *E. coli*. This system represents a

promising approach to discover new cyanobacterial natural products and to produce analogs with biochemically or clinically significant properties.

7.0 THE HAPALOSIN BIOSYNTHETIC GENE CLUSTER FROM HAPALOSIPHON WELWITSCHII UTEX1830 REVEALS A NOVEL SUBSTRATE RECYCLING MECHANISM

Hapalosin is a cyclic depsi-lipopeptide from the stigonematalean cyanobacterium *Hapalosiphon welwitschii* UTEX1830B and exerts multidrug resistance reversing activity via P-glycoprotein inhibition [212]. The biosynthesis of hapalosin proceeds via a hybrid nonribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) enzymatic assembly line. The general biosyntheses of depsipeptides proceed through NRPS enzymatic machinery, condensing α-hydroxy acid substrates to elongate the growing peptide chain. Recent investigations of biosynthetic pathways of depsipeptides have revealed that two distinct mechanisms exist to incorporate α-hydroxy acids. In one scenario, the α-keto acid substrate is selected and activated by its corresponding adenylation domain and is then reduced by a subsequent reductase domain to yield the α-hydroxylacyl-*S*-pantetheinyl carrier protein intermediate. This has been shown in the biosynthesis of the ionophore cereulide from *Bacillus cereus* [223]. In the other scenario, the α-hydroxy acid substrate is selected directly for activation and then incorporated into the growing peptide chain. This mechanism exists in the biosynthesis of fungal cyclodepsipeptides, including the anthelmintic PF1022A from *Rosellinia* sp. PF1022 [237] as well as the enniatins, ionophores from *Fusarium* sp. [238]. Noteworthy from both mechanisms of depsipeptide

biosynthesis is the high selectivity of the adenylation domain for either its cognate α-hydroxy or α-keto acid but not both.

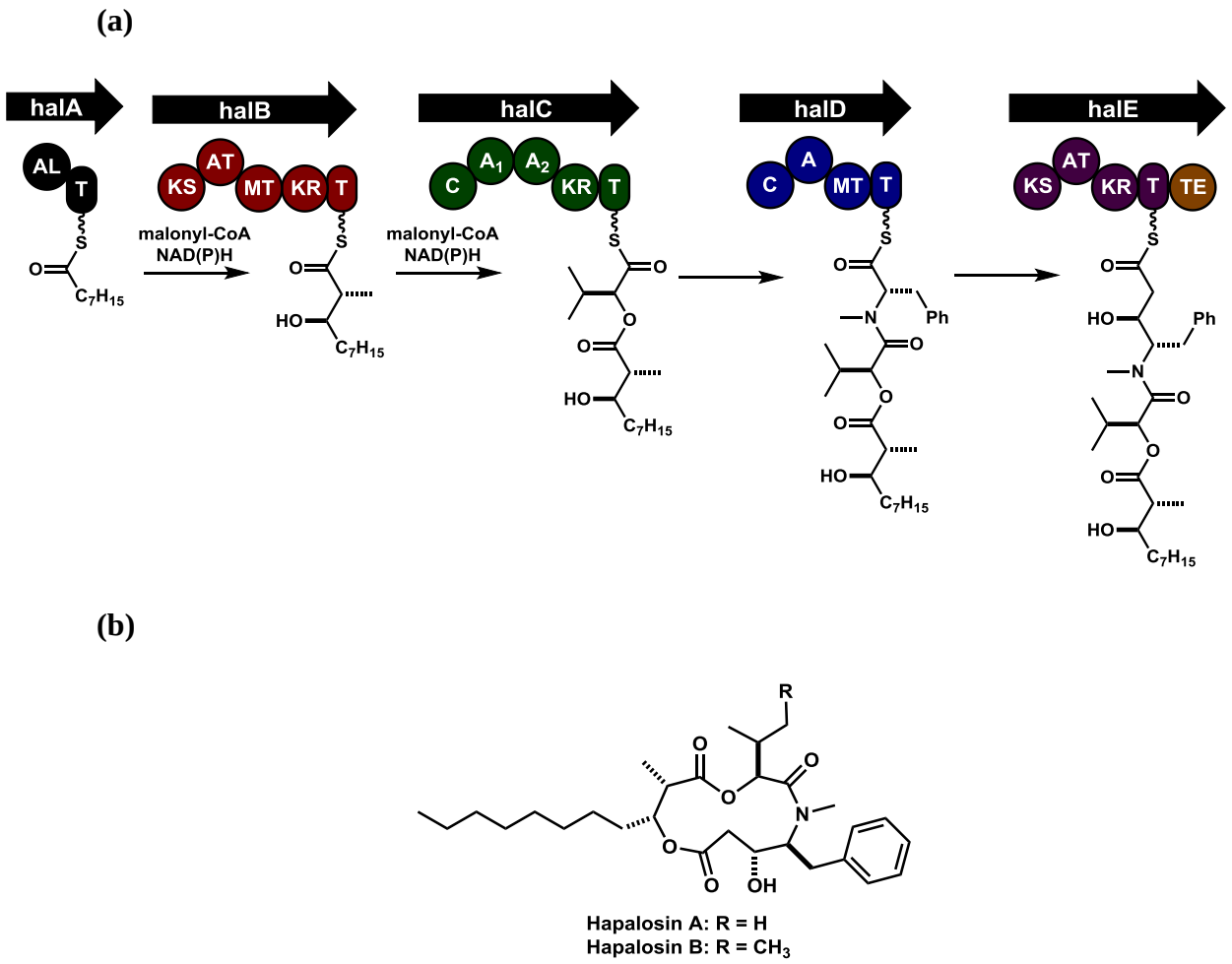


Figure 7.1 (a) Hapalosin biosynthetic gene cluster organization from *H. welwitschii* UTEX1830B. **(b)** Structures of hapalosin A and B.

Recent genome sequencing of the *H. welwitschii* UTEX 1830B genome [219] revealed that hapalosin is biosynthesized via a hybrid pathway containing NRPS and PKS enzymes (**Refer to Hapalosin Gene Cluster, Chapter 6**). The hapalosin (*hal*) biosynthetic pathway

consists of five genes, *hal*ABCDE, corresponding to a fatty acid ligase, NRPS, PKS, NRPS, and NRPS, respectively (**Figure 7.1a**). Interestingly, HalC contains tandem adenylation domains followed by a ketoreductase domain. Canonical NRPS modules contain just a single adenylation domain for the activation and incorporation of a single amino or keto acid substrate. Therefore, the presence of tandem adenylation domains is an unusual feature of this NRPS enzyme. Based on bioinformatic analysis, the adenylation domain(s) of the HalC module are predicted to be responsible for the activations of the α-ketoisovaleric acid (**1**) and α-keto-β-methylvaleric acid (**3**), which are subsequently reduced to the corresponding hydroxy acids by a downstream ketoreductase domain of the HalC domain. The incorporation of activated substrates **1** and **3** lead to the biosynthesis of hapalosin A and B, respectively (**Figure 7.1b**). In this study, the unusual tandem adenylation domains and ketoreductase domain of HalC of the hapalosin biosynthetic pathway are characterized *in vitro* and *in vivo* via heterologous expression in *E. coli*. Insights into this pathway revealed a novel mechanism for the incorporation of the α-keto(hydroxy) acid substrates into the hapalosin structure.

7.1 HALC-A2 DOMAIN EXHIBITS SUBSTRATE ACTIVATION PROMISCUITY

To determine the function of both adenylation domains (A1 and A2) present in the HalC module of the *hal* pathway, the sequences corresponding to each domain were individually cloned and expressed to afford purified proteins of each stand alone domain. The HalC adenylation domains were screened against a panel of substrates including **1** and **3**, as well as their reduced, α-hydroxy acid analogs **2** and **4**, respectively (**Figure 7.2**). Sequence analysis of the HalC A1 domain revealed two point mutations within the amino acid sequence of the HalC module, R922S and

K1011Q, predicted to be essential for substrate adenylation (**Figure E1**). Indeed, *in vitro* activity screening of the HalC A1 domain against substrates **1** – **4** did not exhibit any appreciable ATP hydrolysis activity, corresponding to substrate adenylation.

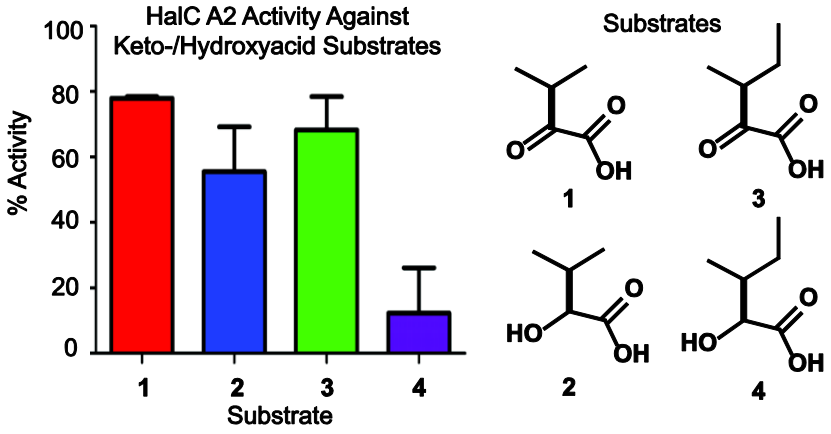


Figure 7.2 HalC_A2 domain adenylation activity towards four substrates.

The HalC A2 domain was therefore expected to be responsible for the activation of substrates **1** and **3** for final biosynthesis of hapalosins A and B, respectively. The *in vitro* adenylation activity of the HalC A2 domain revealed comparable enzyme activity for substrates **1** and **3** (**Figure 7.2**). Reduced native substrate analogs **2** and **4** were also screened for HalC A2 enzymatic activity, and unexpectedly, substrate **2** showed comparable enzyme activity as that of **1** and **3**.

7.2 HALC A2 ENZYMATIC ACTIVITY AGAINST A-KETO AND –HYDROXY ACID SUBSTRATES

The Michaelis-Menten kinetic parameters for the HalC A2 domain against substrates **1** and **2** were determined (**Figure 7.3**). While the turnover number for substrates **1** and **2** were comparable (0.95 ± 0.07 min^{-1} and 1.11 ± 0.16 min^{-1}, respectively), the K_m values for **1** and **2** were determined to be 32.10 ± 5.45 μM and 125.8 ± 34.47 μM, respectively. The catalytic efficiency for the activation of **1** was approximately 3-fold greater than that for **2** (2.97×10^{-2} min^{-1} μM^{-1} versus 0.89×10^{-2} min^{-1} μM^{-1}). The smaller K_m value and greater catalytic efficiency for the activation of **1** over **2** is consistent with the hypothesis that **1** is the native substrate for the HalC A2 domain.

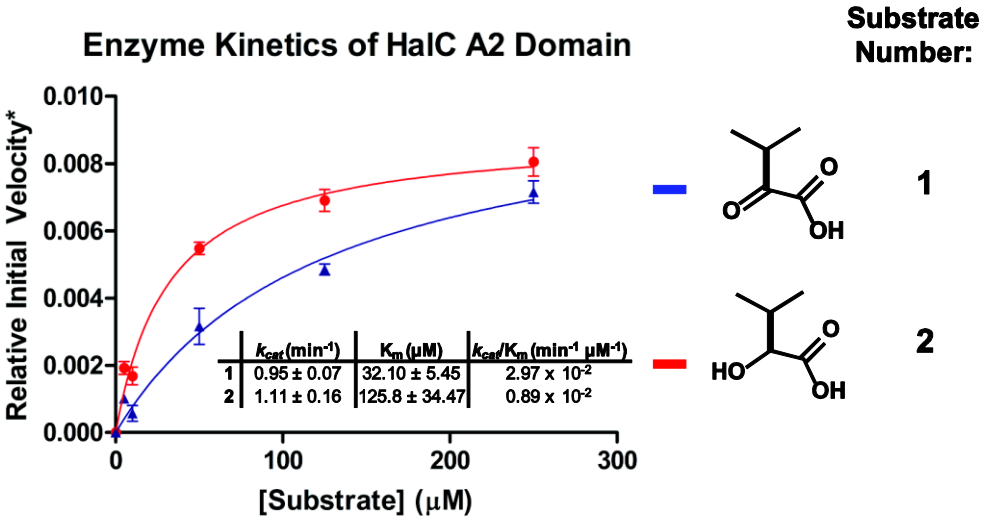


Figure 7.3 HalC_A2 domain kinetic profiles in the activation of α-ketoisovaleric acid and α-hydroxy isovaleric acid.

7.3 CHEMICAL COMPLEMENTATION OF THE HAL MUTANT PATHWAY

The unexpected adenylation domain activity of HalC A2 against substrate **2** suggested that the α-hydroxy acid substrate can be activated and circumvent the ketoreductase domain of the HalC module. To confirm that the biosynthesis of hapalosin A can proceed in absence of HalC ketoreductase activity and with the introduction of **2**, mutations S2071A and Y2086F were introduced into active site of the HalC domain amino acid sequence. The mutant HalC sequence was then used to assembly the remainder of the *hal* pathway for heterologous expression in *E. coli*. Expression of the mutant pathway afforded the production of hapalosin A only with the introduction of **2** (**Figure 7.4**). The absence of hapalosin A production from the expression of the mutant pathway without **2** suggests that the HalC A2 domain can activate both **1** and **2** to afford the final hapalosin A product.

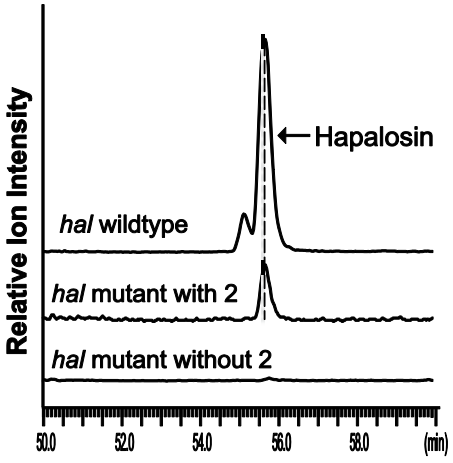


Figure 7.4 Hapalosin production in *E. coli* is abolished by introducing HalC_KR active site mutations S2071A and Y2086F. The mutant pathway can be complemented by the addition of **2**.

7.4 A NOVEL MECHANISM FOR RECYCLING A-KETO AND -HYDROXY ACID SUBSTRATES

Substrate activation of the adenylation domains of the HalC module *in vitro* has revealed that the A1 adenylation domain is not responsible for the activation of either α-keto or -hydroxy acid substrates, which is consistent with predicted inactivity of this domain based on sequence analysis. On the other hand, the A2 adenylation domain was observed to be responsible for the activation of the α-keto acid substrates **1** and **3**, which incorporated into the final structures of hapalosins A and B, respectively (**Figure 7.2**). Unexpectedly, the α-hydroxy acid substrate analog **2** was also activated by the HalC A2 domain. Comparable catalytic efficiency for both substrates (**Figure 7.3**) suggests that the biosynthetic machinery from the native producer *H. welwitschii* UTEX1830B utilizes either substrate in the maturation of hapalosin. Chemical complementation of the HalC ketoreductase mutant of the hapalosin pathway expressed heterologous and in the presence of α-hydroxy acid substrate **2** confirms this finding (**Figure 7.4**). Collectively, these observations led to the proposition of a novel recycling mechanism of the α-hydroxy isovaleric acid by the HalC module (**Figure 7.5**).

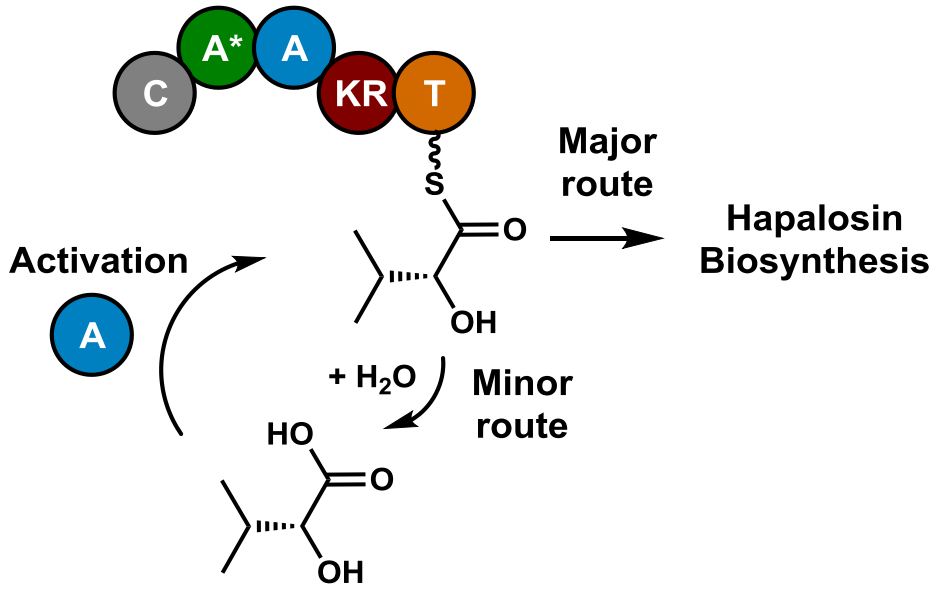


Figure 7.5 Proposed recycling mechanism of α-hydroxy isovaleric acid by HalC module.

Activation and subsequent reduction of the α-keto acid substrate affords the corresponding α-hydroxylacyl-*S*-pantetheinyl carrier protein intermediate. This intermediate can subsequently undergo one of two pathways: 1) it is utilized in elongation to produce the final hapalosin structure or 2) it is hydrolyzed from the thioester tether to yield the corresponding α-hydroxy keto acid. The latter (minor pathway) will then undergo (re)activation to regenerate the α-hydroxylacyl-*S*-pantetheinyl carrier protein intermediate, which would bypass ketoreductase activity to proceed with the subsequent biosynthetic steps in the hapalosin pathway.

7.5 CONCLUSIONS

The hapalosin biosynthetic pathway from cyanobacterial species *H. welwitschii* UTEX1830B presents uncommon evolutionary features. The HalC module contains unusual tandem

adenylation domains, presumably as the result of a gene duplication event; however, the first of these domains (HalC A1) contains active site point mutations that render this adenylation domain inactive. On the other hand, the second domain (HalC A2) unexpectedly displayed substrate promiscuity for both the α-keto and -hydroxy valeric acid substrates, which demonstrated a novel substrate recycling mechanism. The proposed α-hydroxy acid substrate recycling mechanism is the first of such to be characterized. Adenylation domain activation of previously characterized NRPS pathways [212, 223, 237, 238] has demonstrated that these domains are only capable of activating either the α-keto or -hydroxy acid, not both. This finding is especially unusual as adenylation domains are well established to be highly selective in substrate activation [239]. It is conceivable that the ability of the HalC A2 adenylation domain to accept both α-keto and -hydroxy acid substrates is an evolved strategy to utilize the α-hydroxy acid monomer in order to mitigate the energy cost incurred in the re-activation and reduction of α-keto acid substrates that were released prematurely from the biosynthetic assembly line. This novel observation represents an additional mechanism by which microorganisms can utilize substrates in the maturation of NRPS natural products as strategies to enhance their competitiveness in diverse ecological environments.

8.0 HIDDEN OXIDATIVE DIVERSITY OF MICROCYCLAMIDE REVEALED BY PATHWAY REFRACTORING IN E. COLI: IMPLICATION THAT LATE-STAGE OXIDATIVE MATURATION OF THIAZOLE AND OXAZOLE IN CYANOBACTIN BIOGENESIS IS DIRECTIONAL AND PROCESSIVE

Cyanobactins are ribosomally synthesized peptides from cyanobacteria that exhibit bioactivities, including antiviral, antimalarial, and anticancer properties [172, 173, 240]. Cyanobactin secondary metabolites are ubiquitous, as an estimated 30% of all cyanobacterial species produce these molecules [171, 240], evidence to their physiological and ecological significance. The mechanisms of their biosynthesis are of special interest because the machinery to produce compounds of this class of natural products is highly amenable to molecular engineering to generate combinatorial libraries with the potential for novel and enhanced bioactivities [185, 221].

The microcyclamides and aerucyclamides are cyclic hexapeptidic cyanobactins from the cyanobacterium *Microcystis aeruginosa* [174, 209, 241] that have demonstrated cytotoxic activity [174, 209]. The biosynthetic pathway for microcyclamides and aerucyclamides (the *mca* operon) was previously characterized [197] and revealed this operon to share marked sequence similarity to the patellamide biosynthetic gene cluster, demonstrating that the microcyclamides and aerucyclamides are members of the cyanobactin family of secondary metabolites.

Recent genome sequencing of dozens of cyanobacterial strains have revealed that these microorganisms are capable of producing numerous secondary metabolites, most of which are cryptic but may serve as promising scaffolds for therapeutic compounds [171]. However, major challenges in accessing and investigating these molecules arise due to the genetic intractability of these organisms and the lack of genetic tools for their manipulation. To address the shortcoming toward the rapid characterization of biosynthetic pathways of cyanobacterial natural products, our research group has recently developed a designer plasmid that can rapidly assemble complete biosynthetic pathways (**Refer to Plasmid Tools, Chapter 5**). In this study, the *mca* biosynthetic pathway from *M. aeruginosa* PCC7806 was assembled and heterologously expressed. The systematic generation and expression of *mca* pathway mutants allowed for the determination of the genes essential in the biosynthesis of the microcyclamides. Furthermore, the heterologous expression of the *mca* operon revealed the generation of an unexpected diversity of oxidative patterns throughout the microcyclamide structures. The pattern of oxidation revealed that a directionality mechanism exists in the oxidation of azoline groups around the microcyclamide macrocycle. The findings of this study demonstrate that the heterologous expression and investigation of the *mca* gene cluster can reveal biosynthetic mechanisms that are significant in the maturation of azol(in)e-containing cyanobactins.

8.1 MICROCYCLAMIDES GENE CLUSTER ORGANIZATION

The microcyclamide biosynthetic pathway (*mca*) has been previously identified bioinformatically in two strains of *M. aeruginosa*, NIES298 and PCC7806 [197], and resembles the canonical cyanobactin biosynthetic pathway of the patellamides (**Figure 8.1a**). The *mca*

biosynthetic pathway affords three distinct precursor peptides that are encoded for by a single gene *mcaE*. Interestingly, there exists a duplicate copy for the peptide sequence for aerucyclamide A and B. The McaE precursor peptide (**Figure 8.1b**) is predicted to cleave at the N-terminus by McaA. McaG is then expected to both cleave the precursor peptide at the C-terminus and macrocyclize the peptide. McaF shares sequence homology to PatF, a putative prenyltransferase, but its role is unclear in the biosynthesis of the microcyclamides. Microcyclamide and aerucyclamide structures previously characterized are shown in **Figure 8.1c**.

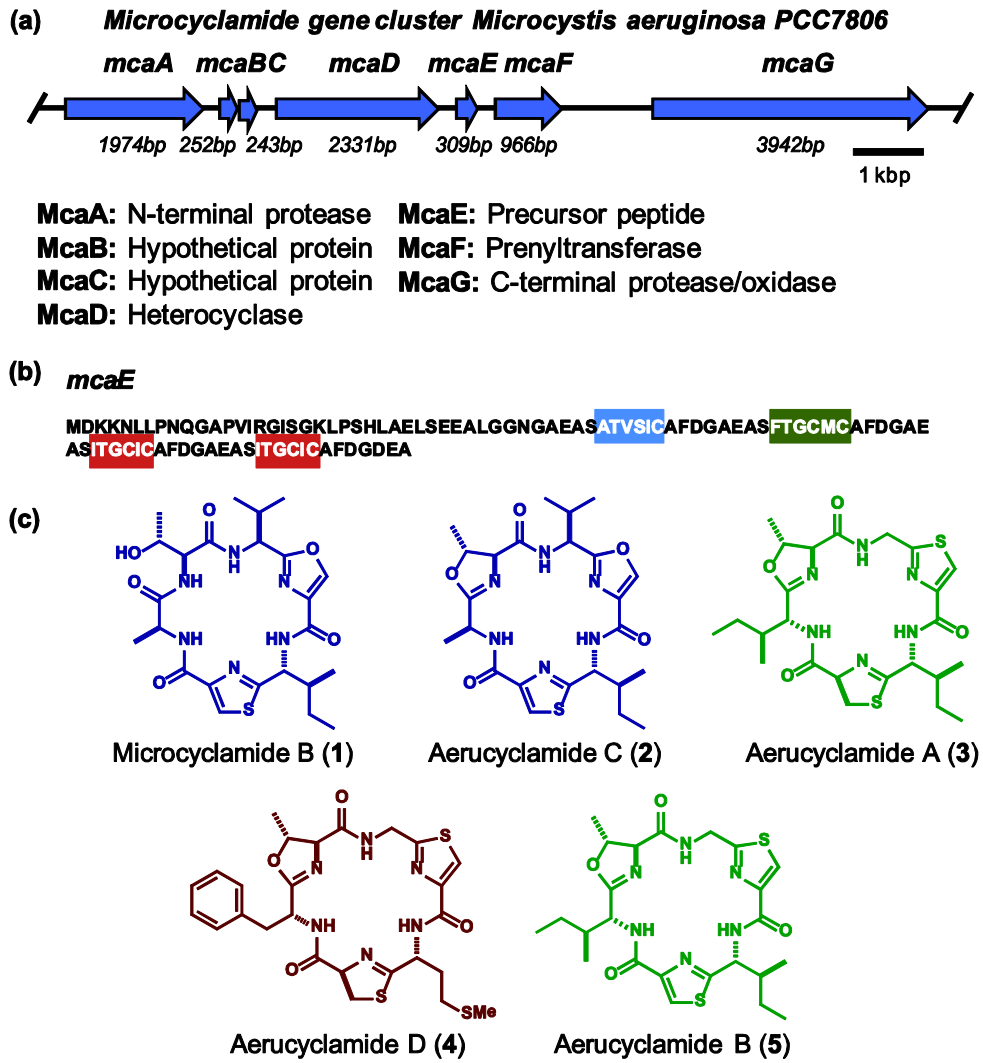


Figure 8.1 (a) Microcyclamide (*mca*) operon organization from *M. aeruginosa* PCC7806. **(b)** McaE precursor peptide sequences with residues presented in the microcyclamide or aerucyclamide structures highlighted; sequences highlighted in blue, green, and red correspond to structures **1**/**2**, **3**/**5**, and **4**, respectively. **(c)** Microcyclamide and aerucyclamide structures previously observed from *M. aeruginosa* PCC7806.

Similar to the patellamides, the microcyclamides include serine, threonine, and cysteine residues that undergo heterocyclization via the PatD-like heterocyclase McaD to afford the corresponding azoline [197]. Following azoline formation, heterocycles undergo further oxidation by the PatG-like oxidase McaG to afford the azole. The directional sequence of azoline oxidation to the corresponding azole throughout the microcyclamide structure is still an unexplored aspect of cyanobactin maturation.

8.2 SYSTEMATIC CLONING AND HETEROLOGOUS EXPRESSION OF THE MCA GENE CLUSTER

The *mca* gene cluster was cloned from PCC7806 using a designer plasmid based on yeast recombineering, as described previously (**Refer to Plasmid Tools, Chapter 5**). Three expression plasmid constructs were generated (**Figure 8.2a**), containing the complete pathway (*mca*A to G) or partial pathways (*mca*ADEGF and *mca*ADEG). As expected, expression of a plasmid containing the *mcaA* to *mcaG* operon resulted in the production of microcyclamides and aerucyclamides that have been previously characterized from *M. aeruginosa* PCC7806 (**Figure 8.2b**). This observation also validates the utility of this designer plasmid and expression platform to rapidly assemble and characterize biosynthetic pathways. In addition to expression of the complete *mca* pathway, re-organized pathways consisting of the *mcaADEG* and the *mcaADEFG* pathways were explored to determine the genes essential for biosynthesis. Genes *mcaB* and *C* were deleted from the plasmid containing the wild type operon, as these homologs of unknown function were discovered to be non-essential for the maturation of cyanobactins [197, 221]. The restoration of microcyclamides production only after the introduction of *mcaF* to *mcaADEG* to

afford the *mcaADEFG* pathway (**Figure 8.2b**) demonstrates that *mcaF* is an essential gene in the *mca* biosynthetic pathway. McaF shares sequence homology to the prenyltransferase PatF. Although the function of the PatF-like protein in the maturation of cyanobactins is unclear [175], it has been shown to be essential in their biosynthesis. The restoration of microcyclamides production by introduction of *mcaF*, therefore, appears to be consistent with previous observations of the essential role of PatF-like proteins.

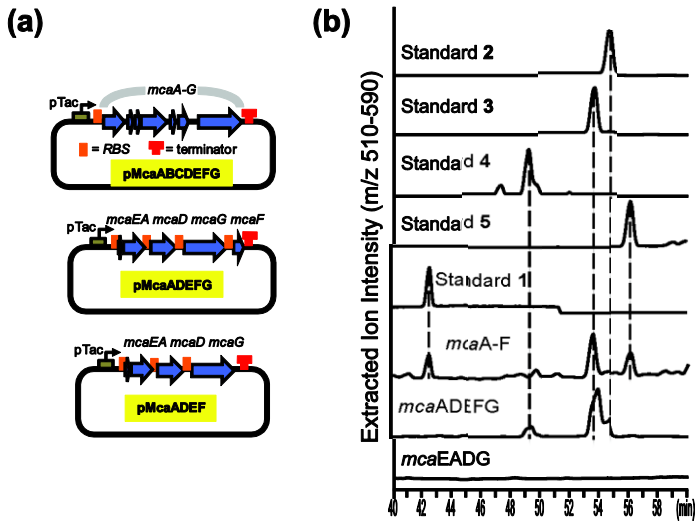


Figure 8.2 (a) Expression plasmid constructs of the *mca* operon to afford the complete pathway (*mca*A to G) or partial pathways (*mca*ADEGF and *mca*ADEG). **(b)** Comparative LCMS analysis between the native producer *M. aeruginosa* PCC7806 and the heterologous expression of the *mca* constructs in *E. coli*.

8.3 UNEXPECTED DISCOVERY OF OXIDATIVE ANALOGS OF MICROCYCLAMIDES

MS analysis of the microcyclamides revealed that these structures readily undergo fragmentation to yield MSMS spectra that further validated the characterization of the microcyclamides and aerucyclamides from the heterologous host *E. coli* compared to those from the native producer *M. aeruginosa* PCC7806. Characteristic ("fingerprint") fragmentation ions were identified for each of the known microcyclamides. The fingerprint fragmentation ions for aerucyclamide A are highlighted in **Figure 8.3a**. Furthermore, LCMS data of the *mcaABCDEFG* metabolite expression profile revealed the co-elution of known microcyclamides in addition to compounds with nominal masses that differ by two Daltons. As an example, the high resolution MS for aerucyclamide A (nominal m/z value of 535) revealed the presence of an unknown compound with a nominal m/z value of 537 (**Figure 8.3b**). In addition, the predicted molecular formula from the high resolution accurate mass values for this putatively related compound suggested that it could contain two additional hydrogen atoms. As oxidation of the azoline groups is a major and shared feature amongst the microcyclamides, this led to the hypothesis that these unknown compounds may correspond to oxidative analogs that feature azoles in place of azolines or vice versa. To determine the structure of these compounds and their potential relation to the microcyclamides, the MSMS fragmentation patterns were characterized for known microcyclamides. Comparative MSMS fragmentation analysis between aerucyclamide A and the ion with m/z value of 537 revealed that this compound was an oxidative analog of aerucyclamide A containing two thiazoline groups instead of one thiazoline and one thiazole, as in the structure of aerucyclamide A.

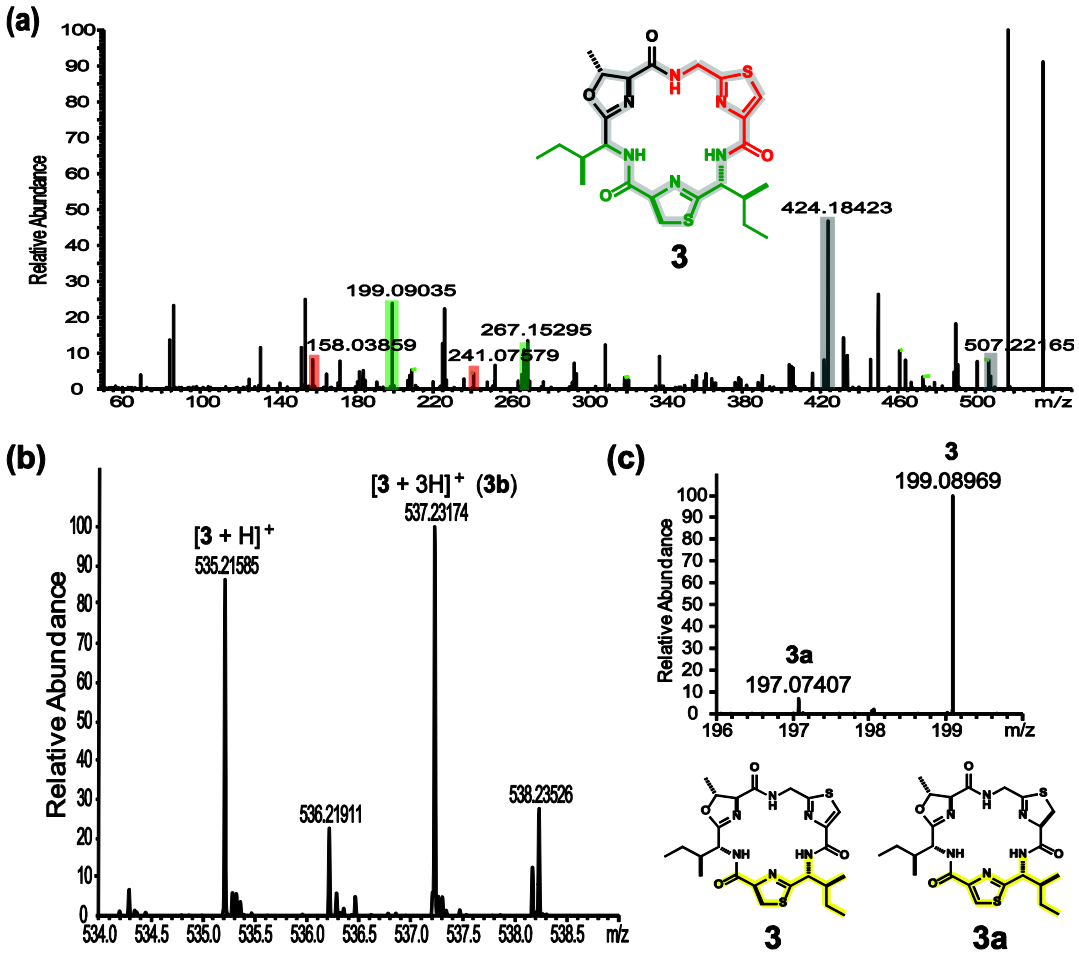


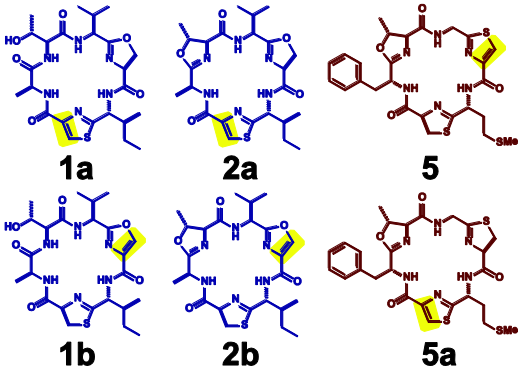
Figure 8.3 (a) MSMS spectrum of **3** with fingerprint ions enabled the structural characterization of the microcyclamide and aerucyclamide compounds. **(b)** MS spectrum of **3** revealed the presence of an ion corresponding to an analog with two additional Daltons. **(c)** MSMS of a chromatographic peak corresponding to both **3** and **3a** revealed the preferential production of **3**.

8.4 MCAG OXIDATION IS A DIRECTIONAL AND LATE-STAGE TRANSFORMATION

Hypothetically, a microcyclamide structure containing one azole and one azoline ring could be arranged in one of two ways to afford analogs with identical molecular formulae. For example, the thiazole of aerucyclamide A could be situated in the position of the thiazoline ring and vice versa to afford two distinct oxidative analogs with identical molecular formulae. To determine the exact oxidative position of the azol(in)e rings around the microcyclamides macrocycles, these structures were characterized via comparative MSMS fragmentation using fingerprint ions determined for known structures. It was observed that indeed positional oxidative analogs (i.e., two analogs that contain an azole in one position and an azoline in the other) exist for each of the microcyclamides and aerucyclamides. It was further determined that the relative yields of each analog were not equivalent, suggesting that directional preference exists for oxidation around the macrocycle. To determine the oxidative preference, an average MSMS spectrum was obtained for the two analogs of interest, which co-elute via LCMS and present identical molecular formulae. Distinct fingerprint ions corresponding to their respective analogs were quantified from these average MSMS spectra, and the relative yield of each analog was estimated based on the ion counts of these fingerprint ions. A characteristic fingerprint ion for the average MSMS spectrum of aerucyclamide A and its oxidative analog are shown in **Figure 8.3c**. Based on ion counts, it was determined that there was approximately 6.5% of fragment A compared to 93.5% of fragment B, which correspond to the aerucyclamide A oxidative analog and aerucyclamide A, respectively. This suggests that oxidative preference exists for heterocycle adjacent to the glycine residues compared to that flanked by the two isoleucine residues of aerucyclamide A.

The oxidative analogs of the other microcyclamide and aerucyclamide structures were also analyzed (**Figure 8.4**).

(a)



Directional Oxidation Analog Pair	Relative Ratios
1a : 1b	80 : 20
2a : 2b	90 : 10
3 : 3a	95 : 5
5 : 5a	100 : 0

(b)

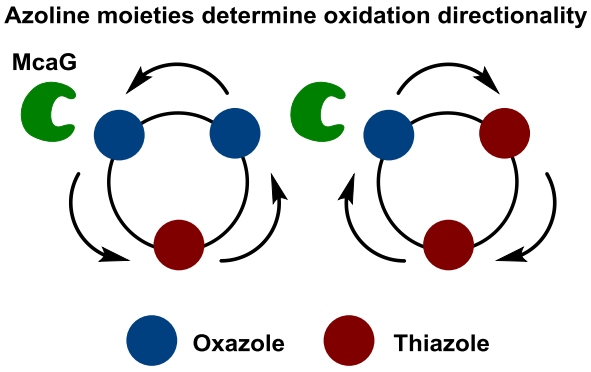


Figure 8.4 (a) Oxidative directional analogs and the relative yields observed from heterologous expression of the *mca* operon in *E. coli*. **(b)** Proposed mechanism for the directionality of azoline oxidation around the microcyclamide-aerucyclamide macrocycle.

Collectively, there are two trends that can be ascertained from the oxidative diversity of these microcyclamide and aerucyclamide analogs (**Figure 8.5**). First, oxidation of the thiazoline is preferred over that of the oxazoline, such as in the case of the microcyclamide 7806B analogs. Second, in structures containing two thiazol(in)e rings, the thiazoline adjacent to the glycine residue is preferentially oxidized.

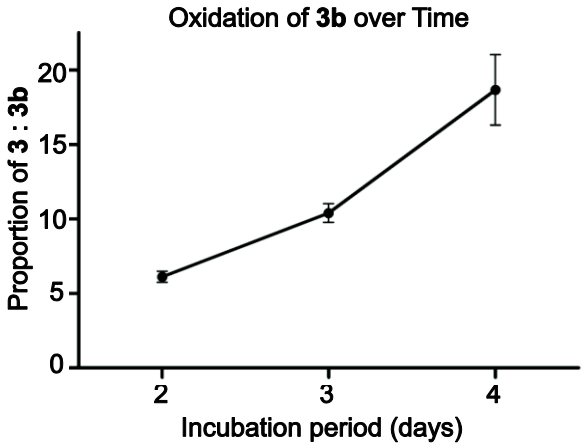


Figure 8.5 Proportion of oxidized analog pair **3** and **3b** observed from heterologous expression of the mca operon over a period of 4 days.

The generation of oxidative analogs of microcyclamide and aerucyclamide from the heterologous expression of the *mca* operon suggested that azoline oxidation is a late-stage transformation. Additional support for this claim arises from the observation of an aerucyclamide D analog that contains two thiazoline rings, which represents the reduced precursor to McaG oxidase activity to afford aerucyclamide D. To further support the hypothesis that McaG oxidase activity is a late-stage transformation, the relative production of aerucyclamide A and its reduced analog was analyzed from the heterologous *mca* operon expression over 2, 3, and 4 days (**Figure**

8.5). The quantity of aerucyclamide A relative to its reduced precursor was determined to be 6.12% ± 0.54, 10.40% ± 0.89, and 18.67% ± 3.34 after 2, 3, and 4 days, respectively. The increased amount of aerucyclamide A detected from culture extracts over time is evidence that McaG-mediate oxidation is a late-stage transformation.

8.5 DISCUSSION AND CONCLUSIONS

Cyanobactins are a promising class of natural products that have the potential to be developed into medicines. They are a continuously growing class of molecules, with nearly 30% of all sequenced strains of cyanobacteria harboring gene clusters that encode the biosynthetic machinery for their production [171, 240]. Understanding the mechanisms of their biosynthesis is important, as previous research findings have shown that these pathways are highly amenable to genetic engineering to generate novel combinatorial products [185, 221]. However, there exists a major challenge in investigating these molecules, as techniques to genetically manipulate cyanobacteria are lacking. To address this gap in technology, a designer plasmid and expression platform has been established. This study has utilized this platform to rapidly assemble and express the microcyclamide (*mca*) gene cluster from *M. aeruginosa* PCC 7806. In addition, selective and systematic assembly of *mca* genes has enabled the identification of the minimal pathway essential to the biosynthesis of the microcyclamides and aerucyclamides.

An unexpected discovery from the biosynthesis of the *mca* pathway was the generation of oxidatively diverse microcyclamides and aerucyclamides. Specifically, positional diversity of azoline ring oxidation to afford the azole was observed. Furthermore, two trends emerged from

the oxidative diversity of these microcyclamide and aerucyclamide analogs (**Figure 8.4**). First, oxidation of the thiazoline is preferred over that of the oxazoline, and second, in structures containing two thiazol(in)e rings, the thiazoline adjacent to the glycine residue is preferentially oxidized. The latter observation could arise from the structure of the active site being more accommodative to smaller amino acid side chains flanking the site of thiazoline oxidation. The preference for McaG-mediated oxidation of thiazoline over oxazoline could be a consequence of enhanced affinity of McaG to the thiazoline group over oxazoline. This is supported by the prevalence of cysteine residues presented in the hexapeptide sequences of the microcyclamides and aerucyclamides (ATVSIC, FTGCMC, ITGCIC). There is only one oxazole-forming residue (serine) of the six total azole-forming residues presented from the three distinct hexapeptides. There are also duplicate copies of the peptide sequence (ITGCIC) encoded in the *mcaE* gene that give rise to aerucyclamides A and B. If the predominance of thiazole-forming amino acids in the precursor peptide is an indication of enzymatic substrate selectivity, then the thiazoline group is likely the preferred substrate for the McaG oxidase. Schneider *et al*. demonstrated *in vitro* the capacity of a single oxidase domain from either the bleomycin or epothilone biosynthetic pathway to catalyze the formation of both oxazoline and thiazoline into their respective azoles [242]. While oxidase affinity was comparable for each azoline substrate, thiazole formation was nearly 10-fold faster over oxazoline formation. This was explained by the exclusive presence of thiazole moieties within the major congeners of bleomycin and epothilone biosynthesis, suggesting that thiazoline is the native substrate of these oxidases. In addition, thiazole formation is more thermodynamically favorable than oxazole formation [243], which may contribute to the increased presence of thiazole moieties in microcyclamide and aerucyclamide structures.

In addition to the directionality of azoline oxidation around the macrocycle, McaG-mediated azoline oxidation was determined to be a late-stage process. McaG is a PatG-like homolog, and PatG is composed of two distinct domains with the N-terminus domain responsible for oxidoreductase function, while the C-terminus domain is responsible for proteolysis of the precursor peptide [175, 221]. The sequence of PatG-mediated azoline oxidation and precursor peptide proteolysis is unclear. Time course experiments of the heterologous expression of the *mca* biosynthetic pathway revealed a relative increase in the production of aerucyclamide A, compared to its reduced precursor (**Figure 8.5**), over time. This suggests that azoline oxidation is a late-stage transformation that follows McaG-mediated proteolysis of the precursor peptide.

This investigation has demonstrated the utility of a novel designer plasmid to rapidly assemble and investigate cyanobacterial biosynthetic pathways. The biosynthetic gene cluster for the cyanobactin microcyclamide from *M. aeruginosa* PCC 7806 was studied using this platform. Heterologous expression of complete and mutant pathways established the genes essential for microcyclamides biosynthesis and also identified the McaG-mediated oxidation of azolines to azoles in microcyclamide structures to be a directional and late-stage transformation. This platform represents a genetic tool that can be utilized to characterize the biochemical mechanisms of other biosynthetic pathways from cyanobacteria.

8.6 PART II: OUTLOOK AND FUTURE DIRECTIONS

Traditional approaches to the discovery of natural products in the development of new drugs have proven limited in recent years. Notably, natural products discovery has hinged on the

development of compounds isolated via bioassay-guided techniques. The major drawback of this approach in microbial natural products discovery is the high rate of replication (i.e., "rediscovery" of known compounds)—in part, a result of horizontal gene transfer events that propagate secondary metabolite biosynthetic gene clusters amongst various microbial species. The consequence of this approach is the rapid decline in the rate of novel natural products discovered each year. A promising alternative relies on the utilization of genome sequencing information to identify orphan biosynthetic pathways predicted to produce novel natural products that may serve as scaffolds for next generation therapeutics.

Cyanobacteria are a promising source of natural products and emerging genomic sequencing data reveals that these organisms dedicate a significant portion of their genome in the biosynthesis of secondary metabolites; however, platforms to investigate their biosyntheses are limited. The advances from Chapters 5 through 8 have established a cloning and heterologous expression system that can rapidly assemble entire gene clusters to correlate biosynthetic operons to their chemical products. A hallmark feature of this platform is its ability to rapidly generate re-organized and mutant pathways. In this way, details into the investigation of biosynthetic mechanisms of cyanobacterial natural products could be explored, particularly in the case of the hapalosins and micro(aeru)cyclamides biosynthesis.

A major step forward in advancing this platform will entail efforts to optimize the expression and production of such secondary metabolites. This would facilitate the discovery of novel secondary metabolites from orphan gene clusters of cyanobacteria and accelerate current understanding of the biological and ecological significance of these molecules and how they can

be developed into medicines and other useful chemicals. However, it is worth noting that expression optimization can be a complex and challenging task, requiring precise control of numerous parameters (e.g., substrate availability, gene expression, growth conditions). Nonetheless, the significance of this source of natural products justifies the need to explore this treasure trove of compounds.

APPENDIX A

SUPPORTING INFORMATION FOR CHAPTER 2

A.1.1 Materials and General Procedures

Synthetic oligonucleotides for gene amplification by PCR were purchased from Life Technologies or Integrated DNA Technology. Kapa HiFi DNA polymerase was obtained from Kapa Biosystems. Restriction endonucleases and Antarctic phosphatase were purchased from New England BioLabs. LB broth and agar used for bacteria culture were obtained from Teknova. Antibiotics were obtained from Gold Biotechnology and Chem-Impex International. DNA sequencing was performed by Elim BioPharm Inc.

A.1.2 Bacteria strains and culture conditions

All plasmid cloning steps were implemented using *E. coli* TOP10 (Invitrogen) and XL1-Blue (Stratagene) strains. Yeast recombineering experiments were implemented using the yeast recombineering strain VL6-48N. *A. baumannii* strain ATCC17978 were used for all mutant generation and phenotypic analysis experiments. All bacteria strains were grown at 37°C in LB

broth with gentle shaking for aeration. Yeast were grown at 30°C in YPD media with gentle shaking. When needed, gentamicin and kanamycin were used at concentrations of 10µg/mL and 50µg/mL for *E. coli*, respectively, and at concentrations of 50µg/mL for *A. baumannii*.

A.1.3 General procedures for plasmid construction via yeast recombineering

The protocol for yeast-assisted recombination to generate plasmid constructs was adapted from a previously reported procedure. [74] DNA fragments of interest were introduced together with SmaI-digested parent plasmid (either pMQ30 for the suicide vector plasmid or the pAB1 for expression plasmids), the PCR amplicons were transformed into yeast strain VL6-48N. Following transformation, the yeast were plated in uracil-deficient SC agar plates and allowed to grow for 3 to 4 days. Following the growth period, all yeast colonies were collected and combined, then plasmid was isolated and transformed into *E. coli* TOP10 (Invitrogen). Transformed *E. coli* TOP10 were then selected on LB agar plates containing gentamicin. Individual colonies were grown in liquid culture and plasmid was isolated and subsequently analyzed via restriction enzyme digest analysis to obtain the desired plasmid construct.

A.1.4 Mutant Generation

The protocol used for kanamycin disruption mutant generation was adapted from a previously reported procedure. [73] The kanamycin resistance cassette was amplified from pCR®-Blunt II-TOPO® vector (Thermo Fisher Scientific). Briefly, primers catAUpFw and KMupintRV were used to amplify approximately the first 500bp of gene *catA*, and primer pair catADwnRV and KMdwintFW were used to amplify the last 500bp of gene *catA*. Additionally, primer pair

catAUpFw and catADwnRV were used to amplify the kanamycin resistance cassette from pCR-BluntII-TOPO vector (Invitrogen) as template. Primers FWnest and DWnest were used for nested OE-PCR using the following condition: 94°C for 15 s, 40°C for 1 min, 72°C for 2 min (10 cycles); 94°C for 15 s, 55°C for 1 min, 72°C for 3 min (20 cycles), and a final extension at 68°C for 10 min. 5µg of the overlap-extended PCR product was purified using isopropanol precipitation and subsequently transformed into electrocompetent *A. baumannii* ATCC17978. Transformed cells were recovered in SOC for 1hr at 37°C and plated on LB agar containing kanamycin. The following day, kanamycin-resistant clones were verified for correct allele-exchange by PCR using primers catAUpFW and KmRV as well as catADwnRV and KmFW. Clones generating PCR products consistent with the correct kanamycin mutant allele were further verified by DNA sequencing.

The protocol used to generate markerless in-frame deletion mutants was adapted from a previously reported procedure. [244] The suicide plasmid vector to delete the *catA* gene was constructed using the yeast-assisted recombination protocol described above. Primers catAKO1FW and catAKO1RV were used to amplify approximately 500bp upstream of gene *catA*, and primers catAKO2FW and catAKO2RV were used to amplify approximately 500bp downstream of gene *catA*. These were introduced into the SmaI restriction site of pMQ30 and introduced to yeast. The resulting mixture of plasmid constructs was isolated from yeast and introduced to *E. coli* TOP10. Desired plasmid constructs were screened from *E. coli* TOP10 clones via restriction enzyme digest analysis. The desired plasmid construct to delete the *catA* gene was transformed into electrocompetent wildtype *A. baumannii* ATCC17978 cells, allowed to recover in SOC media at 37°C for 1hr, and plated on LB agar plates containing gentamicin.

The following day, single colonies were re-streaked on LB agar plates containing gentamicin to ensure stable integration of the suicide vector. Stable gentamicin-resistant clones were then allowed to grow on LB agar plates without antibiotic at 25°C for 2 days to allow loss of the suicide plasmid. Single clones were then streaked on LB agar plate containing sucrose (5% (w/v)) and no NaCl. Sucrose-resistant clones were screened by PCR (using primers catAKOValidationFW and catAKOValidationRV) for the desired in-frame deletion of *catA*.

A.1.5 Construction of *A. baumannii* Expression Plasmids

The *A. baumannii-E. coli-S. cerevisiae* shuttle vector and derivatives thereof were constructed via yeast recombineering. Primers WH1266ORIFW and WH1266ORIRV were used to clone the origin of replication sequence from an *A. baumannii* replicable plasmid pWH1266 [245]. The PCR amplicon containing the *A. baumannii* replicon was introduced into the AatII and SfiI site of plasmid pMQ124, yeast recombineered, and the desired plasmid pXLAB1 was obtained as described above.

The LacZ reporter plasmid was constructed via yeast recombineering. The full length *lacZ* gene was cloned via PCR using primers LacZFW and LacZRV from a pMQ131-derived plasmid containing the full length *lacZ* gene [203]. The amplicon was introduced into the SmaI site of plasmid pXLAB1 and the desired construct pXLAB2 was obtained by screening via restriction enzyme digest analysis.

The *abaR*, *catA*, and *catABCDEFGH* with *sfp* expression plasmids were all constructed via yeast recombineering. The desired amplicons were all introduced into the SmaI site of

plasmid pXLAB1 to generate their corresponding construct. Primers abaRFW and abaRRV were used for amplification of *abaR*. catAFW and catARV were used for amplification of gene *catA*. The full *cat* pathway was assembled by amplifying five 3.5kb fragments with approximately 500bp overlapping regions between each adjacent fragment. The *sfp* gene was also PCR amplified and co-introduced immediately downstream of the P_{BAD} promoter. Primers sfp_RBS_FW and sfp_RBS_RV were used to clone the *sfp* gene. Fragments 1 through 5 were PCR amplified using the following pairs of primers, respectively: catFULL1FW/catFULL1RV, catFULL2FW/catFULL2RV, catFULL3FW/catFULL3RV, catFULL4FW/catFULL4RV, catFULL5FW/catFULL5RV.

A.1.6 LacZ reporter assays

A. baumannii overnight liquid culture (0.1mL) supplemented with arabinose (1% (w/v)) was added to buffer Z (0.9mL of 113mM Na_2HPO_4, 40mM NaH_2PO_4-H_2O, 10mM KCl, 1mM $MgSO_4$-$7H_2O$, 0.0028% (v/v) β-mercaptoethanol, pH 7.0 per 500mL H_2O), sodium dodecyl sulfate (20µL, 0.1% (w/v)), and chloroform (50µL). After adding all components, the mixture was vortexed at room temperature (2 min). Following vortexing, *o*-nitrophenyl-β-D-galactoside (0.2mL, 4 mg/mL in water) was added and allowed to incubate at room temperature (30 min). After incubation, Na_2CO_3 (0.5mL, 1M) was added. The reaction mixture was vortexed briefly, left at room temperature (5min), and diluted with an equal volume of buffer Z. The absorbance of the reaction mixture was then measured at 420nm. β-galactosidase activity (A_{420nm}) was normalized to the OD_{600} measured from the *A. baumannii* overnight liquid culture.

A.1.7 Motility Assays

The procedure for executing motility assays was adapted from a previously reported protocol. [65] Briefly, agar plates were prepared using BD Difco Nobel agar (Sparks, M.D., U.S.) at 0.3% (w/v) concentration in water containing 1% (w/v) tryptone. When necessary, gentamicin (25µg/mL), arabinose (1% (w/v/)), and 3-OH C_{12}-HSL (250nM) were supplemented. Plates were inoculated by dipping a flat-ended wooden stick into a culture grown overnight and gently placed on the surface of the agar. Plates were incubated for 24 to 48hr at 37°C and observed for motility.

A.1.8 Horse blood agar hemolysis and colony morphology

5% (v/v) of defibrinated horse blood agar plates were obtained pre-prepared from Benton-Dickinson (BD) company (Franklin Lakes, N.J., U.S.). To assess hemolysis, 3µL of overnight liquid culture of *A. baumannii* was pipetted onto horse blood agar plates and incubated at 37°C for 48 to 72hr. Both horse blood hemolysis and colony morphology were observed following incubation.

A.1.9 Desiccation Resistance

This assay was adapted from a previously reported procedure. [246] Overnight liquid cultures of *A. baumannii* were diluted 100-fold and allowed to grow to OD_{600} = 0.5. This liquid culture was diluted further 10- for the desiccation assay. 1µL of each diluted culture was spotted onto filter paper in a sterile petri dish, then incubated at 37°C for 24hrs. Desiccated *A. baumannii* inoculum

were then recovered on tryptic soy agar plates and allowed to grow overnight. Desiccation resistance and sensitivity phenotypes were then observed after overnight growth.

A.1.10 Wax Moth *G. mellonella* Cytotoxicity

A. baumannii virulence against the wax moth larvae *G. mellonella* was executed using a procedure previously reported. [89] In this procedure, overnight liquid culture of *A. baumannii* were serially diluted by 10-fold and resuspended in PBS buffer. 10μL volume inoculum of each diluent was introduced to the wax worm larvae and allowed to incubate at 37°C for up to 96hr. After each 24hr period, the number of surviving larvae was quantified to assess virulence of each *A. baumannii* strain against *G. mellonella*. The percent survival was calculated against a negative control *G. mellonella* group subjected to PBS buffer only.

A.1.11 Heat Shock Viability

A previously reported procedure was adapted for these experiments. [246] Overnight liquid culture of *A. baumannii* strains were diluted 100-fold in LB media and allowed to grow to OD_{600} = 1.0. When the appropriate cell density was reached, *A. baumannii* liquid culture was heated to 55°C in a heat block for 10 minutes; the number of viability cells was counted before heating and after heating for 10 minutes. Viable cells were counted by serial dilution of the heated culture by 10-fold increments and plating the dilutions onto tryptic soy agar plates. The percent viability was normalized to the initial number of viable cells (i.e., without heating).

A.1.12 Reverse Transcriptase (RT) PCR Analysis

To obtain cells for RNA extraction, cultures were first inoculated into 3 mL of LB medium. The cultures were grown overnight in LB medium, subcultured to OD600 = 0.05, grown to OD600 = 0.6. RNA was isolated from A. baumannii cells using an RNeasy kit (Qiagen). RNA was treated first with an on column DNase I treatment (Qiagen), followed by 2 U of RQ1 DNase (Promega) for 30 min at 37°C, followed by a second purification using a Zymo RNA cleanup kit. RNA was normalized to 100 ng/µl using DNase-free water and 5 µl was used in each reverse transcriptase (RT) reaction to generate cDNA using Superscript III RT (Invitrogen) following the manufacturers specifications. A PCR reaction with a 94°C hold for 5 minutes, followed by 28 cycles of 30 seconds at 94°C, 30 seconds at 55°C, and 30 seconds at 72°C, followed by a 72°C hold for 7 minutes was used to detect the amount of transcript. The 16S rRNA gene was used as a control.

A.1.13 Generation and Verification of Δ*cat*A::Km and Δ*cat*A mutants

To verify the correct generation of mutants, colony PCR was implemented to yield PCR products consistent with the expected mutant region. To screen for correct Δ*cat*A::Km mutants, two sets of primers were designed. For each pair, one anneals to the kanamycin resistance cassette and the other anneals to the chromosome of the *A. baumannii* strain, yielding two PCR products of approximately 1.5kb in size. The relative locations of these primers are denoted by the blue and red arrows in **Figure A1**. Therefore, the wildtype locus does not yield any PCR products with this primer design, while two PCR products can be obtained from the mutant.

The in-frame deletion mutant was screened using PCR primers complementary to regions flanking the deleted gene. The relative location of these primers are indicated by the purple arrows in **Figure A2**. Therefore, the wild type locus yields a larger PCR product of approximately 2.5kb, while the in-frame deleted gene yields a PCR product of approximately 1.0kb. All mutants

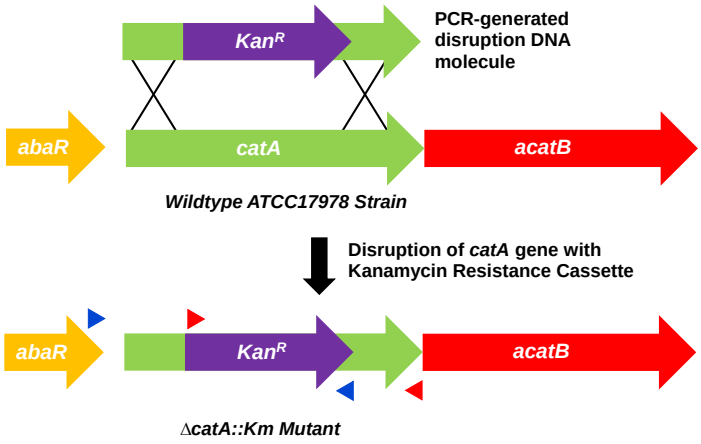


Figure A1 Disruption mutant Δ*catA*::Km was generated by introduction of a kanamycin resistance cassette to replace the internal sequences of the *catA* gene.

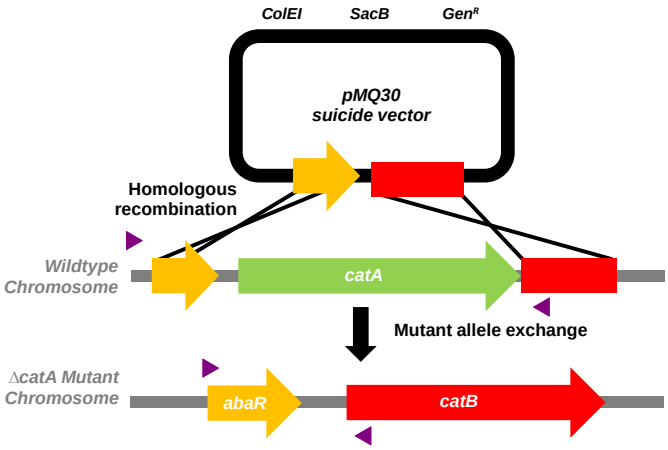
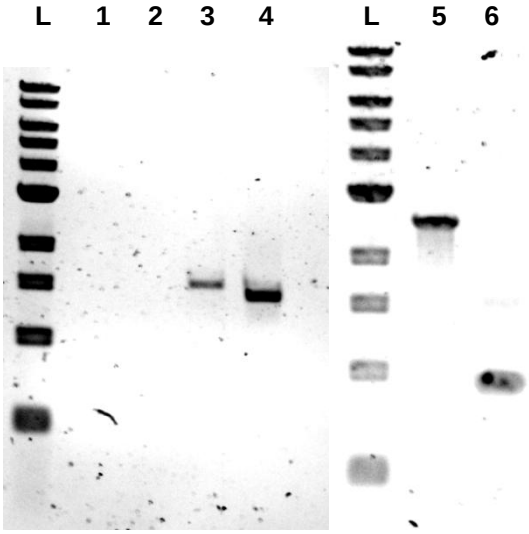
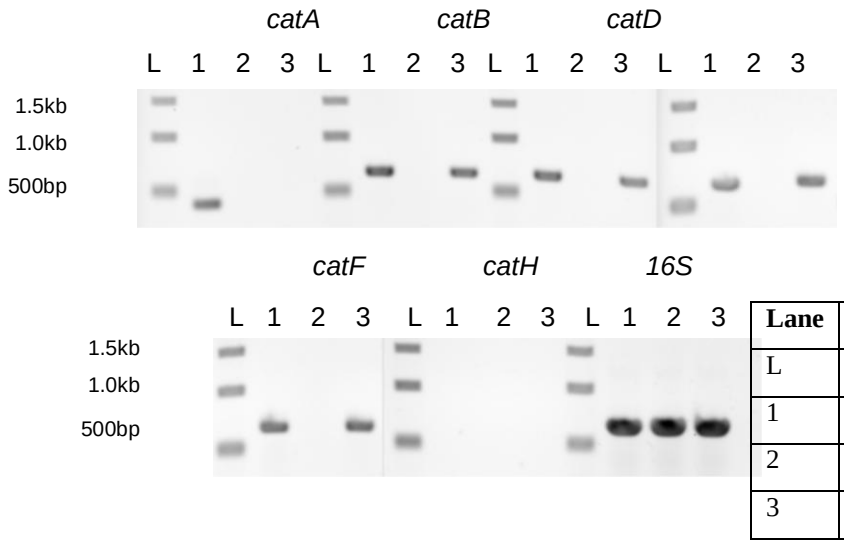


Figure A2 Markerless in-frame deletion mutant ΔcatA. The suicide plasmid vector contains an origin of replication for *E. coli* (ColE1), a counterselectable *sacB* gene which confers toxicity to its host in the presence of sucrose, and a selectable *aacC1* gene.



Lane	Genomic DNA template	Expected PCR product
1	ATCC17978 WT	None
2	ATCC17978 WT	None
3	Δ*catA*::Km	1548 bp
4	Δ*catA*::Km	1376 bp
5	ATCC17978 WT	2525 bp
6	Δ*catA*	914 bp

Figure A3 PCR products obtained from colony PCR of Δ*catA*::Km and Δ*catA* mutants. The expected PCR products for each mutant are listed on the left. Lanes "L" contained the reference DNA ladder.



Lane	Sample
L	DNA Ladder
1	ATCC17978 WT
2	Δ*catA*::Km
3	Δ*catA*

Figure A4 RT-PCR analysis of wildtype and mutants Δ*catA* and Δ*catA*::Km. Transcription of genes *catA*, *catB, catD, catE, catF, catH,* and the housekeeping (control) 16S rRNA gene were monitored for each strain. PCR amplicons for each gene are approximately 500bp in size.

Table A1 Plasmids used in this study

Plasmid	Major features	Reference
pMQ124	$_{ori}$ColE1/pRO1600, *aacC1*, P_{BAD}-*lacZα*, *CEN6*	[74]
pXLAB1	pMQ124[74] derivative containing *ori* from *A. calcoaceticus*	This work, [245]
pXLAB2	pAB1 containing full length *LacZ* gene under P_{BAD} control	This work
pXLAB3	Suicide vector; contains approx. 500bp flanking regions of *catA* gene.	This work
pXLAB4	pAB1 containing *catA* gene at SmaI position	This work
pXLAB5	pAB1 containing *abaR* gene at SmaI position	This work
pXLAB6	pAB1 containing *sfp* and *catABCDEFGH*	This work

Table A2 Primers used in this study

Primer name	Sequence	Application
catAUpFW	CGAACCTGGGTTGGACTTTA	5' primer to clone first 500bp of *catA*
KMupintRV	CCCAGCTGGCAATTCCGGTGATGCGATGTAGGCACAAT	5' primer to clone kanamycin resistance cassette for OE-PCR
catADwnRV	CGATCACGGATGACGTCTAA	3' primer to clone last 500bp of *catA*
KMdwintFW	CTTGACGAGTTCTTCTGAACGGGACAACCTAGCCAAGTA	3' primer to clone kanamycin resistance cassette for OE-PCR
FWnest	TCGTTCTTTGCCTGTCTGAA	5' primer for nested OE-PCR
DWnest	ACCGTTCTACGGTTTGGATG	3' primer for nested OE-PCR
KmFW	CCGGAATTGCCAGCTGGG	5' primer to clone kanamycin resistance
KmRV	TTCAGAAGAACTCGTCAAG	3' primer to clone kanamycin resistance
catAKO1FW	TTGCATGCCTGCAGGTCGACTCTAGA	5' primer to clone 500bp upstream *catA*
catAKO1RV	TGATCATTGAGTTTTCGAACAGCAATA	3' primer to clone 500bp upstream *catA*
catAKO2FW	ATTGCTGTTCGAAAACTCAATGATCAA	5' primer to clone 500bp downstream catA
catAKO2RV	AGCTATGACCATGATTACGAATTCGA	3' primer to clone 500bp downstream *catA*
WH1266ORIFW	GGACGAGGCAAGCTAAACAGATCTCTAGACCTAGATCGTAGAAATATCTATGATTATC	5' primer to clone *Acinetobacter* ori of pWH1266
WH1266ORIRV	CGCCCTTCCCAACAGTTGCGCAGCCTGAAAGGCAGGATTTTAACATTTTGCGTTGTTCC	3' primer to clone *Acinetobacter* ori of pWH1266

LacZFW	ACTCTCTACTGTTTCTCCATACCCGTAGGAGGAAAAAATGATAGATCCCGTCGTTTTAC	5’ primer to clone full length *lacZ*
LacZRV	TATCAGACCGCTTCTGCGTTCTGATTTAATCTGTATCATTACTTTTGACACCAGACCAACTGG	3’ primer to clone full length *lacZ*
abaRFW	TACTGTTTCTCCATACCCGTAGGAGGAAAAAATGGAAAGTTGGCAAGAAGATTTATTATC	5’ primer to clone *abaR*
abaRRV	CAAATTCTGTTTTATCAGACCGCTTCTGCGTTCTGATCTACAAAAGCCCTAGCATTAC	3’ primer to clone *abaR*
catAKOValidationFW	TGCGCGTATCTAAATTAAG	5’ primer to validate in-frame deletion of *catA*
catAKOValidationRV	TCACAGACCAAGCATTCGT	3’ primer to validate in-frame deletion of *catA*
sfp_RBS_FW	CTCTCTACTGTTTCTCCATACCCGTAGGA	5’ primer to clone *sfp* gene and introduce RBS
sfp_RBS_RV	TTATATCCTCCTACGGGTATGGAGAAT	3’ primer to clone *sfp* gene and introduce RBS
catFULL1FW	TTCTCCATACCCGTAGGAGGATATAAATG	5’ primer to clone first 3.5kb fragment of the full *cat* gene cluster
catFULL1RV	CGGCATAGTCACGGGCGA	3’ primer to clone first 3.5kb fragment of the full *cat* gene cluster
catFULL2FW	TCGCTTGCCAAAATCACATCC	5’ primer to clone second 3.5kb fragment of the full *cat* gene cluster
catFULL2RV	TTCGCCAGTGGTTCACCTAAA	3’ primer to clone second 3.5kb fragment of the full *cat* gene cluster
catFULL3FW	ATGATAGATTGCAAGTAACGAC	3’ primer to clone third 3.5kb fragment of the full *cat* gene cluster
catFULL3RV	ACACCATAACCGGACCCACACG	3’ primer to clone third 3.5kb fragment of the full *cat* gene cluster
catFULL4FW	CGTAGAAAGCGAACAATCTAA	3’ primer to clone fourth 3.5kb fragment of the full *cat* gene cluster
catFULL4RV	TACACGCCTACTAAACTCAGC	3’ primer to clone fourth 3.5kb fragment of the full *cat* gene cluster
catFULL5FW	GATTCCATTGCGTCAGATTGCT	3’ primer to clone fifth 3.5kb fragment of

		the full *cat* gene cluster
catFULL5RV	TATCAGACCGCTTCTGCGTTCTGATTTAATCTGTATCACTCGTTTCCCGCTCCAGATT	3’ primer to clone fifth 3.5kb fragment of the full *cat* gene cluster
catAFW	ACTCTCTACTGTTTCTCCATACCCGTAGGAGGAAAAAATGTGTACGGATAATATCATTG	5’ primer to clone *catA*
catARV	CAAATTCTGTTTTATCAGACCGCTTCTGCGTTCTGATTTAAAACAAAGCACTTTTCAG	3’ primer to clone *catA*
RTcatAFW	GTGCGTGTGCTGTTTAATG	5’ primer for RT-PCR of *catA*
RTcatARV	CTTCATTAAGGATCGTTGCTTC	3’ primer for RT-PCR of *catA*
RTcatBFW	AGTACATGGTCACCAATCGTCAA	5’ primer for RT-PCR of *catB*
RTcatBRV	TCAGCCATGCACCATTCTGTAAA	3’ primer for RT-PCR of *catB*
RTcatDFW	ATTTCCTGAGCCGATCTTTGA	5’ primer for RT-PCR of *catD*
RTcatDRV	CTTCAGGTACATTTTCTGGAATT	3’ primer for RT-PCR of *catD*
RTcatEFW	AGTGTATTTGCAACCTTAATTATG	5’ primer for RT-PCR of *catE*
RTcatERV	AGTATGCACATACGACGTTTG	3’ primer for RT-PCR of *catE*
RTcatFFW	CGCGTAGTCGTTCAGATGTGATTCC	5’ primer for RT-PCR of *catF*
RTcatFRV	CGATCTCCGCCTGCATTCAACT	3’ primer for RT-PCR of *catF*
RTcatHFW	AAGATGGGCATCACAGACCTG	5’ primer for RT-PCR of *catH*
RTcatHRV	GCTTAGACACTGACATGCGGC	3’ primer for RT-PCR of *catH*
RT16SFW	ATGAAGTCTGTTTTATCAGATAAG	5’ primer for RT-PCR of 16S rRNA
RT16SRV	TTAGTTATGTTCGCAAGGTTCT	3’ primer for RT-PCR of 16S rRNA

APPENDIX B

SUPPORTING INFORMATION FOR CHAPTER 3

B.1.1 General Materials

Synthetic oligonucleotides for gene amplification by PCR were purchased from Life Technologies or Integrated DNA Technology. Kapa Robust 2G DNA polymerase was obtained from Kapa Biosystems. Restriction endonucleases and Antarctic phosphatase were purchased from New England BioLabs. LB broth and agar used for bacteria culture were obtained from Teknova. Antibiotics were obtained from Gold Biotechnology and Chem-Impex International. Polyvinyl chloride plastic 96-well plates were purchased from Corning. DNA sequencing was performed by Elim BioPharm Inc. *P. aeruginosa* PAO1 laboratry strain and plasmids pMQ124 and pMQ30 were generously donated by the research laboratory of Professor Robert Shanks of the Ear and Eye Institute at the University of Pittsburgh Medical Center.

B.1.2 Bacteria strains and culture conditions

All plasmid cloning steps were implemented using *E. coli* TOP10 (Invitrogen). Yeast recombineering experiments were implemented using the yeast recombineering strain VL6-48N. The common laboratory strain *P. aeruginosa* PAO1 was used for all mutant generation and phenotypic analysis experiments. All bacteria strains were grown at 37°C in LB broth with gentle shaking for aeration. Yeast were grown at 30°C in YPD media with gentle shaking. When needed, gentamicin was used at concentrations of 10µg/mL for *E. coli* and at concentrations of 50µg/mL for *P. aeruginosa*.

B.1.3 General procedures for plasmid construction via yeast recombineering

DNA fragments of interest (generated by PCR amplification) were introduced together with SmaI-digested parent plasmid (either pMQ30 for the suicide vector plasmid or pMQ124 for expression plasmids), the PCR amplicons were transformed into yeast strain VL6-48N. Following transformation, the yeast were plated in uracil-deficient SC agar plates and allowed to grow for 3 to 4 days. Following the growth period, all yeast colonies were collected in 1mL of TE buffer, then plasmid was isolated and transformed into *E. coli* TOP10. Transformed *E. coli* TOP10 were then selected on LB agar plates containing gentamicin. Individual colonies were grown in liquid culture and plasmid was isolated and subsequently analyzed via restriction enzyme digest analysis to obtain the desired plasmid construct.

B.1.4 Δ*cpt*AB mutant generation

The suicide plasmid vector to delete the *cpt*AB genes was constructed using the yeast-assisted recombination protocol described above. Primers cptABKO1FW and cptABKO1RV were used to amplify approximately 500bp upstream of genes cptAB, and primers cptABKO2FW and cptABKO2RV were used to amplify approximately 500bp downstream of genes cptAB. These were introduced into the SmaI restriction site of pMQ30 and transformed into yeast. The resulting mixture of plasmid constructs was isolated from yeast and introduced to E. coli TOP10. Desired plasmid constructs were screened from *E. coli* TOP10 clones via restriction enzyme digest analysis.

To prepare electrocompetent cells, 6mL of wildtype PAO1 was grown in LB overnight at 37°C. The following day, the overnight culture was washed three times with sterile sucrose solution (500mM) at room temperature. Approximately 1μg of the suicide vector was mixed with the electrocompetent cells and allowed to recover in LB media at 37°C for 1hr. The transformed cells were plated on LB agar plates containing gentamicin and allowed to grow overnight at 37°C.

The following day, stable gentamicin-resistant clones were then allowed to grow on LB agar plates without antibiotic at 25°C for 2 days to allow loss of the suicide plasmid. Single clones were then streaked on LB agar plate containing sucrose (5% (w/v)) and no NaCl. Sucrose-resistant clones were screened by PCR (using primers cptABKOValidationFW and cptABKOValidationRV) for the desired in-frame deletion of catA. The wildtype PAO1 yielded an amplicon that was 4780bp in size, and the Δ*cpt*AB mutant yielded an amplicon that was

1912bp in size (**Figure B1**). Mutant clones identified by colony PCR were sequence verified using primers cptABKOValidationFW / cptABKOValidationRV and cptABKOValidation2FW / cptABKOValidation2RV.

B.1.5 Construction of the cptAB complementation plasmid

The *cpt*AB complementation plasmid was constructed via yeast recombineering, as described above. Primers cptAB1into124_FW / cptAB1into124_RV and cptAB2into124_FW / cptAB2into124_RV were used to amplify the two fragments that contained the entire *cpt*AB locus. Primers were designed such that fragments recombining with the backbone of the plasmid overlap by approximately 60 nucleotides. PCR amplicons that recombine with each other overlap with one another by approximately 100 nucleotides. Restriction enzyme digest verification of this plasmid is shown in **Figure B2**.

B.1.6 Twitching motility

The bacteria were stabbed through an agar layer of LB (1% agar) to the bottom of the Petri dish and incubated for 48hr at 37°C and then left at 25°C for 24hr. Following incubation, the agar was carefully cut around the edges of the plate and removed. After removal of the agar, crystal violet (1% (w/v), ethanol solution) was used to stain attached cells and the radius was determined for the wildtype and mutant strains. All motility assays were performed in triplicate.

B.1.7 Colony Morphology

The colony morphology of *P. aeruginosa* strains were observed from the twitching motility LB agar plates that were prepared above. After incubation of the stab-inoculated LB agar plates at 37°C for 48hr, the colony morphologies were observed for autolysis.

B.1.8 Abiotic Biofilm Formation

Overnight cultures of *P. aeruginosa* were diluted 1000-fold in LB media containing the appropriate antibiotics. 100µL volume aliquots of the diluted inoculum were pipetted into wells of a polyvinyl chloride 96-well plate. The plate was then covered with a sterile Kimwipe and placed in a sealed bag containing a damp paper towel. The plates were then incubated at 30°C for 18hr under static conditions.

The following day, bacterial growth was assessed by measuring the optical density (OD) at 600nm using a plate reader (Molecular Devices SpectraMax 340). The overnight bacteria cultures from the 96-well plate were decanted and washed 10 times with nanopure water to ensure all planktonic bacteria were discarded from the wells. Then, 125µL of 1% (w/v) crystal violet in ethanol solution was introduced into each well and allowed to incubate for 20-30 minutes at room temperature under static conditions. Following the CV stain, the wells were decanted of all stain and washed 10 times with nanopure water to ensure all CV not associated with biofilm was removed from the wells. Following the post-CV stain wash, 100µL of aqueous 1% (w/v) SDS was pipetted into each well and allowed to shake gently for 20-30 minutes at room temperature, until all biofilm-associated CV was solubilized. CV concentration was then

assessed by measuring the absorbance of each well at 540nm. Biofilm formation was compared between each *P. aeruginosa* sample by normalizing the CV response (A540nm) by bacteria growth (OD600nm).

Table B1 Primers used in this study

Primer name	**Sequence**	**Application**
cptABKO1_FW	ccagtgccaagcttgcatgcctgcaggtcgactctagaggatgttctgatcagcggccag	Clone ~500bp flanking sequence upstream of *cptAB*
cptABKO1_RV	gattctgttcttggcttattcgtgaatgcctgtcgcttcgtcgattatcggaaccagccg	Clone ~500bp flanking sequence upstream of *cptAB*
cptABKO2_FW	ggagtctcagccatgtcgaaagattctgttcttggcttattcgtgaatgcctgtcgcttc	Clone ~500bp flanking sequence downstream of *cptAB*
cptABKO2_RV	cacaggaaacagctatgaccatgattacgaattcgagctctgagtcagtgcggttctttg	Clone ~500bp flanking sequence downstream of *cptAB*
cptAB1into124_FW	actctctactgtttctccatacccgtaggaggaaaaa_atgtcgaaagattctgttc	Clone fragment #1 of *cptAB* into pMQ124
cptAB1into124_RV	acaatgcaggtcggcctc	Clone fragment #1 of *cptAB* into pMQ124
cptAB2into124_FW	atttcagccgccccacgg	Clone fragment #2 of *cptAB* into pMQ124
cptAB2into124_RV	tatcagaccgcttctgcgttctgatttaatctgta_tcaccagggcagcggctgg	Clone fragment #2 of *cptAB* into pMQ124
cptABKOValidationFW	tccgccatagtcgtattcccag	Validate mutant generation
cptABKOValidationRV	tattgccggatgccctgacg	Validate mutant generation
cptABKOValidation2FW	ccccgctcgatacgctgaaaaac	Validate mutant generation
cptABKOValidation2RV	agaccacaggtttcctggtaacg	Validate mutant generation

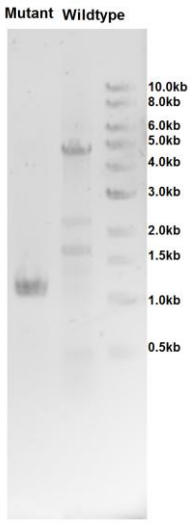


Figure B1 Colony PCR screening of identified Δ*cpt*AB mutant (left lane), compared to PAO1 wildtype. Primers used for colony PCR: cptABKOValidationFW and cptABKOValidationRV

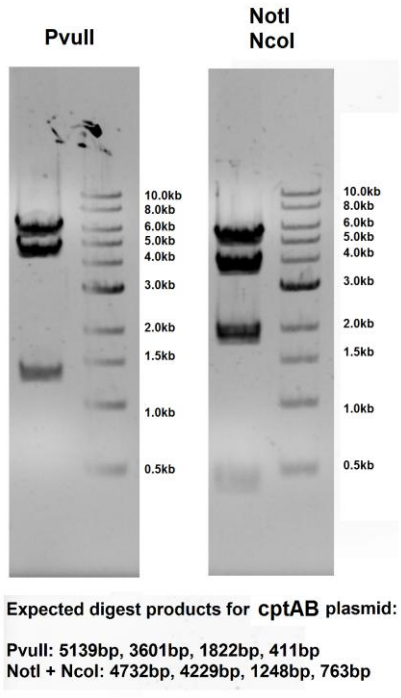


Figure B2 Restriction enzyme digest analysis for cptAB plasmid. Either PvuII (left) or NotI and NcoI (right) were used to verify the assembly of these expression plasmids.

APPENDIX C

SUPPORTING INFORMATION FOR CHAPTER 5

C.1.1 General methods and materials

PCRs were executed using a C1000 thermal cycler (Bio-Rad). HRMS analysis was conducted using a Q Exactive Benchtop Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific) equipped with a Luna® 5 µm C18(2), 100 Å, 250 x 4.6 mm LC Column (Phenomenex). All MS experiments utilized electrospray ionization modes for analysis. Synthetic oligonucleotides for gene amplification by PCR were purchased from Life Technologies or Integrated DNA Technology. Kapa HiFi DNA polymerase was obtained from Kapa Biosystems. Plasmid DNA purification was executed using column purification kits obtained from either SydLabs or Qiagen. Restriction endonucleases and Antarctic phosphatase were purchased from New England BioLabs. LB broth and agar used for culturing *E. coli* were obtained from Teknova. All *E. coli* strains were cultured in LB media supplemented with antibiotics kanamycin (50 µg/mL), gentamicin (10 µg/mL), and tetracycline (10 µg/mL), when necessary. Biosynthetic precursor substrates were introduced to cultures at the time of induction at the following final concentrations (when necessary): 5mM for halogenated phenylalanine substrates, 1mM for homo-L-tyrosine, and 1mM mevalonate. Homo-L-tyrosine was purchased from Combi-Blocks.

p-Chlorophenylalanine and *p*-fluorophenylalanine were purchased from Chem-Impex International. All cyanobacterial genomic DNA were acquired from Professor David Fewer's research laboratory (University of Helsinki). Plasmid pMBI was acquired from Professor Jay Keasling's research laboratory (University of California, Berkeley). LaserGene SeqBuilder (DNAStar) software was used for all *in silico* plasmid DNA sequence manipulation.

Table C1 Primers used in this study

Primer name	**Sequence**	**Purpose**
age_frag1_5'_toMQ123i	cgagctcggtacccggggaaggagatatacatatgctagatagattac atac	1-step cloning of *age* pathway (fragment #1)
age_frag1_3'	ctgtttattaatcttggcg	1-step cloning of *age* pathway (fragment #1)
age_frag2_5'	ctttacggcagcctgtaacc	1-step cloning of *age* pathway (fragment #2)
age_frag2_3'	gagggctggcattgtcatc	1-step cloning of *age* pathway (fragment #2)
age_frag3_5'	ctctatatatcacaggtcttttg	1-step cloning of *age* pathway (fragment #3)
age_frag3_3'	gttcactgagttgttctatg	1-step cloning of *age* pathway (fragment #3)
age_frag4_5'	ttgaacagtggtctgactttcc	1-step cloning of *age* pathway (fragment #4)
age_frag4_3'_toMQ123i	gttttatcagaccgcttctgcgttctgatttatcaggtttgttcagaacgcg	1-step cloning of *age* pathway (fragment #4)
kaw_frag1_5'_toMQ123i	cgagctcggtacccggggaaggagatatacatATGTCAGAC ATAACGGAAATACTG	1-step cloning of *kaw* pathway (fragment #1)
kaw_frag1_3'	TTACTTCAGCAAAGAGCCACG	1-step cloning of *kaw* pathway (fragment #1)
kaw_frag2_5'	CTGGCGGTCGATGTAGTTC	1-step cloning of *kaw* pathway (fragment #2)
kaw_frag2_3'	CTGTTCGGGGATGCACAC	1-step cloning of *kaw* pathway (fragment #2)
kaw_frag3_5'	GGAATTTCAGCAAACACAAGTTTG	1-step cloning of *kaw* pathway (fragment #3)
kaw_frag3_3'	gttttatcagaccgcttctgcgttctgatttaattagttgggagaatccca ccaag	1-step cloning of *kaw* pathway (fragment #3)
123i_spu_Fg1_FW	agctcggtacccggggaaggagatatacatatgcagttagacaacca tagtc	1-step cloning of *spu* pathway (fragment #1)
123i_spu_Fg1_RV	atcatcccacatactccaac	1-step cloning of *spu* pathway (fragment #1)
123i_spu_Fg2_FW	CATCAATTAGCAAAAGAAAATCAG	1-step cloning of *spu* pathway (fragment #2)
123i_spu_Fg2_RV	CAGCAGTAGTTTCGGGACG	1-step cloning of *spu* pathway (fragment #2)
123i_spu_Fg3_FW	aatccacggtgacggctac	1-step cloning of *spu* pathway

		(fragment #3)
123i_spu_Fg3_RV	catcaaatcggctaatatcatag	1-step cloning of *spu* pathway (fragment #3)
123i_spu_Fg4_FW	ccagaggaacgtctttatttg	1-step cloning of *spu* pathway (fragment #4)
123i_spu_Fg4_RV	ctcgatggtgctatggtc	1-step cloning of *spu* pathway (fragment #4)
123i_spu_Fg5_FW	cctctattccaagtcatgtttaac	1-step cloning of *spu* pathway (fragment #5)
123i_spu_Fg5_RV	tcagaccgcttctgcgttctgatttattagttagtcctattaccaaatac	1-step cloning of *spu* pathway (fragment #5)
apnA_fg1_FW_into_123i	cctctagattctccatacAGGAGGAATAATatgttcaagcaa tctatccac	2-step cloning of *apn* pathway (2^{nd} step)
apnA_fg1_RV	ttgacttgttcttcaaagag	2-step cloning of *apn* pathway (2^{nd} step)
apnA_fg2_FW	gaaggaattacgatggtttg	2-step cloning of *apn* pathway (2^{nd} step)
apnA_fg2_RV	tgactgaccatttgcagagg	2-step cloning of *apn* pathway (2^{nd} step)
apnB_FW	gaagatatttattctctctc	2-step cloning of *apn* pathway (2^{nd} step)
apnB_RV_into_apnC	atgtatatctccttccccgggtaccgagtcataaaaataattgttgtg	2-step cloning of *apn* pathway (2^{nd} step)
apnCD_fg1_FW_into_123i	cgaattcgagctcggtacccggggaaggagatatacatATGAGT GATCTGCTAAAACGTTTAG	2-step cloning of *apn* pathway (1^{st} step)
apnCD_fg1_RV	3' AAGTTATCCCGACGACCCAC	2-step cloning of *apn* pathway (1^{st} step)
apnCD_fg2_FW	CATGATTCCAAGTCAGATTATGATC	2-step cloning of *apn* pathway (1^{st} step)
apnCD_fg2_RV	ACTGGGAACAATATAGGCGAC	2-step cloning of *apn* pathway (1^{st} step)
apnCD_fg3_FW	AGGAGACTTAGCTAAATGGCGAAC	2-step cloning of *apn* pathway (1^{st} step)
apnCD_fg3_RV	CTTCTAAACAAGTTTGACGAGTTTGC	2-step cloning of *apn* pathway (1^{st} step)
apnCD_fg4_FW	GGCAGTTTTTAACCTTTTACTCTCCC	2-step cloning of *apn* pathway (1^{st} step)
apnCD_fg4_RV_into123i	tatcagaccgcttctgcgttctgatttaatctgtatcaTTAAAGAT CAGCAAGCAAAGCTTTG	2-step cloning of *apn* pathway (1^{st} step)

C.1.2 Plasmid construction

Cyanobacterial biosynthetic constructs were either assembled in a single cloning step or stepwise. PCR amplicons were generated using cyanobacterial genomic DNA as template, unless otherwise specified. The following biosynthetic pathways contained in pMQ123i were

constructed in a single step: *ageABCDEFG*, *kawABCDEFG*, *mcaABCDEFG*, *mcaADEFG* and *spuAB*. Details for the construction of the *mcaABCDEFG*, *mcaADEFG*, and *mcaADEG* pathways are described in **Microcyclamide Chapter 8**. The *kaw* pathway was assembled as three separate fragments that were amplified using the following primers, respectively: kaw_ frag1_5'_toMQ123i / kaw_ frag1_3', kaw_ frag2_5 '/ kaw_ frag2_3', kaw_ frag3_5'/kaw_ frag3_3'. The *spu* pathway was assembled as five separate fragments that were amplified using the following primers, respectively: 123i_spu_ Fg1_FW / 123i_spu_ Fg1_RV, 123i_spu_ Fg2_FW / 123i_spu_ Fg2_RV, 123i_spu_ Fg3_FW / 123i_spu_ Fg3_RV, 123i_spu_ Fg4_FW / 123i_spu_ Fg4_RV, 123i_spu_ Fg5_FW / 123i_spu_ Fg5_RV.

The following biosynthetic pathways contained in pMQ123i were constructed in two steps: *mcaADEF*, *mcaADEFG*, *apnABCD*, and *halABCDE*. Details for the construction of the complete *hal* pathway are described in **Hapalosin Chapter 6**. Details for the construction of the *mca* pathways are described in **Microcyclamide Chapter 8**. The apnABCD pathway was constructed by first generating the *apnCD* construct in pMQ123i. Four fragments corresponding to genes *apnCD* were generated via PCR using the following primers to recombineer at the HindIII and SpeI sites of pMQ123i: apnCD_fg1_FW_into_123i/apnCD_fg1_RV, apnCD_fg2_FW / apnCD_fg2_RV, apnCD_fg3_FW / apnCD_fg3_RV, apnCD_fg4_FW / apnCD_fg4_RV_into123i. The *apnAB* genes were subsequently cloned using primers apnA_fg1_FW_into_123i/apnA_fg1_RV and apnA_fg2_FW/apnA_fg2_RV, generating three fragments, which were cloned at the Eco53kI restriction site of the *apnCD* plasmid construct.

PCR amplification of fragments was performed using an annealing temperature gradient that ranged from 55°C to 65°C. PCR conditions were as follows: 95°C initial denaturation for 3:00min, 98°C denaturation for 20sec, 55°C to 65°C annealing for 15sec, 72°C extension for 1 min per 1kbp, and 72°C final extension for 5:00min. The denaturation, annealing, and extension steps were performed a total of 35 times. Volume per PCR reaction was 10µL. The PCR reactions were stored at 4°C or -20°C thereafter. Approximately 10ng of purified genomic DNA was utilized per reaction tube. PCR amplification was verified by agarose gel (1% (w/v)) electrophoresis, followed by ethidium bromide staining and visualization under UV illumination. When the desired amplicon was generated as a mixture, the amplicon of interested was excised from the agarose gel following electrophoresis.

PCR primers were designed such that recombineering fragments overlap by at least 30bp between each insert or with the plasmid backbone. Non-native and synthetic *E. coli* ribosome binding site (RBS) sequences were introduced within the pathway and upstream of genes by incorporating these sequences into the middle of a 3' primer sequence of one recombineering fragment and also into the middle of a 5' primer sequence of a downstream recombineering fragment.

The procedures for yeast-assisted homologous recombination to construct cyanobacterial biosynthetic pathways was adopted from procedures previously report [74, 203]. An overnight culture (1.0mL) of *Saccharomyces cerevisiae* cloning strain VL6-48N grown in yeast-peptone dextrose media was washed twice with equal volumes of sterile TE (pH 7.4) buffer. After washing, 0.5mL of Lazy Bones solution (40% (w/v) PEG MW 3350, 100mM lithium acetate,

10mM Tris-HCl pH 7.5, 1mM EDTA) was added to the yeast cells. Then, 20µL salmon sperm DNA (2mg/mL) denatured at 98°C for 10min and cooled at 4°C was added to the yeast cells. Lastly, 1µg of pMQ123i plasmid (linearized via restriction enzyme and dephosphorylated) and all biosynthetic pathway fragments were added to the transformation mixture. The mixture was then vortexed for 2 minutes and then allowed to incubate at approximately 25° overnight. The following day, the yeast transformation mixture was heated to 42°C for approximately 12 minutes and then washed three times with equal volumes of sterile TE prior to a final resuspension of the yeast cells in 200µL, which were plated on synthetic complete minus uracil (SC-URA) selection agar plates. The yeast were allowed to grow on selection plates incubate at 30°C for three to four days.

Yeast recombineered plasmids were isolated from yeast as follows. Following growth on SC-URA selective plates, approximately 50µL yeast cells were resuspended in sterile TE (200µL), then an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) and approximately 100µL of glass disruption beads (0.5mm diameter) were added to the yeast suspension. This mixture was vortexed for 2 minutes and then centrifuged at 21.1×10^3 g for 5 minutes at room temperature. 150µL of the aqueous phase from the centrifuged mixture is then added to 105µL of isopropanol, which is mixed by gently inverting the tube and allowed to sit at room temperature for 5 minutes to allow precipitation of plasmid DNA. Following precipitation, the mixture was centrifuged at 21.1×10^3 g for 5 minutes and the supernatant was carefully decanted with gentle aspiration. To wash the resultant DNA precipitate, 1mL of ethanol (70% (v/v)) is added to the tube to wash, and then the tube is centrifuged at 21.1×10^3 g for 5 minutes. The DNA precipitate was washed a total of two times before completely drying the DNA under

gentle heating (37°C). The plasmid DNA was dissolved in sterile TE (30µL) and between 2 to 5µL of DNA was used for transformation into *E. coli*.

C.1.3 Heterologous expression of cyanobacterial biosynthetic pathways

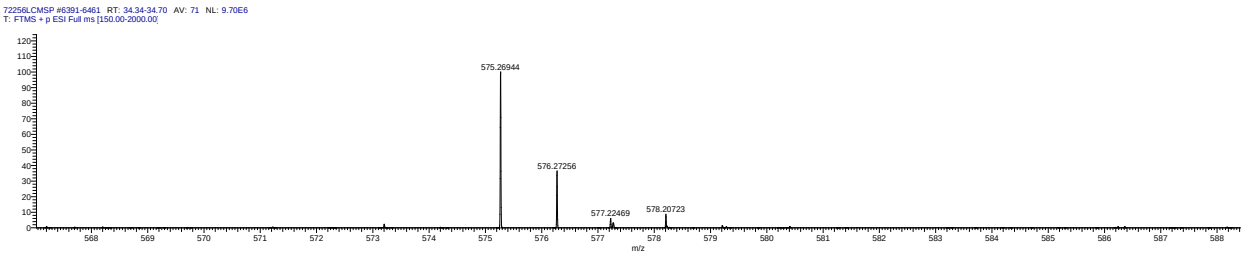
Electrocompetent *E. coli* TOP10 cells were transformed with pMQ131 (containing the *sfp* gene and only used when necessary) or pMBI (when necessary) and pMQ123i constructs and allowed to recover in SOC media for 1hr at 37°C before plating on LB agar plates containing kanamycin (50 µg/mL), gentamicin (10 µg/mL), and/or tetracycline (10 µg/mL). A single colony of the transformant was used to inoculate 1mL of LB containing the appropriate antibiotics and allowed to grow at 37°C overnight. 500µL of the seed culture was subsequently used to inoculate fresh LB media (50mL) containing the appropriate antibiotics and allowed to grow at 37°C with shaking (200 rpm). When the culture reached mid-log phase exponential growth (OD_{600} = 0.6), the flask was cooled to 25°C and IPTG (1mM) was introduced. At the time of IPTG induction, substrates including mevalonolactone (1mM final concentration), homo-L-tyrosine (1mM final concentration), *p*-chlorophenylalanine (5mM final concentration), or *p*-fluorophenylalanine (5mM final concentration) were introduced to the culture media, when necessary. After 48hr of induction at 25°C, the culture was centrifuged (4696 g, 30min) and the supernatant was extracted with ethyl acetate (2×50mL). The ethyl acetate layers were combined and dried over Na_2SO_4 and the organic solvent was evaporated *in vacuo* under gentle heating (37°C).

The ethyl acetate-supernatant crude extract was dissolved in methanol to a concentration of approximately 10 mg/mL and centrifuged (21.1×10^3 g, 5min). The crude extract was dissolved in methanol to a concentration of approximately 10 mg/mL and centrifuged (21.1×10^3 g, 5min).

The organic supernatant was then used for subsequent high resolution liquid chromatography mass spectrometric (HR-LC-MS) analysis. MeOH-dissolved supernatant extract (2 µL) was separated by passing through a reverse phase HPLC column (Phenomenex Jupiter® C18, 370Å, 3µm, 2×150mm) using a Dionex® Ultimate3000 HPLC and analyzed on a Thermo Scientific™ Q-Exactive HRMS instrument using electrospray ionization. Thermo Scientific™ Excalibur™ software was used for MS data analysis. LC method used for analysis (0.2 mL/min flow rate): 5% (v/v) MeCN in H_2O (including 0.1% (v/v) formic acid) for 3 min, ramp to 95% (v/v) MeCN in H_2O (including 0.1% (v/v) formic acid) over 45 min, and then a final wash with 95% MeCN in H_2O (including 0.1% (v/v) formic acid) for 15 min.

C.1.4 High resolution mass spectrometry analysis of heterologous expression constructs

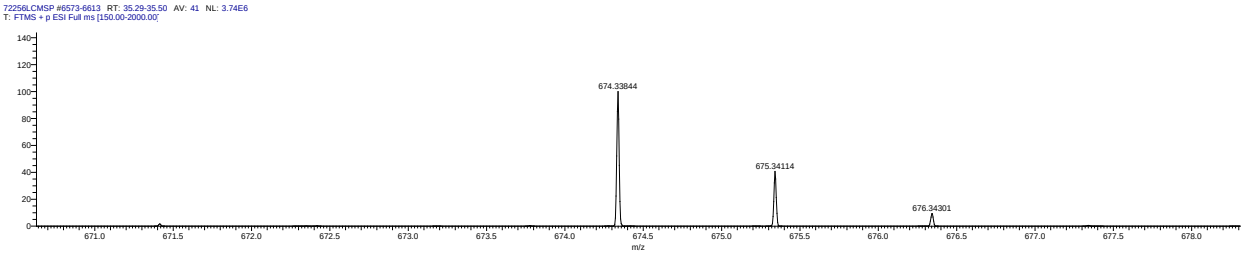
MS analysis for heterologous expression of *mca* (microcyclamide) and *hal* (hapalosin) gene cluster constructs can be found within the Appendices of **Chapter 8, Microcyclamide** and **Chapter 6, Hapalosin**, respectively.



72256LCMSP#6391-6461 RT: 34.34-34.70 AV: 71
T: FTMS + p ESI Full ms [150.00-2000.00]
m/z= 575.00-576.00

m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
575.26944	9985890.0	100.00	575.26865	0.79	$C_{32}H_{39}O_4N_4S$

Figure C1 Accurate mass for **9** obtained from heterologous expression of *age* gene cluster in *E. coli* TOP10.



72256LCMSP#6574-6614 RT: 35.29-35.50 AV: 41
T: FTMS + p ESI Full ms [150.00-2000.00]
m/z= 674.00-675.00

m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
674.33844	3766890.5	100.00	674.33707	1.37	$C_{37}H_{48}O_5N_5S$

Figure C2 Accurate mass for **10** obtained from heterologous expression of *age* gene cluster in *E. coli* TOP10.

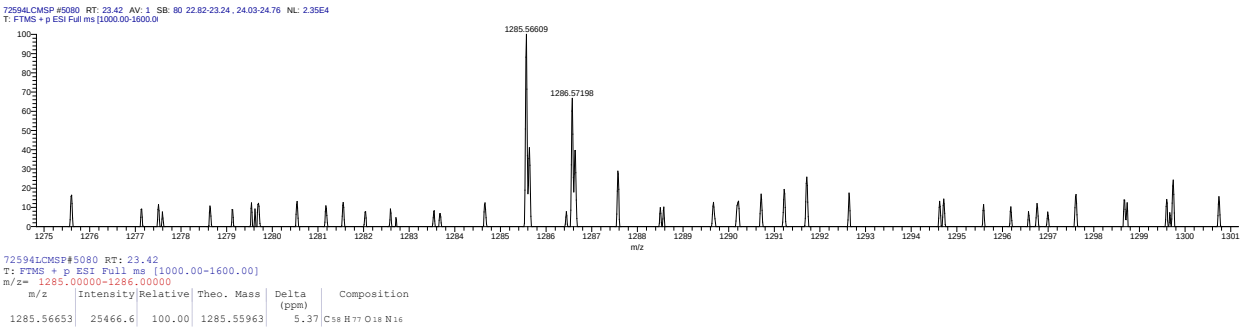


Figure C3 Accurate mass for **5** obtained from heterologous expression of *apn* gene cluster in *E. coli* TOP10.

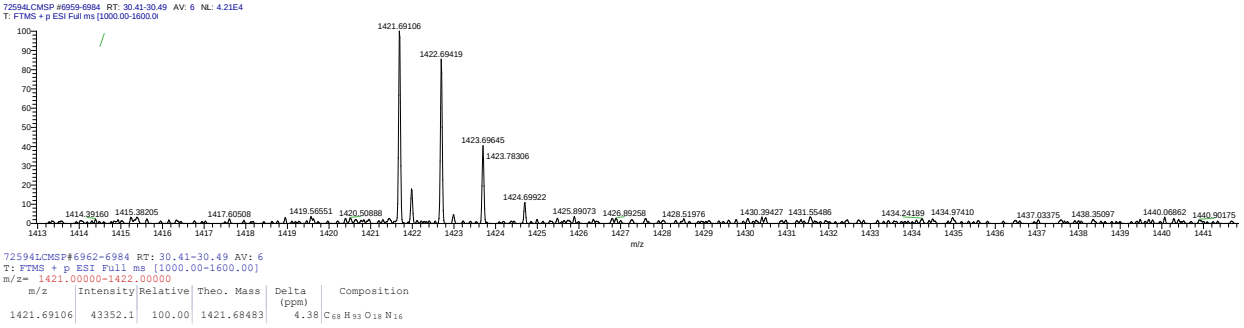
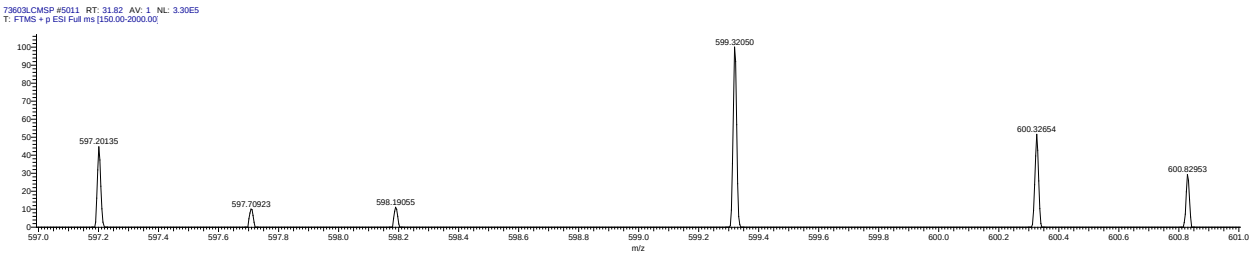


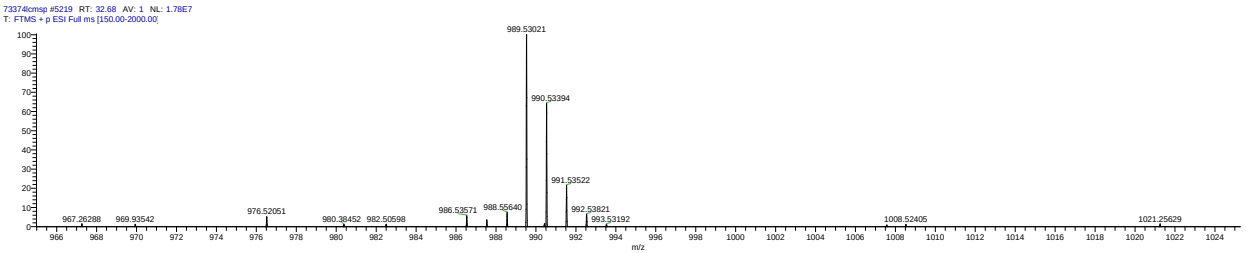
Figure C4 Accurate mass for **6** obtained from heterologous expression of *apn* gene cluster in *E. coli* TOP10.



73603LCMSP#5011 RT: 31.82
T: FTMS + p ESI Full ms [150.00-2000.00]
m/z= 599.00-600.00

m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
599.32	336170.5	100.00	599.32	1.72	$C_{30}H_{43}O_7N_6$

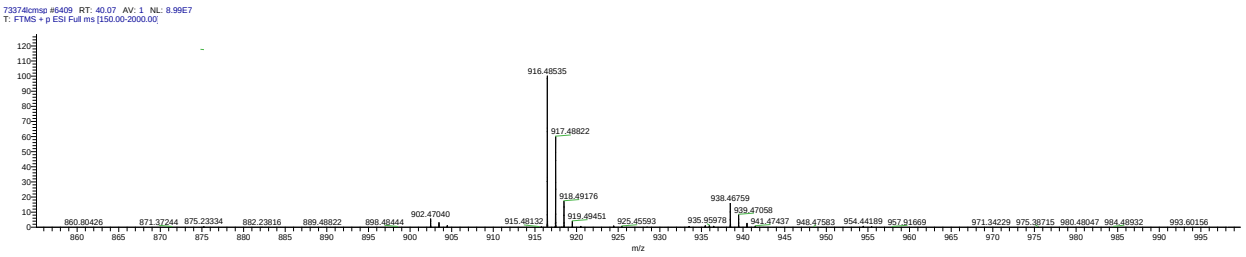
Figure C5 Accurate mass for **2** obtained from heterologous expression of *apn* gene cluster in *E. coli* TOP10.



73374lcmsp#5217 RT: 32.67
T: FTMS + p ESI Full ms [150.00-2000.00]
m/z= 909.00-910.00

m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
909.52246	14946783.0	100.00	909.51926	3.20	$C_{45}H_{69}O_{10}N_{10}$

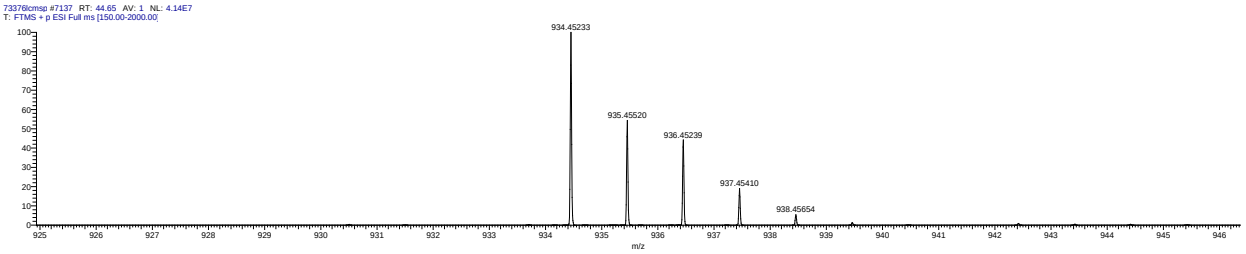
Figure C6 Accurate mass for **3** obtained from heterologous expression of *apn* gene cluster in *E. coli* TOP10.



73374lcmsp#6409 RT: 40.07
T: FTMS + p ESI Full ms [150.00-2000.00]
m/z= 916.00-917.00

m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
916.48535	93307496.0	100.00	916.48148	3.87	$C_{48}H_{66}O_{11}N_7$

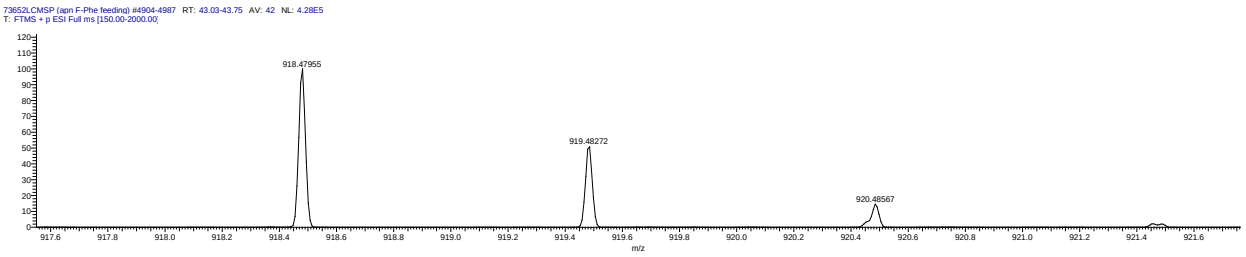
Figure C7 Accurate mass for **4** obtained from heterologous expression of *apn* gene cluster in *E. coli* TOP10.



73376lcmsp#7135 RT: 44.64
T: FTMS + p ESI Full ms [150.00-2000.00]
m/z= 934.00-935.00

m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
934.45221	37466812.0	100.00	934.44760	4.61	$C_{48}H_{65}O_{10}N_7Cl$

Figure C8 Accurate mass for **4a** obtained from heterologous expression of *apn* gene cluster in *E. coli* TOP10.



73652LCMSP (apn F-Phe feeding)#4905-4987 RT: 43.03-43.75 AV: 42
T: FTMS + p ESI Full ms [150.00-2000.00]
m/z= 918.00-919.00

m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
918.47955	432929.6	100.00	918.47715	2.40	$C_{48}H_{65}O_{10}N_7F$

Figure C9 Accurate mass for fluorinated **4** obtained from heterologous expression of *apn* gene cluster in E. coli TOP10.

APPENDIX D

SUPPORTING INFORMATION FOR CHAPTER 6

D.1.1 General methods and materials

PCRs were executed using a C1000 thermal cycler (Bio-Rad). DNA sequencing was performed by Elim BioPharm Inc. Preparative-scale reverse-phase HPLC was performed using a Dionex instrument equipped with a 21 × 250 mm Luna C18 column (Phenomenex). LRMS analysis was conducted using a Shimadzu LCMS-2020 instrument equipped with an Acclaim® C18 column (Dionex, 120Å, 3µm, 2.1 × 150mm). HRMS analysis was conducted using a Q Exactive Benchtop Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific) equipped with a Jupiter® C18, 370Å, 3µm, 2×150mm column (Phenomenex). All MS experiments utilized electrospray ionization modes for analysis. The NMR spectrum was recorded on a Bruker Avance III 700 MHz spectrometer equipped with a 1H/13C/15N triple-resonance inverse probe (1.7 mm 'microprobe').

Synthetic oligonucleotides for gene amplification by PCR were purchased from Life Technologies or Integrated DNA Technology. Kapa HiFi DNA polymerase was obtained from Kapa Biosystems. Restriction endonucleases, Antarctic phosphatase was purchased from New

England BioLabs. LB broth and agar used for culturing E. coli were obtained from Teknova. 6-bromohexanoic and 8-bromooctanoic acids were both obtained from Sigma. 7-bromoheptanoic acid was obtained from Matrix Scientific. Sodium azide was obtained from Fisher Scientific.

D.1.2 Plasmid construction

To reconstruct the entire hapalosin biosynthetic pathway, genes halA and sfp were cloned by PCR and yeast recombineered into parent vector pMQ131 to yield daughter plasmid pXL2. These genes were located downstream of the isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible Lac promoter. Included within the design of the PCR primers were sequences to synthetically insert an *E. coli* ribosome binding site (RBS) immediately upstream of gene *halA*.

A sequential cloning procedure was implemented to yeast recombineer genes *halBCDE* into parent vector pMQ123i to yield the final daughter plasmid pXL4. Sequential construction of the final expression plasmid enabled the genetic engineering of an *E. coli* RBS sequence immediately upstream of genes halDE as well as a site-directed mutation of the native TTG start codon of *halD* to the high frequency start codon ATG of E. coli. This was accomplished by first recombineering genes *halDE* into pMQ123i, yielding the partial expression plasmid pXL3. Similar to the construction of plasmid pXL2, PCR primers were designed to introduce the RBS sequences upstream of the *halDE* genes. Following successful construction of the halDE initial plasmid, known as pXL3, *halBC* were introduced to afford the *halBCDE* expression plasmid, pXL4. These genes were situated downstream of the IPTG-inducible Tac promoter. A summary of the plasmids used in this study are listed in **Table D2**.

D.1.3 Heterologous expression of the hapalosin biosynthetic pathway

Electrocompetent *E. coli* XL1-Blue cells were transformed with pMQ131 and pMQ123i constructs and allowed to recover in SOC media for 1hr at 37°C before plating on LB agar plates containing kanamycin (50 µg/mL) and gentamicin (10 µg/mL). A single colony of the transformant was used to inoculate 1mL of LB containing the appropriate antibiotics and allowed to grow at 37°C overnight. 500µL of the seed culture was subsequently used to inoculate fresh LB media (50mL) containing the appropriate antibiotics and allowed to grow at 37°C with shaking (200 rpm). When the culture reached mid-log phase exponential growth (OD_{600} = 0.6), the flask was cooled to 25°C and IPTG (1mM) was introduced. After 48hr of induction at 25°C, the culture was centrifuged (5000 g, 30min) and the supernatant was extracted with ethyl acetate (2×50mL). The ethyl acetate layers were combined and dried over Na_2SO_4 and the organic solvent was evaporated in vacuo under gentle heating (37°C).

D.1.4 LC-MS analysis of hapalosin expression in *E. coli* and LC-HR-MS^2

The ethyl acetate-supernatant crude extract was dissolved in methanol to a concentration of approximately 10 mg/mL and centrifuged (21.1×10^3 g, 5min). The organic supernatant was then used for subsequent LR-LCMS analysis. 2µL of MeOH-dissolved supernatant extract was separated by R-HPLC prior to LR-LCMS analysis. LC method used for analysis (0.2 mL/min flow rate): 5% (v/v) MeCN in H_2O (including 0.1% (v/v) formic acid) for 5 min, ramp to 95% (v/v) MeCN in H_2O (including 0.1% (v/v) formic acid) over 45 min, and then a final wash with 95% MeCN in H_2O (including 0.1% (v/v) formic acid) for 5 min.

D.1.5 High resolution tandem-MS analysis of hapalosin analog production in *E. coli*

The same MeOH-soluble extract was subjected to HR-LCMS and HR-LCMS^2 analysis. LC method used for analysis (0.2 mL/min flow rate): 5% (v/v) MeCN in H_2O (including 0.1% (v/v) formic acid) ramp to 90% (v/v) MeCN in H_2O (including 0.1% (v/v) formic acid) over 55 min, and then a final wash with 90% MeCN in H_2O (including 0.1% (v/v) formic acid) for 5 min.

A precursor ion screening method analogous to ones reported previously [247, 248] was implemented to identify novel hapalosin compounds. Hypothesized structures of hapalosin compounds based on their chemical formulae were initially identified by their expected precursor ion accurate mass within an error of 5 ppm. Accurate masses were subsequently analyzed for fingerprint fragmentation peaks. Characteristic fragment ions frequently consisted of two sets of ions: 188, 206, 224 *m/z* and 288, 306, 324 *m/z*, which differ by the number of dehydrations. Additionally, fragment ions were correlated structurally by their accurate masses.

D.1.6 Large scale purification of hapalosin from *E. coli* extracts

Large scale (3×1.0L) liquid cultures of *E. coli* XL1-Blue strains containing the hapalosin pathway were grown and induced as discussed previously for small scale (50mL) expression cultures. Following induction, liquid cultures were centrifuged (7000 g, 30min) and supernatant was discarded. Cell pellets obtained from centrifugation were then extracted by adding acetone including 0.1% (v/v) formic acid (300mL) to the cell pellet. The cell-acetone mixture was then vortexed thoroughly and mixed at 25°C overnight. After mixing, the organic supernatant was filtered through Celite®545 (Acros Organics) under reduced pressure and the organic extract

was evaporated *in vacuo* under gentle heating (37°C), yielding between 200-250mg of crude extract. The crude was then re-dissolved in 2mL of 4:1 MeOH/DCM, centrifuged (21.1×10^3 g, 5min), and the organic supernatant was applied to a size exclusion column (Sephadex LH-20, Fluka, 120 cm × 1.5 cm diameter, flow rate 48 mL/hr, 4:1 MeOH/DCM mobile phase). Fractions were screened by LC-MS. Hapalosin-containing fractions were found to elute between 160-180mL of mobile phase.

All hapalosin-containing fractions from size-exclusion purification were combined, yielding between 5 to 10mg of enriched crude. The combined fractions were dissolved in 1.2mL of MeOH, centrifuged (21.1×10^3 g, 5min), and the organic supernatant was used for preparative HPLC. Method used for purification (10.620 mL/min flow rate, 208nm wavelength monitoring): 50% (v/v) MeOH in H_2O (including 0.1% (v/v) formic acid) for 4 min, ramp to 100% (v/v) MeOH in H_2O (including 0.1% (v/v) formic acid) over 100 min, and then a final wash with 100% MeOH in H_2O (including 0.1% (v/v) formic acid) for 5 min. Hapalosin eluted at approximately 84 min retention time under these conditions. Approximately 100µg of pure hapalosin was obtained using this procedure.

D.1.7 Cultivation of *Hapalosiphon welwitschii* UTEX 1830 and metabolite analysis

UTEX B1830 was cultured as described previously. Briefly, 1L volume cultures were grown in BG-11 media adjusted to pH = 7.8. Cyanobacteria were separated via filtration and the cells were lyophilized to yield a dry, filamentous green mass. To 5.0g of lyophilized cells, 400mL of 1:1 MeOH/DCM was added and stirred overnight at room temperature. The organic solvent was separated from the cell mass via filtration and evaporated *in vacuo* under gentle heating (37°C)

to yield the dried crude extract. The crude extract was redissolved in MeOH and analyzed via LC-MS and HR-MS^2 using the same protocol as described above for hapalosin obtained from *E. coli*.

D.1.8 NMR experiments

Purified hapalosin was analyzed by ^1H NMR ($CDCl_3$) and ^1H-^1H correlation spectroscopy (COSY) NMR. For all NMR experiments, Bruker Avance III 700MHz NMR instrument with a $^1H/^{13}C/^{15}N$ triple-resonance inverse probe (1.7mm microprobe) was used for analysis.

D.1.9 Azido-alkanoic acid synthesis

The procedure to synthesize azido-alkanoic acid substrates for feeding experiments was adapted from a protocol previously reported. [234] Briefly, the corresponding 6-, 7-, or 8-bromo-alkanoic acid (4mmol) was dissolved in DMF (20mL) and sodium azide (0.82g, 12.6mmol) was added and allowed to stir overnight at room temperature. Upon completion of the reaction as monitored by TLC, the reaction was diluted with DCM (100mL) and washed with ice cold 1N HCl (7×100mL), followed by brine (2×100mL) washing. The organic layer was dried over Na_2SO_4 and evaporated *in vacuo* under gentle heating (37°) to yield yellow viscous oil. See below Appendix figures for ^1H NMR of azidoalkanoic acid products.

D.1.10 Azido-alkanoic acid feeding experiments

Overnight cultures of *E. coli* XL1-Blue strains containing the hapalosin biosynthetic genes were diluted 100-fold in fresh LB media (50mL), supplemented with the corresponding azido-alkanoic acid (500µM for 6- and 7-azido-alkanoic acids and 1mM for 8-azidooctanoic acid; 500mM azido-alkanoic acid stock solutions were prepared in DMSO), incubated at 37°C with shaking (200 rpm), to allow growth to cell density OD_{600} = 0.6. At mid-log phase exponential growth, cultures were cooled to 25°C and induced with IPTG (1mM). After 48hr of induction, the culture supernatant was extracted and analyzed as described previously for the analysis of native hapalosin.

Table D1 Function annotation of individual gene product in *hal* biosynthetic gene cluster

Gene	Size, aa	Deduced role	Protein Homolog	Number	Identity/positivity %/%
halA	708	Fatty acid-CoA ligase (AL-T)	acyl-CoA synthetase [Rivularia sp. PCC 7116]	YP_007059074.1	56/71
halB	1981	Polyketide synthase (KS-AT-MT-KR-T)	beta-ketoacyl synthase [Nostoc punctiforme PCC 73102]	YP_001865643.1	49/64
halC	2348	Non-ribosomal peptide synthetase/polyketide synthase hybrid (C-A-A-KR-T)	HctF [Lyngbya majuscula]	AAY42398.1	47/63
halD	1526	Non-ribosomal peptide synthetase (A-A-MT-T)	anabaenopeptilide synthetase ApdB [Anabaena sp. 90]	YP_006996377.1	64/77
halE	1818	Polyketide synthase (KS-AT-KR-T-TE)	beta-ketoacyl synthase [Nostoc punctiforme PCC 73102]	YP_001866779.1	57/73

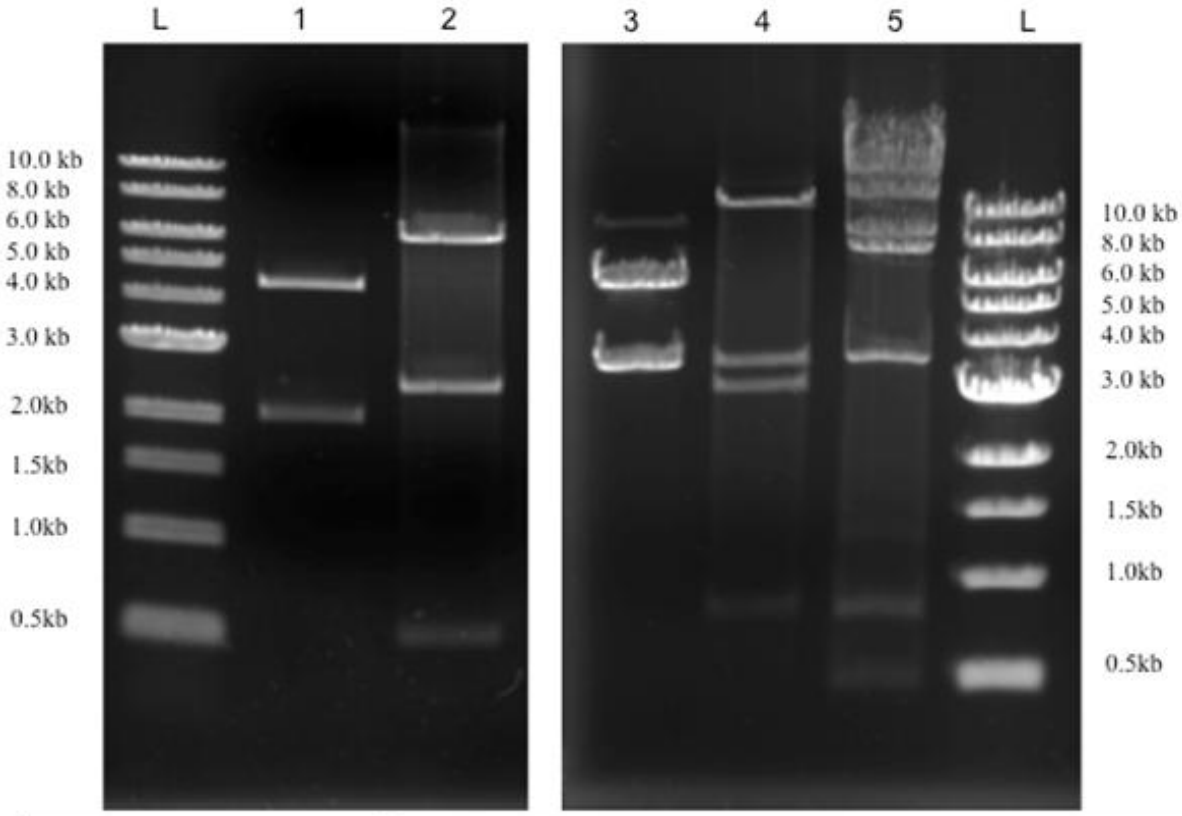
Table D2 Plasmids used in this study

Plasmid name	Features	References
pMQ131	$_{ori}$ColE1/pC194, *aphA-3*, P_{lac}-*lacZα, CEN6*	Reference [74]
pHal2	Derived from pMQ131; contains *halA* and *sfp*	This work
pMQ123i	$_{ori}$ColE1/pRK2, *aacC1*, P_{tac}-*lacZα, 2µm*	Reference [74]
pHal3	Derived from pMQ123i; contains *halDE*	This work
pHal4	Derived from pHal3; contains *halBC*	This work

Table D3 Primers used in this study

Primer name	Sequence	Application
sfpRBSFW	TGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGAAGATTTACGGAATTTATATG	5’ primer to clone *sfp* and RBS
sfpRBSRV	TTATATCCTCCTACGGGTATGGAGAATTATAAAAGCTCTTCGTACGAG	3’ primer to clone *sfp* and RBS
HapAFW	TTCTCCATACCCGTAGGAGGATATAAATGAACAACAATTTATTTTCAG	5’ primer to clone *halA*
HapARV	CTGTTTTATCAGACCGCTTCTGCGTTCTGATGGGCCCTTAAACACTAACTGAAGAGG	3’ primer to clone *halA*
HalDE_Frag1_FW	CGAATTCGAGCTCGGTACCCGGGGAAGGAGATATACATATGAGCAATATCTCTAAAAAAACC	5’ primer to clone fragment 1 of *halDE*
HalDE_Frag1_RV	TTCTTCTTCGGCGGCTACACTAGAG	3’ primer to clone fragment 1 of *halDE*
HalDE_Frag2_FW	GCGTTGAGTATTTGTTGCAGG	5’ primer to clone fragment 2 of *halDE*
HalDE_Frag2_RV	CCTTGTTCTAGTAGCTTCGAG	3’ primer to clone fragment 2 of *halDE*
HalDE_Frag3_FW	GTCAATTGATGCAGCAATTGC	5’ primer to clone fragment 3 of *halDE*
HalDE_Frag3_RV	TATCAGACCGCTTCTGCGTTCTGATTTAATCTGTATCATCAATCATCTGCCTGTGACCG	3’ primer to clone fragment 3 of *halDE*
HalBC_Frag1_FW	GAATTGGATCCTCTAGATTCTCCATACAGGAGGAATAATATGAATAGAGAACCAATCGCC	5’ primer to clone fragment 1 of *halBC*
HalBC_Frag1_RV	CTTGTGCTAAATTTGACCCAC	3’ primer to clone fragment 1 of *halBC*
HalBC_Frag2_FW	CACTCGCGGCTAATGCATCGTC	5’ primer to clone fragment 2 of *halBC*
HalBC_Frag2_RV	GTTTGGGGTGGGAGTGCTACC	3’ primer to clone fragment 2 of *halBC*
HalBC_Frag3_FW	gagcaacgaacccgtaatacgc	5’ primer to clone fragment 3 of *halBC*
HalBC_Frag3_RV	gtaaaattattggcttgagcggtagcag	3’ primer to clone fragment 3 of *halBC*
HalBC_Frag4_FW	ggtcaaagagattgaagaaacaaaac	5’ primer to clone fragment 4of *halBC*

HalBC_Frag4_RV	catatgtatatctccttccccgggtaccgagtcaagtattttcggcttttagc	3' primer to clone fragment 4 of *halBC*



Lane	Plasmid	Digest analysis	Expected products (bp)
1	*pMQ131*	HindIII	4461, 2019, 13
2	*pHal2*	HindIII	6124, 2406, 499, 13
3	*pMQ123i*	BglII	6076, 3621
4	*pHal3*	BglII	11426, 3621, 3090, 837
5	*pHal4*	BglII	11426, 8299, 7286, 3671, 837, 479

Figure D1 The successful construction of expression plasmids containing the genes of the hapalosin biosynthetic pathway were verified by restriction enzyme digest analysis. *sfp-halA* plasmid construct pHal2 (lane 2) was verified by restriction enzyme digest analysis. Additionally, the initial, empty plasmid with no genes, pMQ131, was also analyzed for

comparison (lane 1). The expression plasmid containing the partial pathway *halDE* in plasmid pMQ123i (lane 4) was compared to the expression plasmid containing the full biosynthetic pathway (lane 5). The empty plasmid pMQ123i was also analyzed as a control (lane 3).

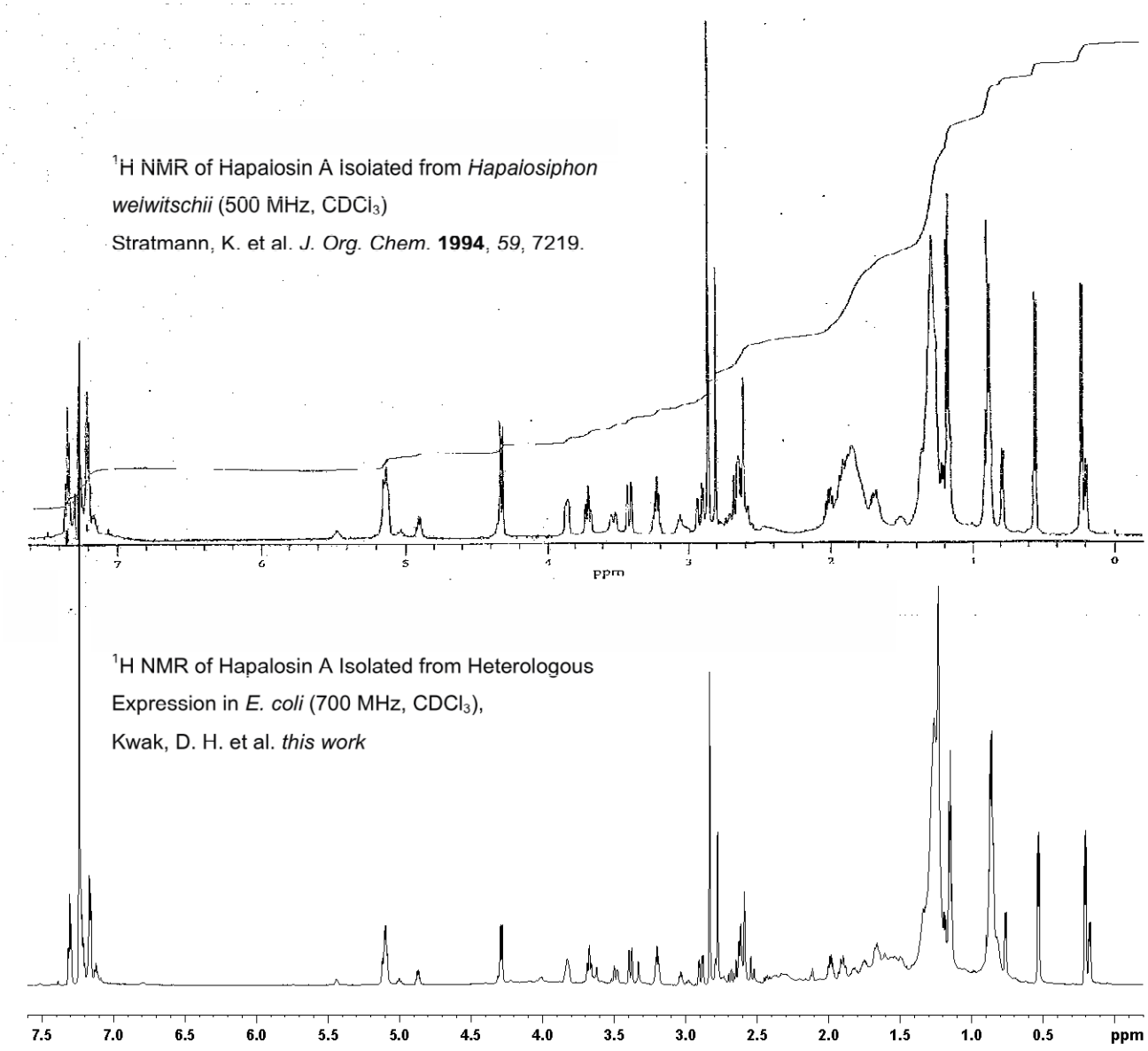


Figure D2 Comparative 1H-NMR with hapalosin standard from Moore *et al.*

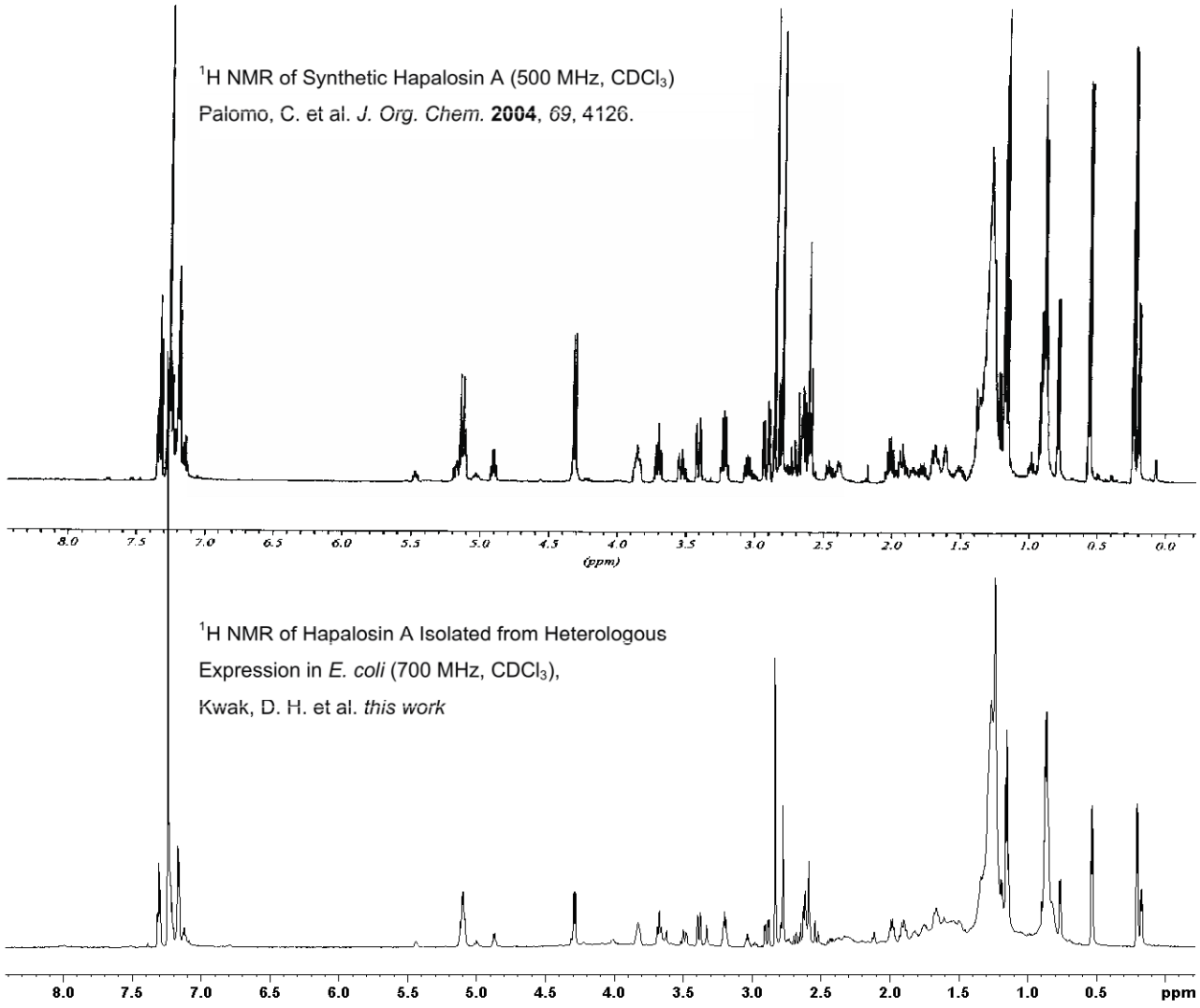


Figure D3 Comparative 1H-NMR with hapalosin standard derived via synthetic methods.

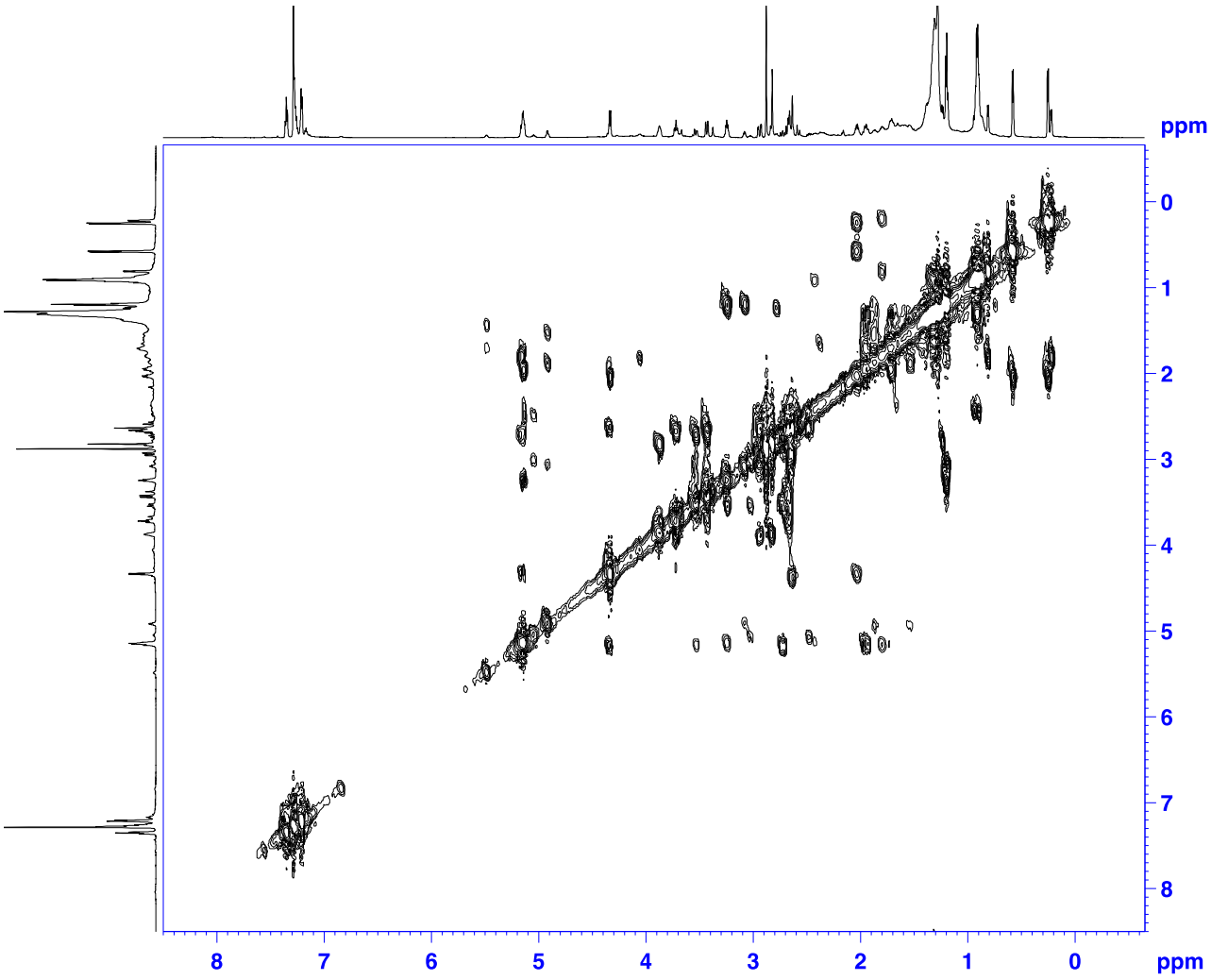


Figure D4 1H-1H COSY NMR analysis of hapalosin isolated from *E. coli* in this study.

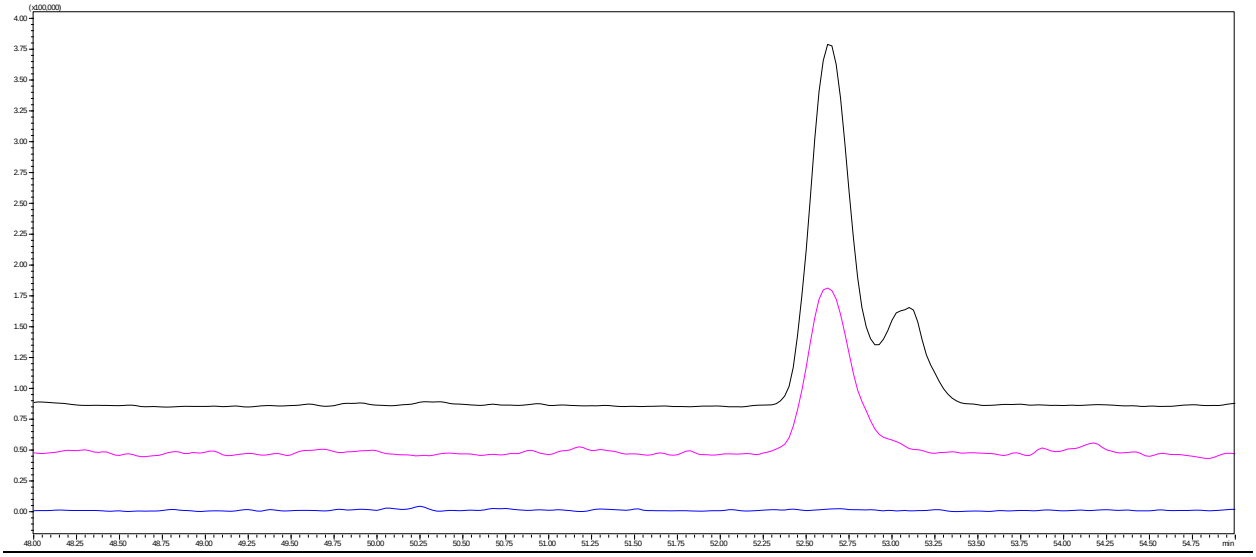


Figure D5 Low resolution LC-MS analysis of hapalosin observed from native producer *H. welwitschii* UTEX1830B (black), *E. coli* heterologous expression (pink). *E. coli* containing no genes of the hapalosin pathway in pMQ123i was used as a negative control (blue).

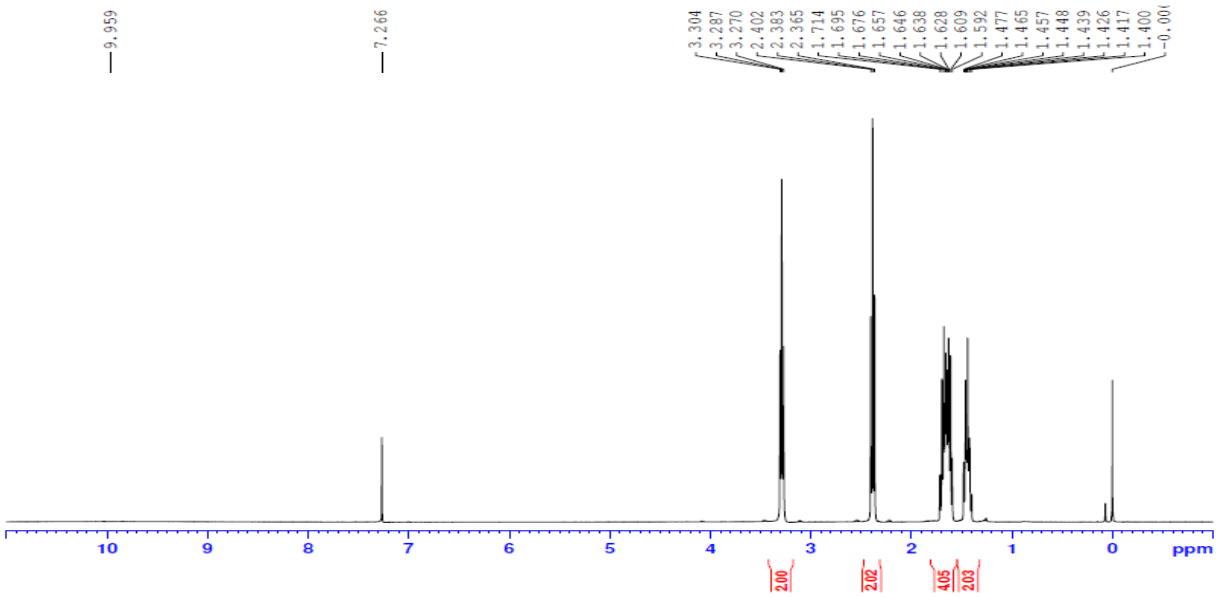


Figure D6 1H NMR spectrum of 6-azidohexanoic acid (400MHz, $CDCl_3$).

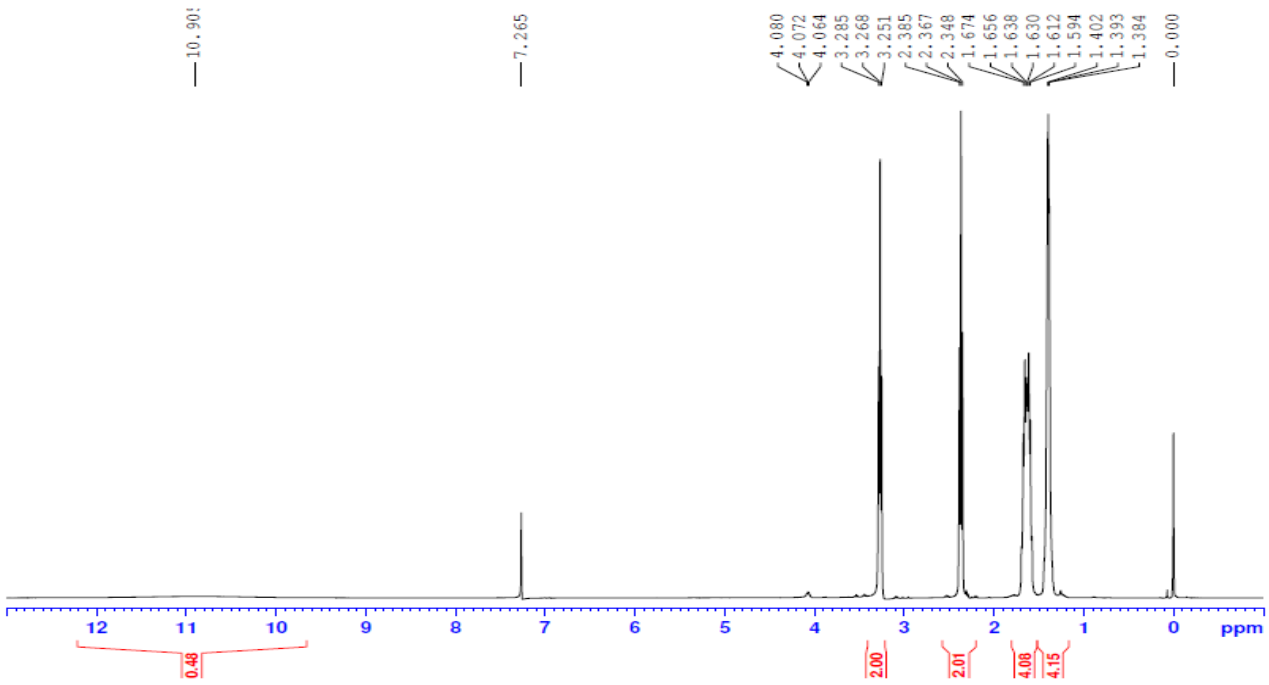


Figure D7 1H NMR spectrum of 7-azidoheptanoic acid (400MHz, $CDCl_3$).

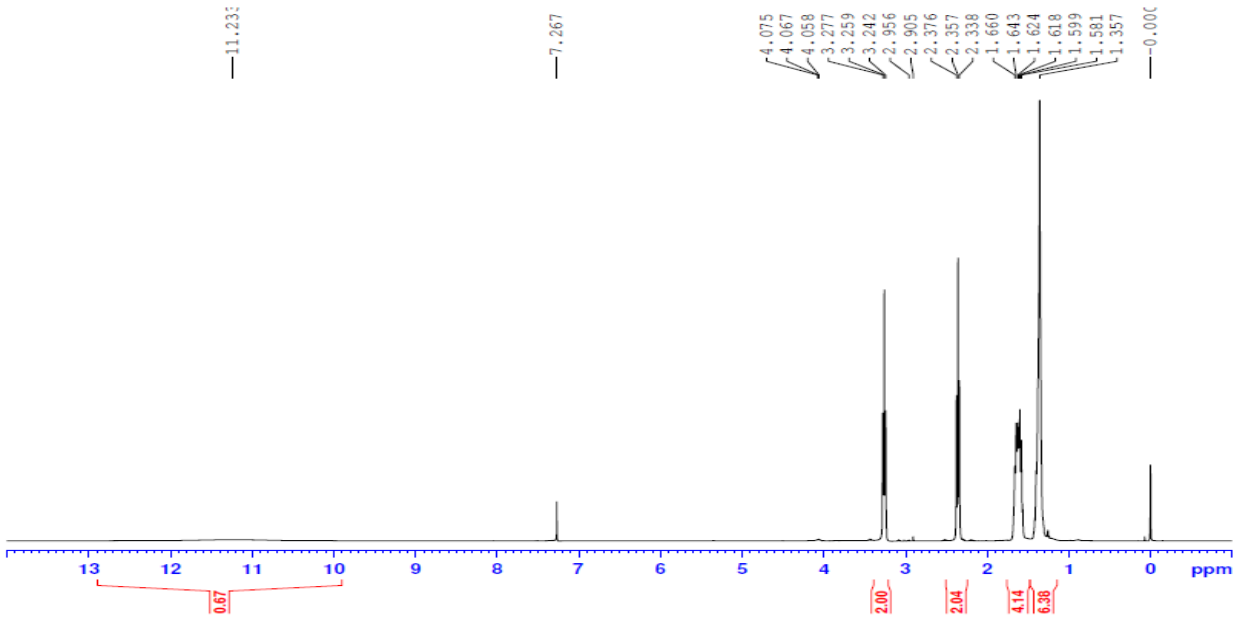
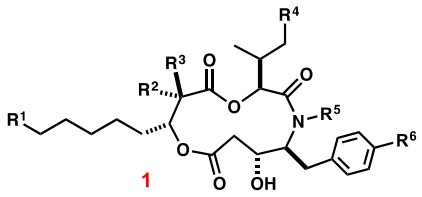
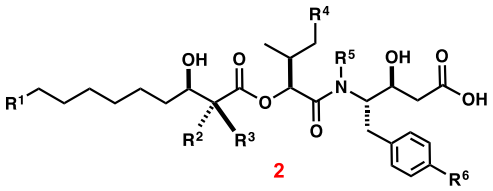


Figure D8 1H NMR spectrum of 8-azidooctanoic acid (400MHz, $CDCl_3$).

Table D4 List of compounds detected from heterologous hapalosin pathway in *E. coli*



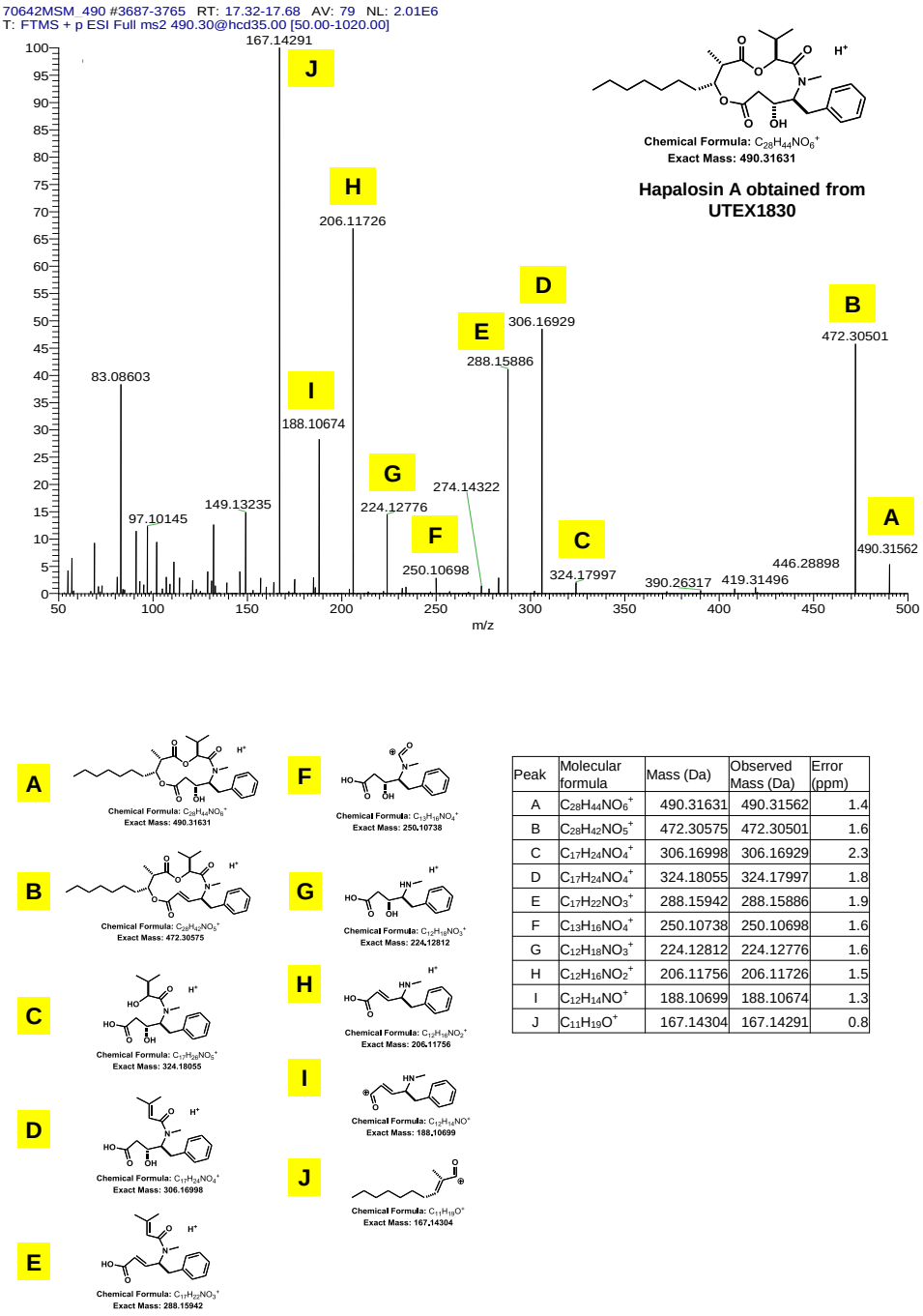
1



2

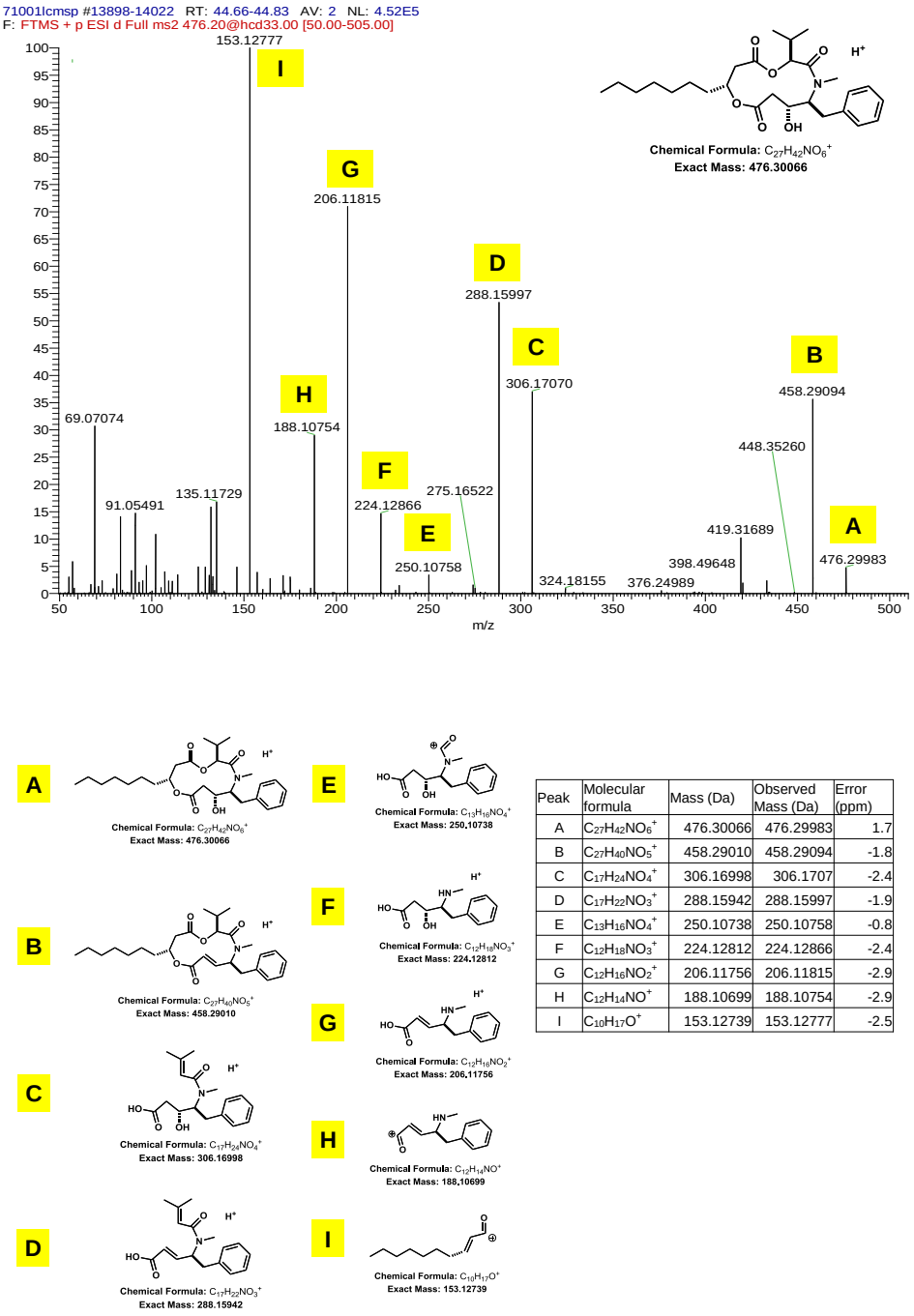
Compound number	R^1	R^2	R^3	R^4	R^5	R^6
1a	CH_3CH_2	CH_3	H	H	CH_3	H
1aa	CH_3CH_2	H	H	H	CH_3	H
1ab	CH_3CH_2	CH_3	H	H	H	H
1ac	CH_3CH_2	CH_3	CH_3	H	CH_3	H
2a	CH_3CH_2	CH_3	H	H	CH_3	H
2aa	CH_3CH_2	H	H	H	CH_3	H
2ab	CH_3CH_2	CH_3	H	H	H	H
1b	CH_3CH_2	CH_3	H	CH_3	CH_3	H
1ba	CH_3CH_2	H	H	CH_3	CH_3	H
2b	CH_3CH_2	CH_3	H	CH_3	CH_3	H
1c	H	CH_3	H	H	CH_3	H
1ca	H	H	H	H	CH_3	H
2c	H	CH_3	H	H	CH_3	H
2cb	H	CH_3	H	H	H	H
1d	H	CH_3	H	CH_3	CH_3	H
2d	H	CH_3	H	CH_3	CH_3	H
1e	$CH_3(CH_2)_3$	CH_3	H	H	CH_3	H

Compound number	R^1	R^2	R^3	R^4	R^5	R^6
1f	CH_3CH_2	CH_3	H	H	CH_3	F
1g	CH_3CH_2	CH_3	H	H	CH_3	Cl
1h	N_3	CH_3	H	H	CH_3	H
1ha	N_3	H	H	H	CH_3	H
1hb	N_3	CH_3	H	H	H	H
2h	N_3	CH_3	H	H	CH_3	H
2hb	N_3	CH_3	H	H	H	H
1i	N_3	CH_3	H	CH_3	CH_3	H
1j	N_3CH_2	CH_3	H	H	CH_3	H
1ja	N_3CH_2	H	H	H	CH_3	H
1jb	N_3CH_2	CH_3	H	H	H	H
2j	N_3CH_2	CH_3	H	H	CH_3	H
1k	N_3CH_2	CH_3	H	CH_3	CH_3	H
1l	$N_3CH_2CH_2$	CH_3	H	H	CH_3	H
1m	$N_3CH_2CH_2$	CH_3	H	CH_3	CH_3	H



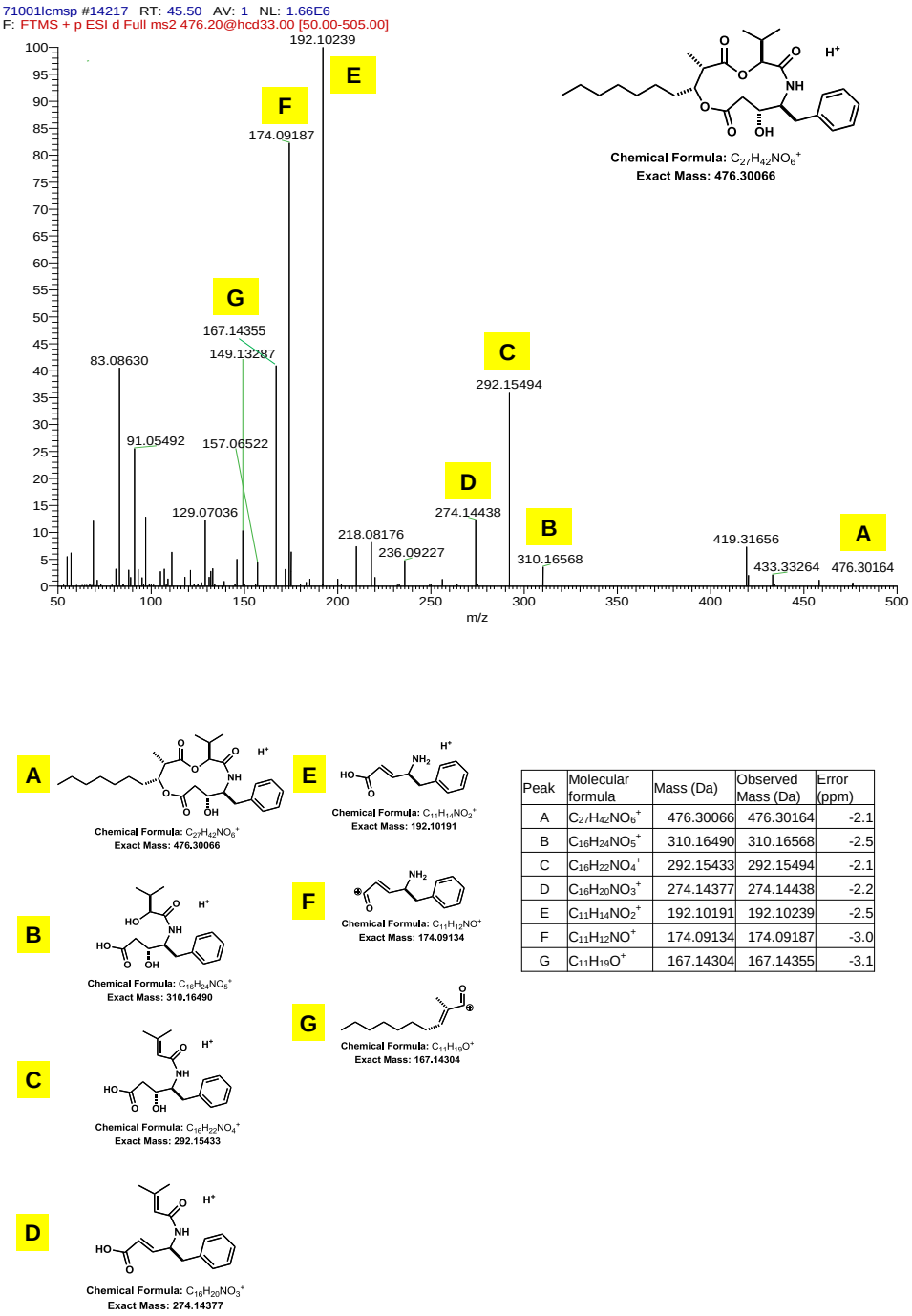
Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{28}H_{44}NO_6^+$	490.31631	490.31562	1.4
B	$C_{28}H_{42}NO_5^+$	472.30575	472.30501	1.6
C	$C_{17}H_{24}NO_4^+$	306.16998	306.16929	2.3
D	$C_{17}H_{24}NO_4^+$	324.18055	324.17997	1.8
E	$C_{17}H_{22}NO_3^+$	288.15942	288.15886	1.9
F	$C_{13}H_{16}NO_4^+$	250.10738	250.10698	1.6
G	$C_{12}H_{18}NO_3^+$	224.12812	224.12776	1.6
H	$C_{12}H_{16}NO_2^+$	206.11756	206.11726	1.5
I	$C_{12}H_{14}NO^+$	188.10699	188.10674	1.3
J	$C_{11}H_{19}O^+$	167.14304	167.14291	0.8

Figure D9 HR-MS-MS of compound **1a.**



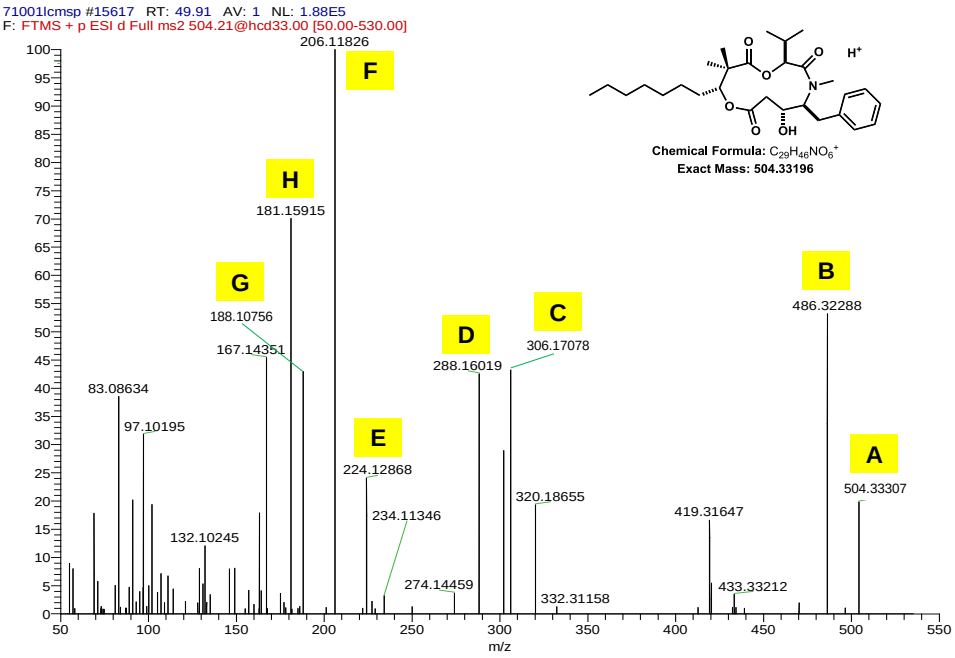
Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{27}H_{42}NO_6^+$	476.30066	476.29983	1.7
B	$C_{27}H_{40}NO_5^+$	458.29010	458.29094	-1.8
C	$C_{17}H_{24}NO_4^+$	306.16998	306.1707	-2.4
D	$C_{17}H_{22}NO_3^+$	288.15942	288.15997	-1.9
E	$C_{13}H_{16}NO_4^+$	250.10738	250.10758	-0.8
F	$C_{12}H_{18}NO_3^+$	224.12812	224.12866	-2.4
G	$C_{12}H_{16}NO_2^+$	206.11756	206.11815	-2.9
H	$C_{12}H_{14}NO^+$	188.10699	188.10754	-2.9
I	$C_{10}H_{17}O^+$	153.12739	153.12777	-2.5

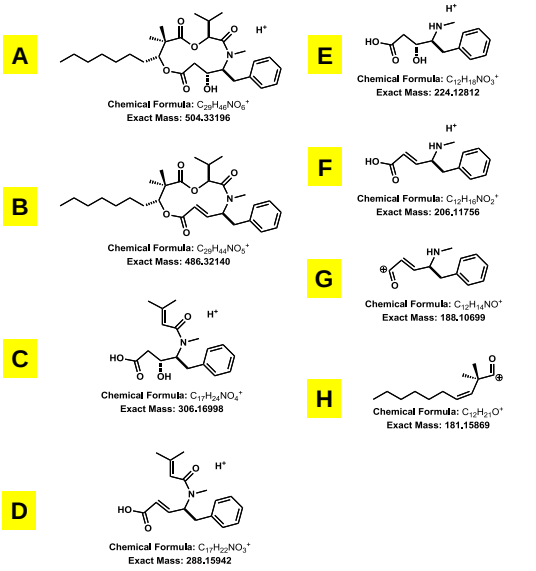
Figure D10 HR-MS-MS of compound **1aa**.



Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{27}H_{42}NO_6^+$	476.30066	476.30164	-2.1
B	$C_{16}H_{24}NO_5^+$	310.16490	310.16568	-2.5
C	$C_{16}H_{22}NO_4^+$	292.15433	292.15494	-2.1
D	$C_{16}H_{20}NO_3^+$	274.14377	274.14438	-2.2
E	$C_{11}H_{14}NO_2^+$	192.10191	192.10239	-2.5
F	$C_{11}H_{12}NO^+$	174.09134	174.09187	-3.0
G	$C_{11}H_{19}O^+$	167.14304	167.14355	-3.1

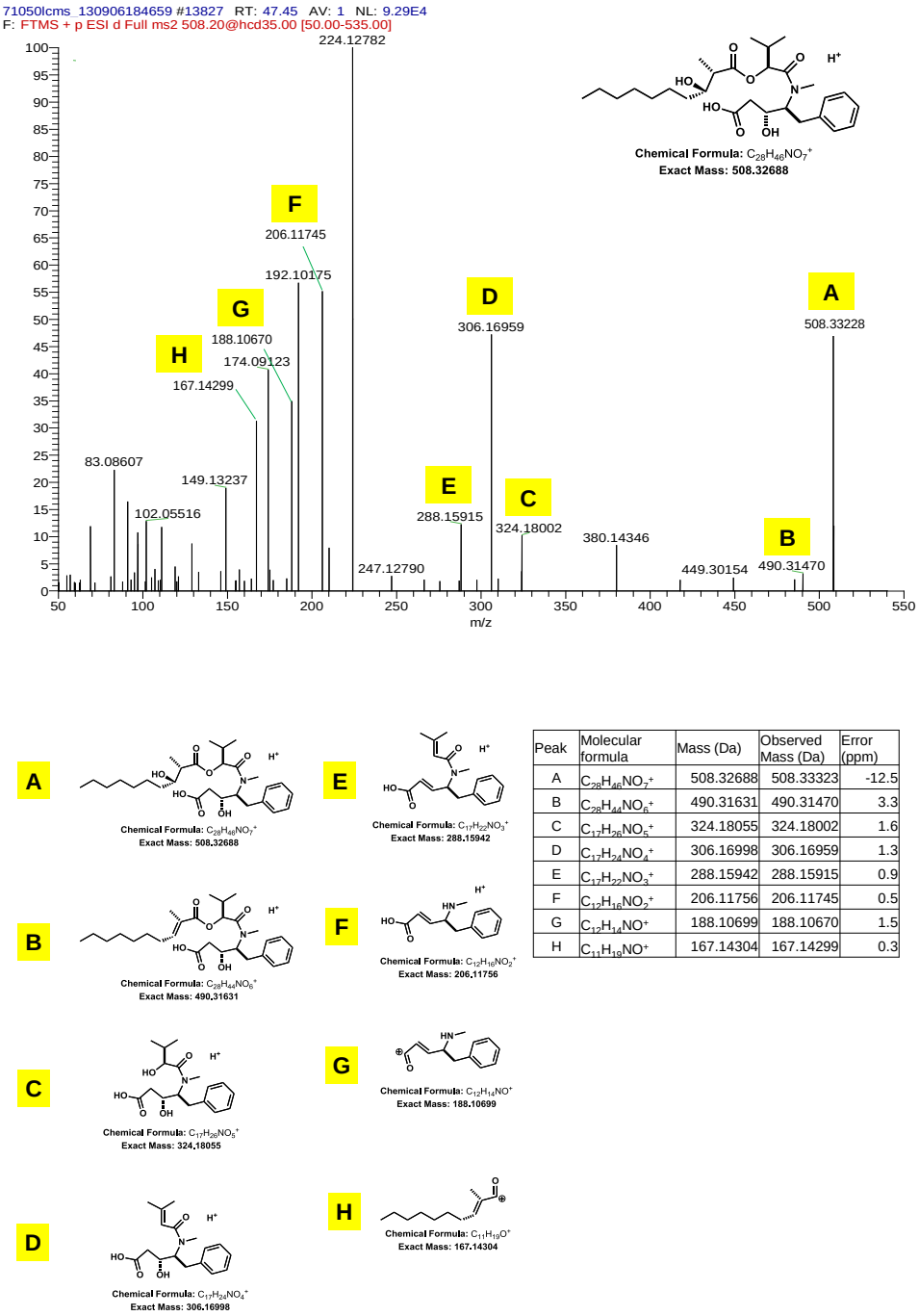
Figure D11 HR-MS-MS of compound **1ab**.





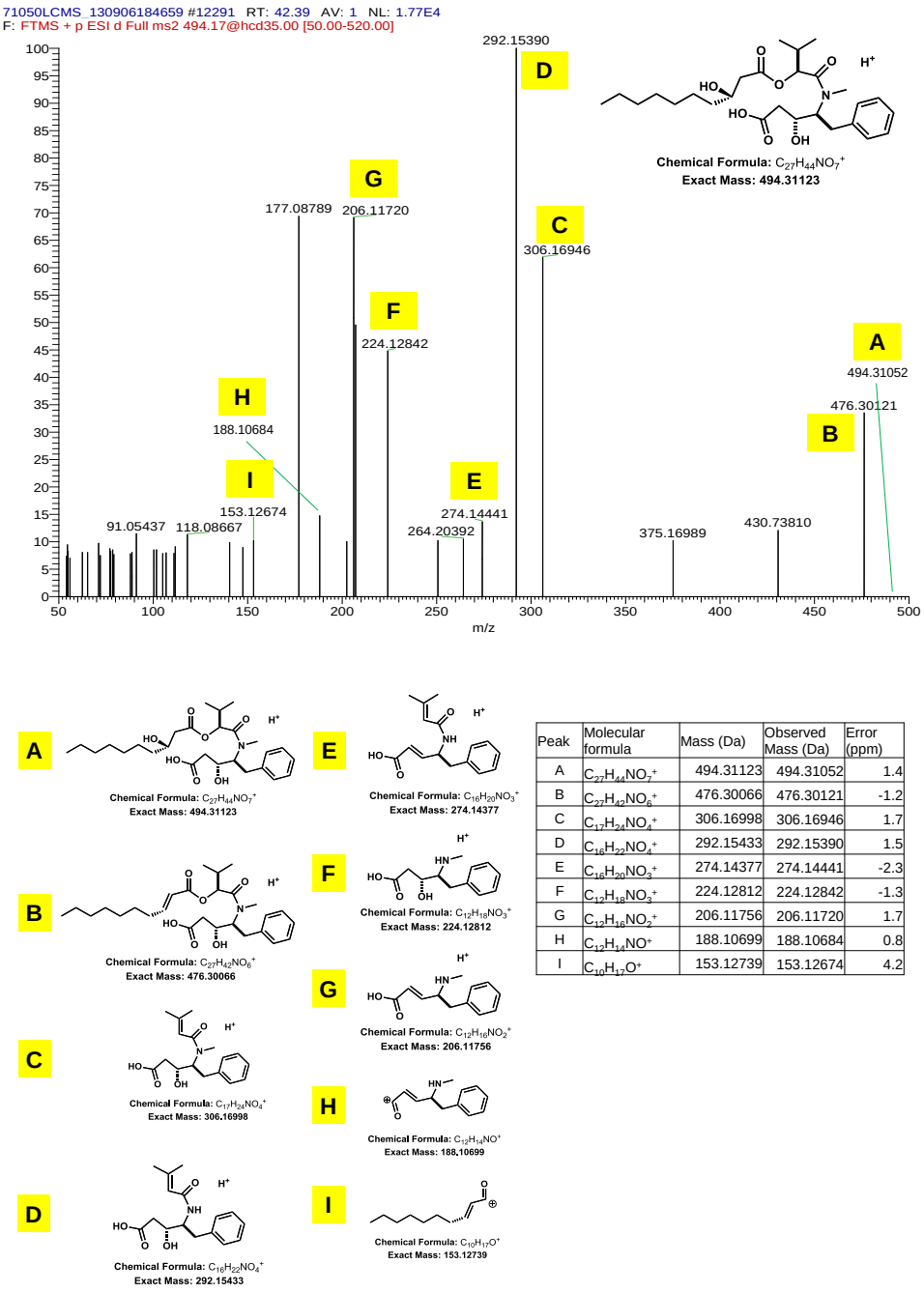
Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{29}H_{46}NO_6^+$	504.33196	504.33307	-2.2
B	$C_{29}H_{44}NO_5^+$	486.32140	486.32288	-3.0
C	$C_{17}H_{24}NO_4^+$	306.16998	306.17078	-2.6
D	$C_{17}H_{22}NO_3^+$	288.15942	288.16019	-2.7
E	$C_{12}H_{18}NO_3^+$	224.12812	224.12868	-2.5
F	$C_{12}H_{16}NO_2^+$	206.11756	206.11826	-3.4
G	$C_{12}H_{14}NO^+$	188.10699	188.10756	-3.0
H	$C_{12}H_{21}O^+$	181.15869	181.15915	-2.5

Figure D12 HR-MS-MS of compound **1ac**.



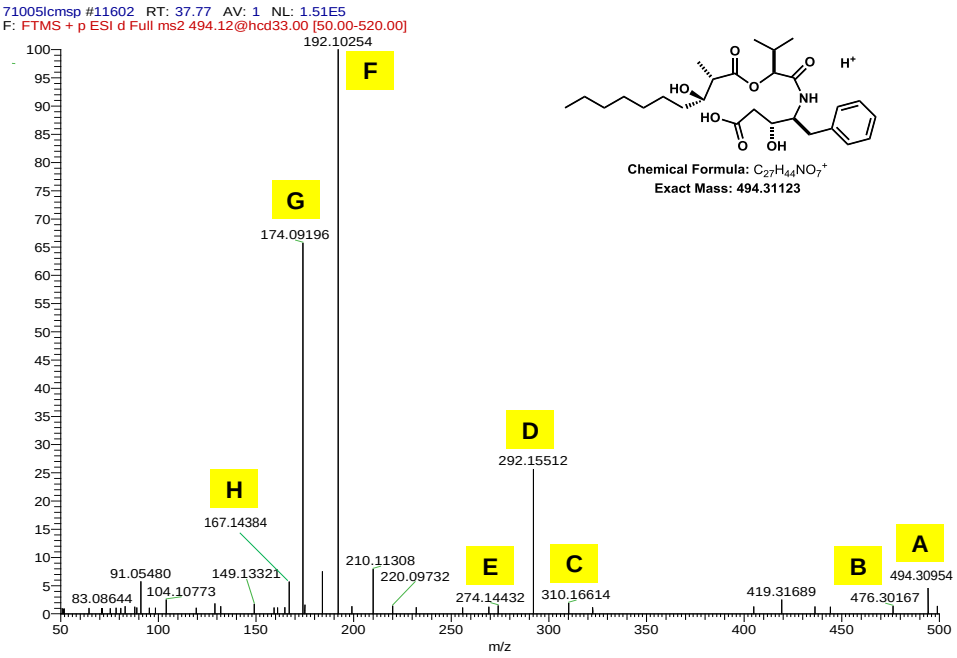
Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{28}H_{46}NO_7^+$	508.32688	508.33323	-12.5
B	$C_{28}H_{44}NO_6^+$	490.31631	490.31470	3.3
C	$C_{17}H_{26}NO_5^+$	324.18055	324.18002	1.6
D	$C_{17}H_{24}NO_4^+$	306.16998	306.16959	1.3
E	$C_{17}H_{22}NO_3^+$	288.15942	288.15915	0.9
F	$C_{12}H_{16}NO_2^+$	206.11756	206.11745	0.5
G	$C_{12}H_{14}NO^+$	188.10699	188.10670	1.5
H	$C_{11}H_{19}NO^+$	167.14304	167.14299	0.3

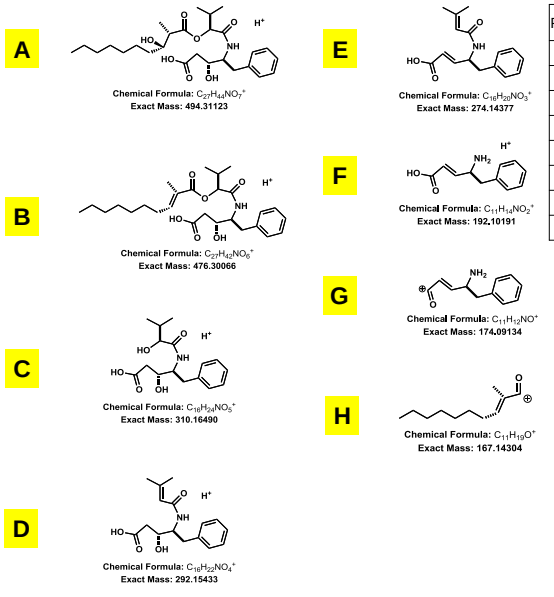
Figure D13 HR-MS-MS of compound **2a**.



Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{27}H_{44}NO_7^+$	494.31123	494.31052	1.4
B	$C_{27}H_{42}NO_6^+$	476.30066	476.30121	-1.2
C	$C_{17}H_{24}NO_4^+$	306.16998	306.16946	1.7
D	$C_{16}H_{22}NO_4^+$	292.15433	292.15390	1.5
E	$C_{16}H_{20}NO_3^+$	274.14377	274.14441	-2.3
F	$C_{12}H_{18}NO_3^+$	224.12812	224.12842	-1.3
G	$C_{12}H_{16}NO_2^+$	206.11756	206.11720	1.7
H	$C_{12}H_{14}NO^+$	188.10699	188.10684	0.8
I	$C_{10}H_{17}O^+$	153.12739	153.12674	4.2

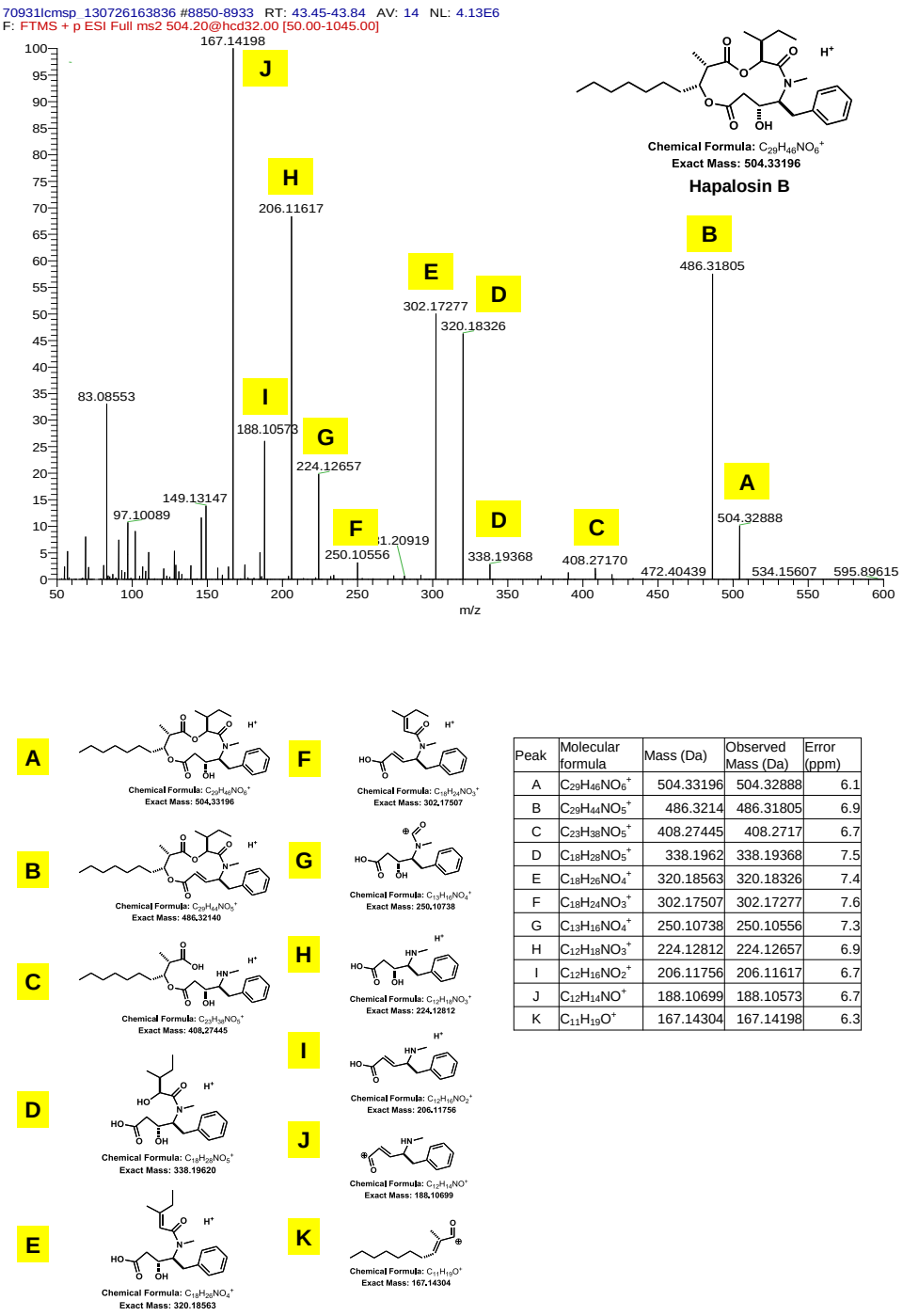
Figure D14 HR-MS-MS of compound **2aa**.





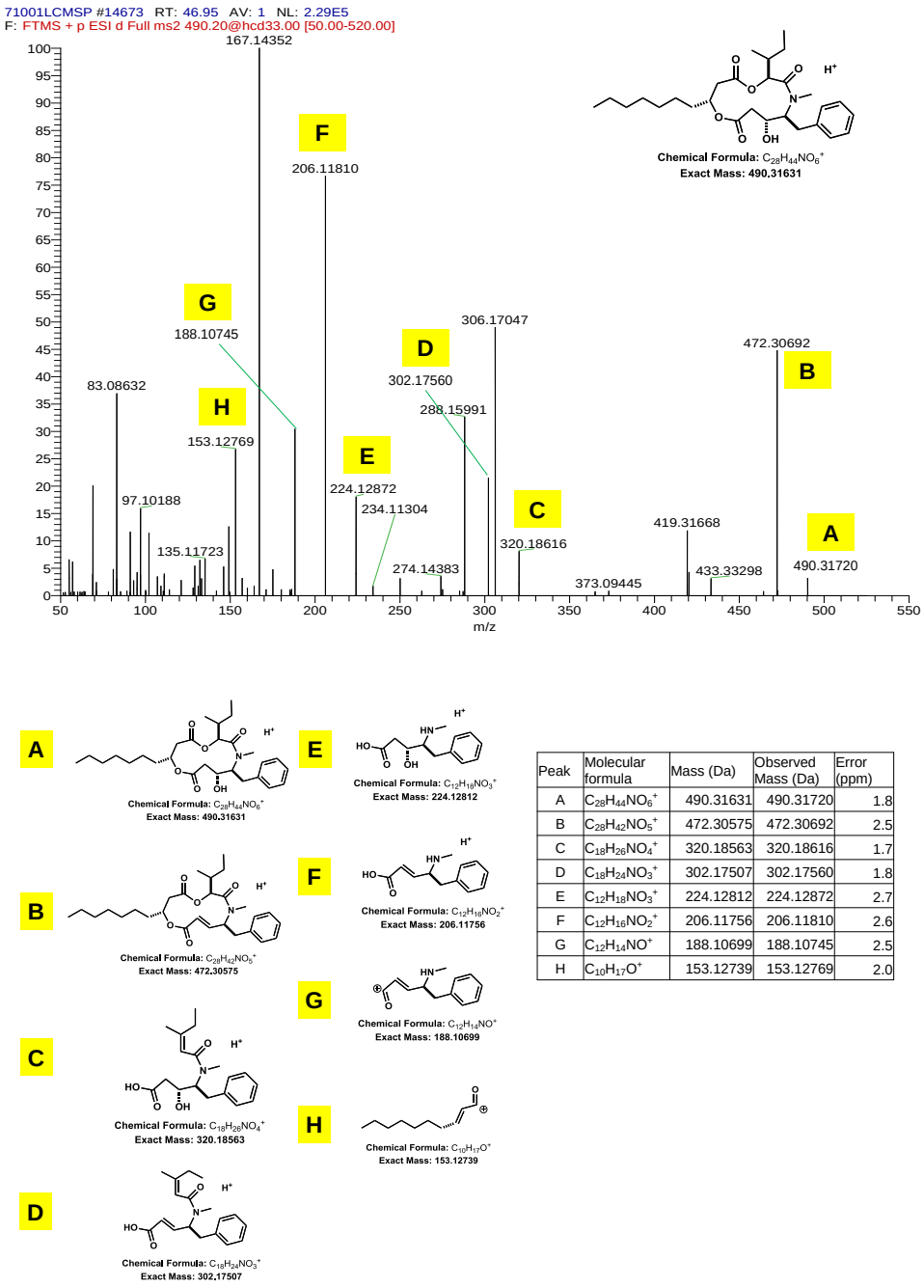
Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{27}H_{44}NO_7^+$	494.31123	494.30954	3.4
B	$C_{27}H_{42}NO_6^+$	476.30066	476.30167	-2.1
C	$C_{16}H_{24}NO_5^+$	310.16490	310.16610	-3.9
D	$C_{16}H_{22}NO_4^+$	292.15433	292.15512	-2.7
E	$C_{16}H_{20}NO_3^+$	274.14377	274.14432	-2.0
F	$C_{11}H_{14}NO_2^+$	192.10191	192.10254	-3.3
G	$C_{11}H_{12}NO^+$	174.09134	174.09196	-3.6
H	$C_{11}H_{19}O^+$	167.14304	167.14384	-4.8

Figure D15 HR-MS-MS of compound **2ab**.



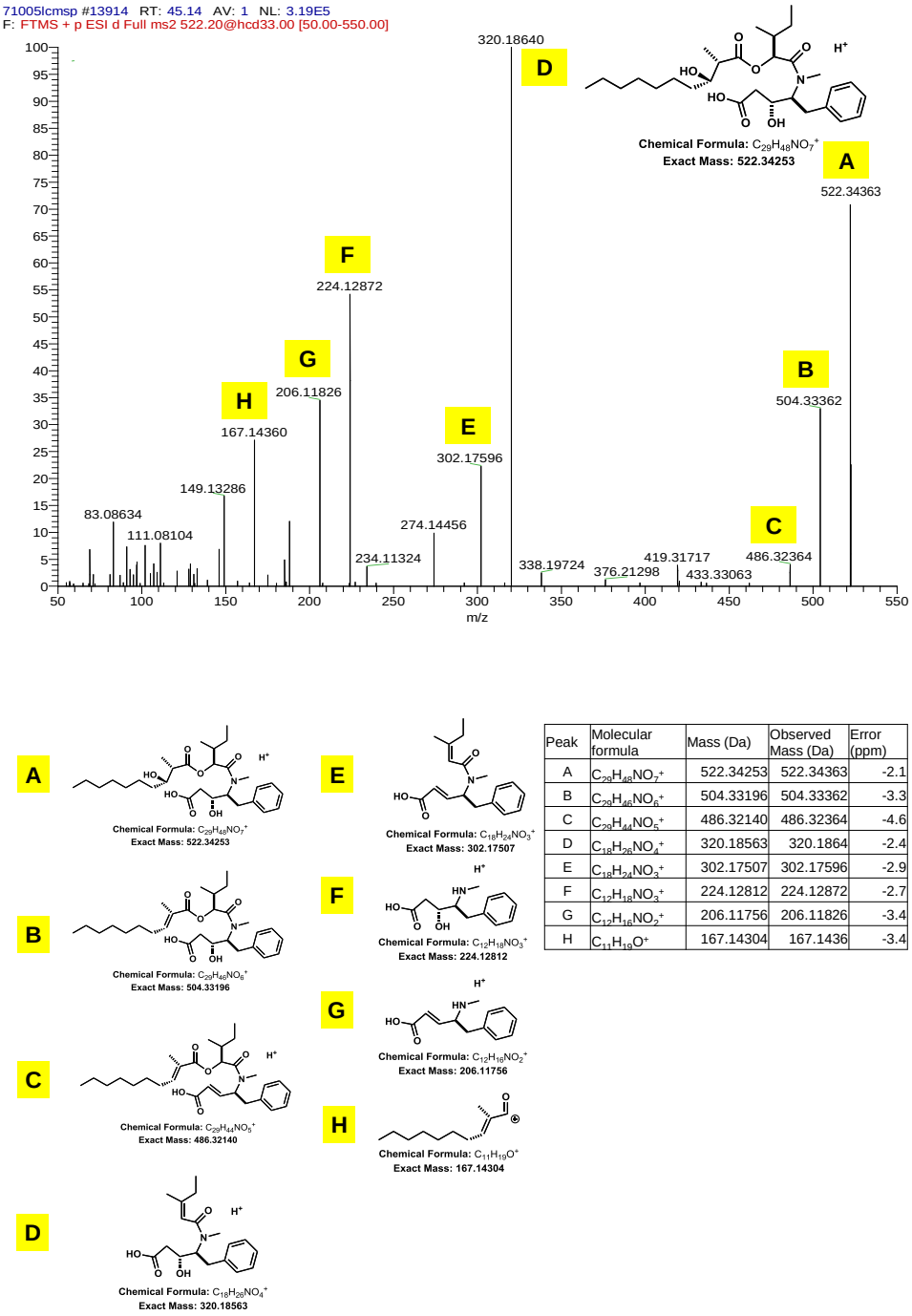
Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{29}H_{46}NO_6^+$	504.33196	504.32888	6.1
B	$C_{29}H_{44}NO_5^+$	486.3214	486.31805	6.9
C	$C_{23}H_{38}NO_5^+$	408.27445	408.2717	6.7
D	$C_{18}H_{28}NO_5^+$	338.1962	338.19368	7.5
E	$C_{18}H_{26}NO_4^+$	320.18563	320.18326	7.4
F	$C_{18}H_{24}NO_3^+$	302.17507	302.17277	7.6
G	$C_{13}H_{16}NO_4^+$	250.10738	250.10556	7.3
H	$C_{12}H_{18}NO_3^+$	224.12812	224.12657	6.9
I	$C_{12}H_{16}NO_2^+$	206.11756	206.11617	6.7
J	$C_{12}H_{14}NO^+$	188.10699	188.10573	6.7
K	$C_{11}H_{19}O^+$	167.14304	167.14198	6.3

Figure D16 HR-MS-MS of compound **1b**.



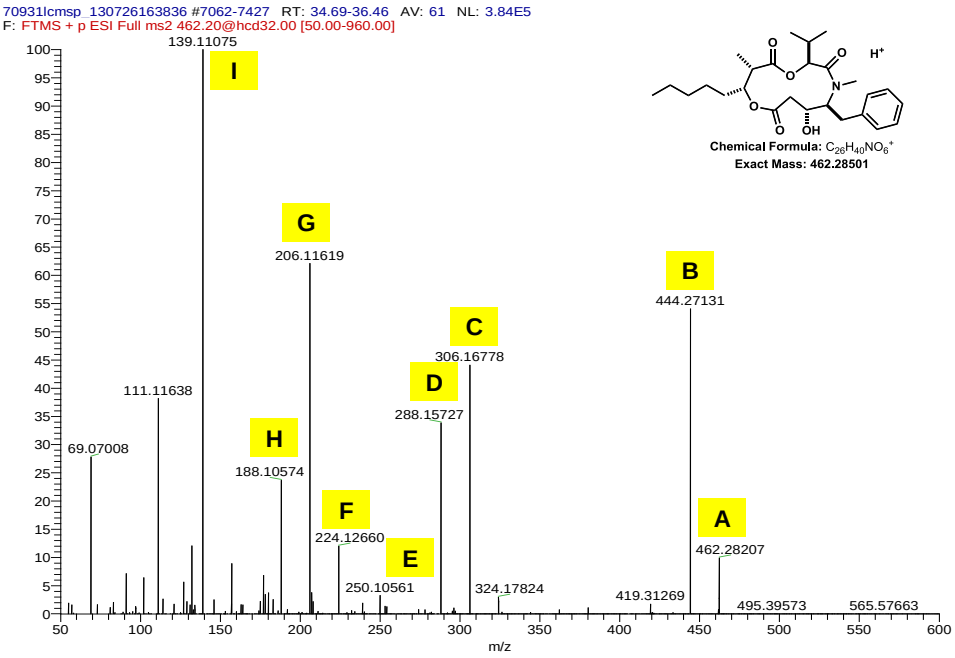
Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{28}H_{44}NO_6^+$	490.31631	490.31720	1.8
B	$C_{28}H_{42}NO_5^+$	472.30575	472.30692	2.5
C	$C_{18}H_{26}NO_4^+$	320.18563	320.18616	1.7
D	$C_{18}H_{24}NO_3^+$	302.17507	302.17560	1.8
E	$C_{12}H_{18}NO_3^+$	224.12812	224.12872	2.7
F	$C_{12}H_{16}NO_2^+$	206.11756	206.11810	2.6
G	$C_{12}H_{14}NO^+$	188.10699	188.10745	2.5
H	$C_{10}H_{17}O^+$	153.12739	153.12769	2.0

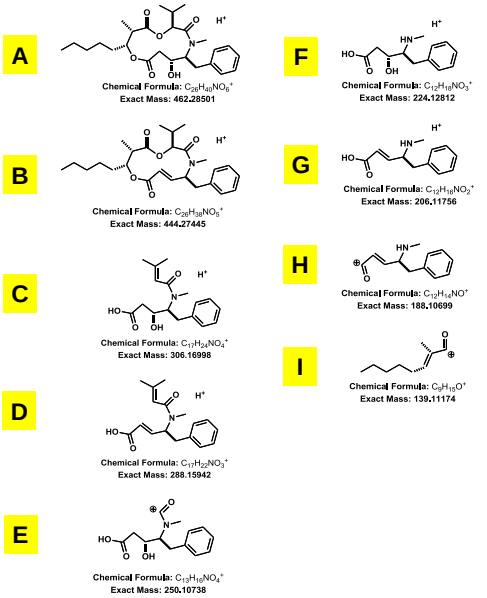
Figure D17 HR-MS-MS of compound **1ba**.



Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{29}H_{48}NO_7^+$	522.34253	522.34363	-2.1
B	$C_{29}H_{46}NO_6^+$	504.33196	504.33362	-3.3
C	$C_{29}H_{44}NO_5^+$	486.32140	486.32364	-4.6
D	$C_{18}H_{26}NO_4^+$	320.18563	320.1864	-2.4
E	$C_{18}H_{24}NO_3^+$	302.17507	302.17596	-2.9
F	$C_{12}H_{18}NO_3^+$	224.12812	224.12872	-2.7
G	$C_{12}H_{16}NO_2^+$	206.11756	206.11826	-3.4
H	$C_{11}H_{19}O^+$	167.14304	167.1436	-3.4

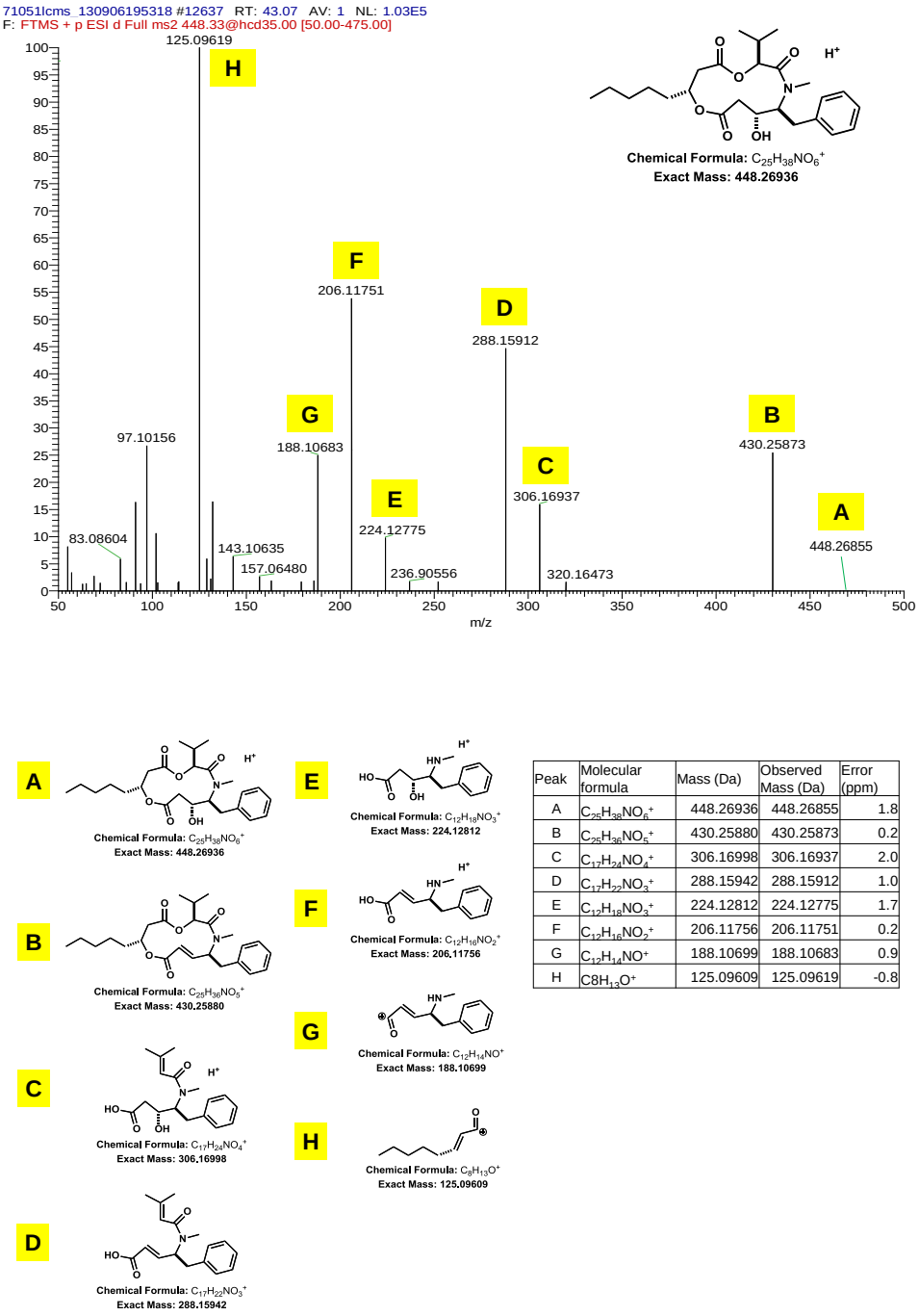
Figure D18 HR-MS-MS of compound **2b**.





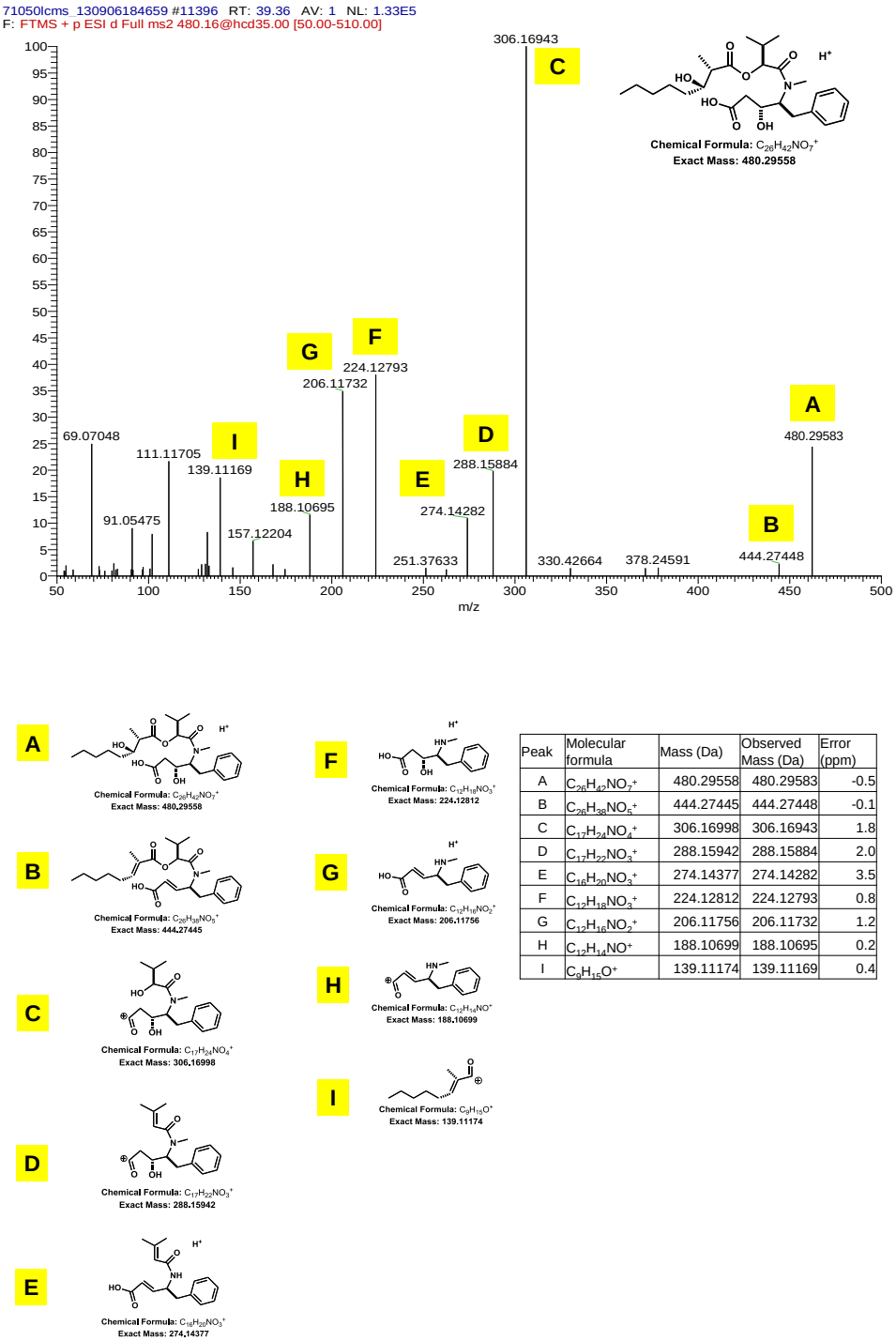
Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{26}H_{40}NO_6^+$	462.28501	462.28207	6.4
B	$C_{26}H_{38}NO_5^+$	444.27445	444.27131	7.1
C	$C_{17}H_{24}NO_4^+$	306.16998	306.16778	7.2
D	$C_{17}H_{22}NO_3^+$	288.15942	288.15727	7.5
E	$C_{13}H_{16}NO_4^+$	250.10738	250.10561	7.1
F	$C_{12}H_{18}NO_3^+$	224.12812	224.1266	6.8
G	$C_{12}H_{16}NO_2^+$	206.11756	206.11619	6.6
H	$C_{12}H_{14}NO^+$	188.10699	188.10574	6.6
I	$C_9H_{15}O^+$	139.11174	139.11075	7.1

Figure D19 HR-MS-MS of compound **1c**.



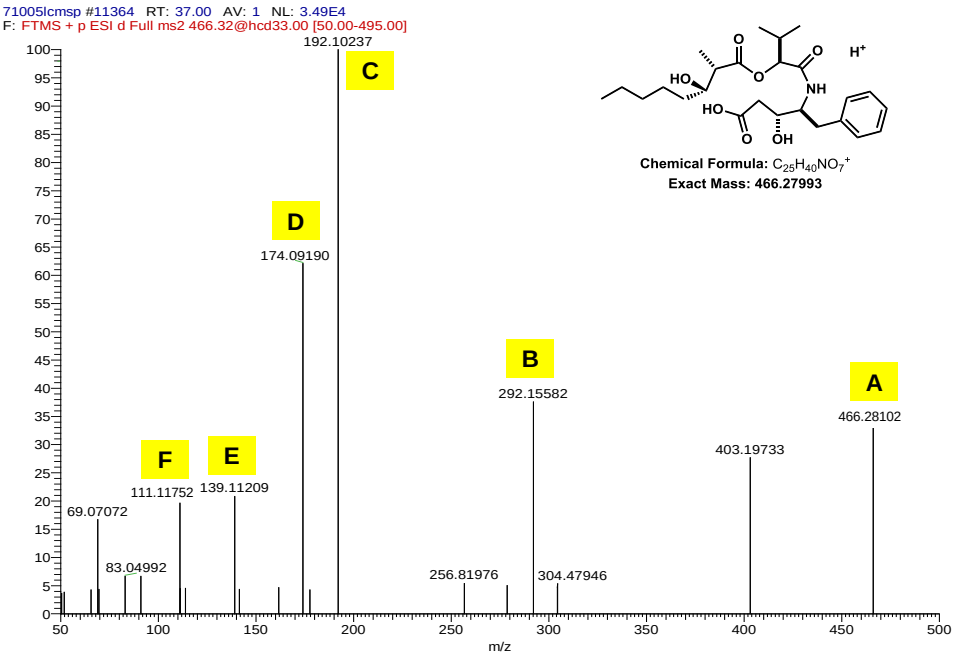
Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{25}H_{38}NO_6^+$	448.26936	448.26855	1.8
B	$C_{25}H_{36}NO_5^+$	430.25880	430.25873	0.2
C	$C_{17}H_{24}NO_4^+$	306.16998	306.16937	2.0
D	$C_{17}H_{22}NO_3^+$	288.15942	288.15912	1.0
E	$C_{12}H_{18}NO_3^+$	224.12812	224.12775	1.7
F	$C_{12}H_{16}NO_2^+$	206.11756	206.11751	0.2
G	$C_{12}H_{14}NO^+$	188.10699	188.10683	0.9
H	$C8H_{13}O^+$	125.09609	125.09619	-0.8

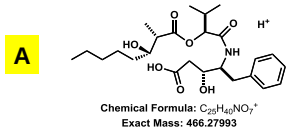
Figure D20 HR-MS-MS of compound **1ca**.

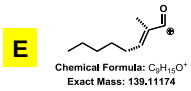


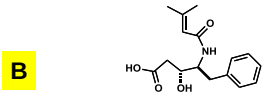
Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{26}H_{42}NO_7^+$	480.29558	480.29583	-0.5
B	$C_{26}H_{38}NO_5^+$	444.27445	444.27448	-0.1
C	$C_{17}H_{24}NO_4^+$	306.16998	306.16943	1.8
D	$C_{17}H_{22}NO_3^+$	288.15942	288.15884	2.0
E	$C_{16}H_{20}NO_3^+$	274.14377	274.14282	3.5
F	$C_{12}H_{18}NO_3^+$	224.12812	224.12793	0.8
G	$C_{12}H_{16}NO_2^+$	206.11756	206.11732	1.2
H	$C_{12}H_{14}NO^+$	188.10699	188.10695	0.2
I	$C_9H_{15}O^+$	139.11174	139.11169	0.4

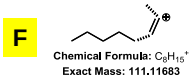
Figure D21 HR-MS-MS of compound **2c**.

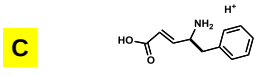


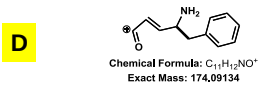






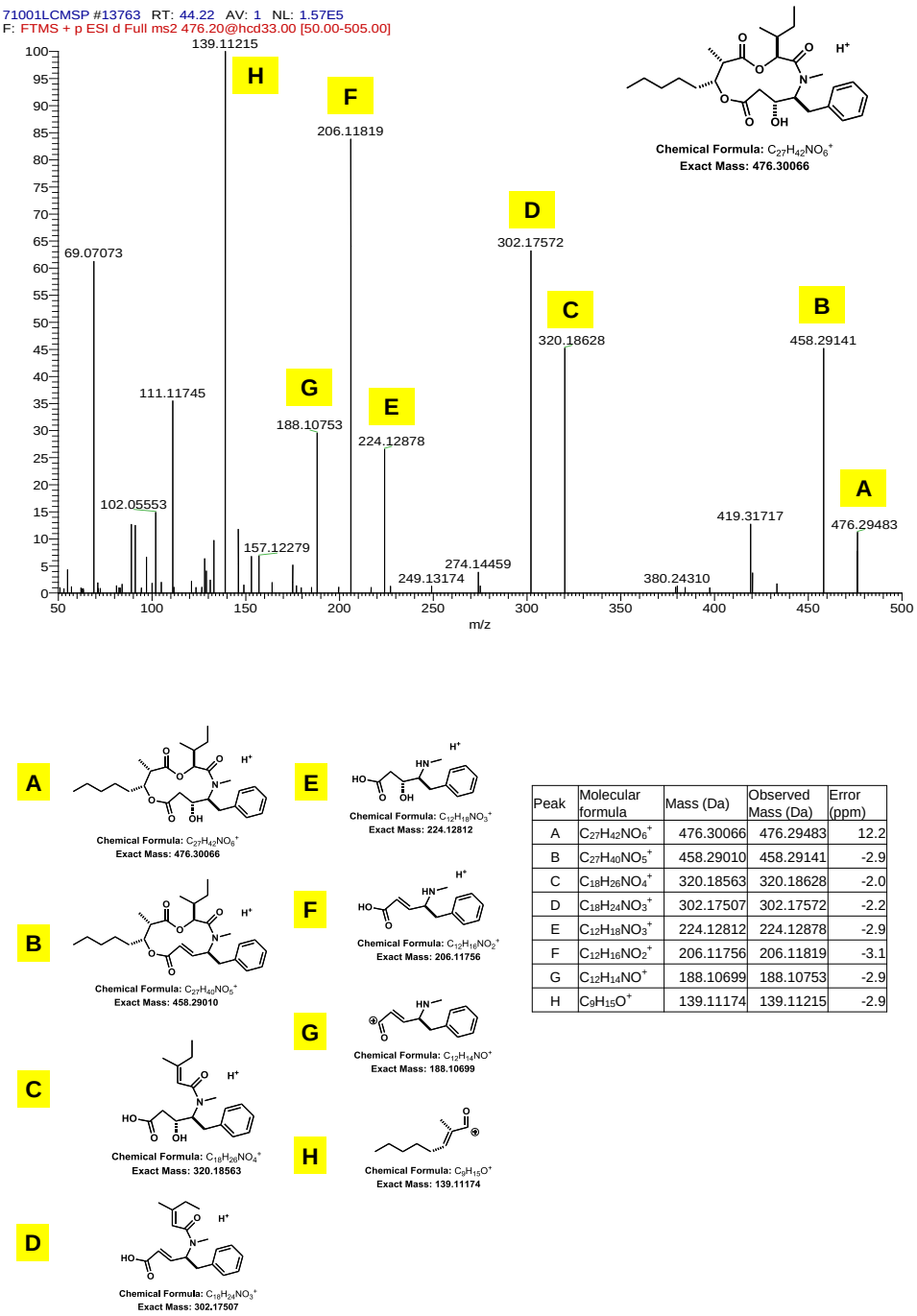






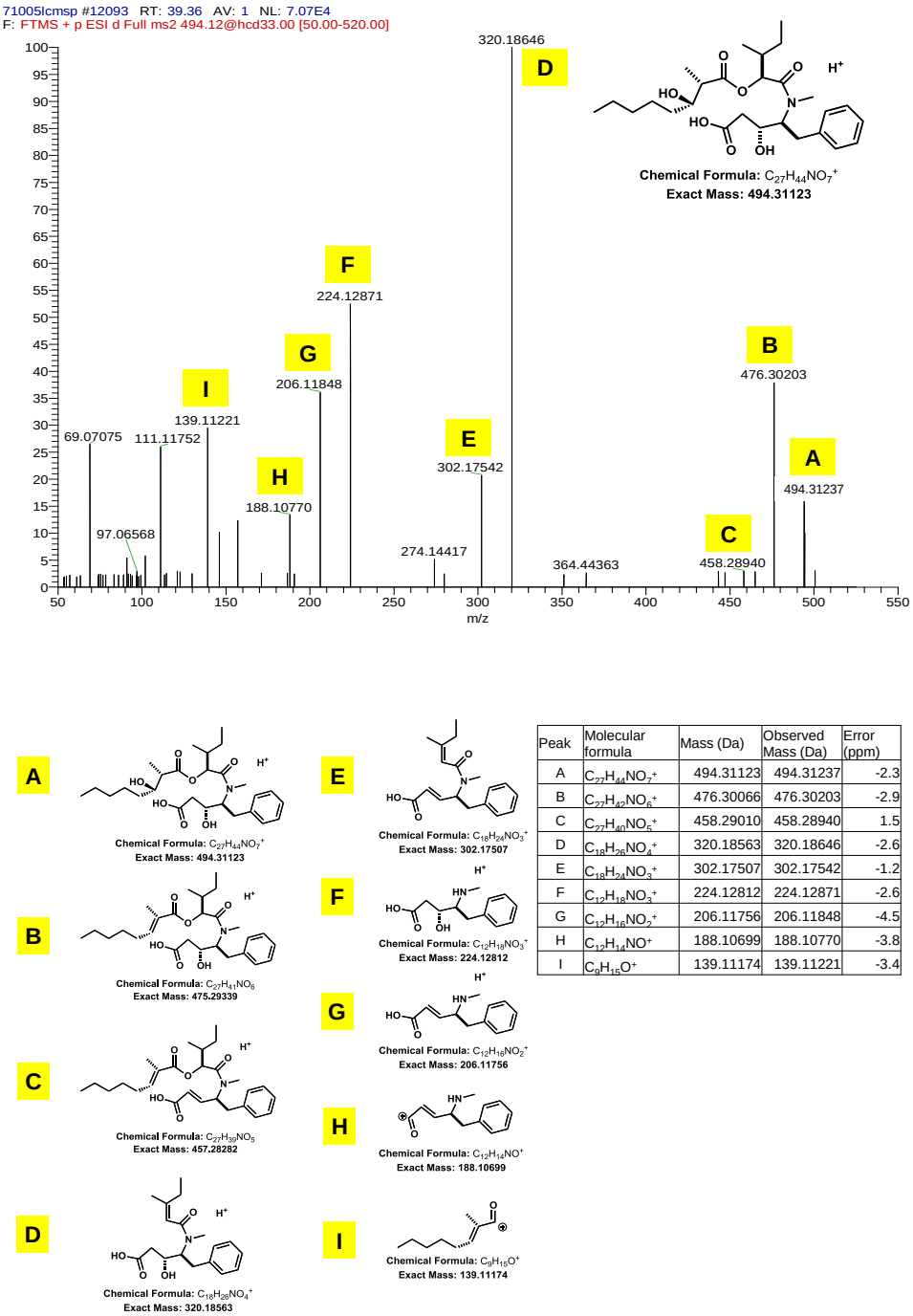
Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{25}H_{40}NO_7^+$	466.27993	466.28102	-2.3
B	$C_{16}H_{22}NO_4^+$	292.15433	292.15582	-5.1
C	$C_{11}H_{14}NO_2^+$	192.10191	192.10237	-2.4
D	$C_{11}H_{12}NO^+$	174.09134	174.0919	-3.2
E	$C_8H_{15}O^+$	139.11174	139.11209	-2.5
F	$C_8H_{15}^+$	111.11683	111.11752	-6.2

Figure D22 HR-MS-MS of compound **2cb**.



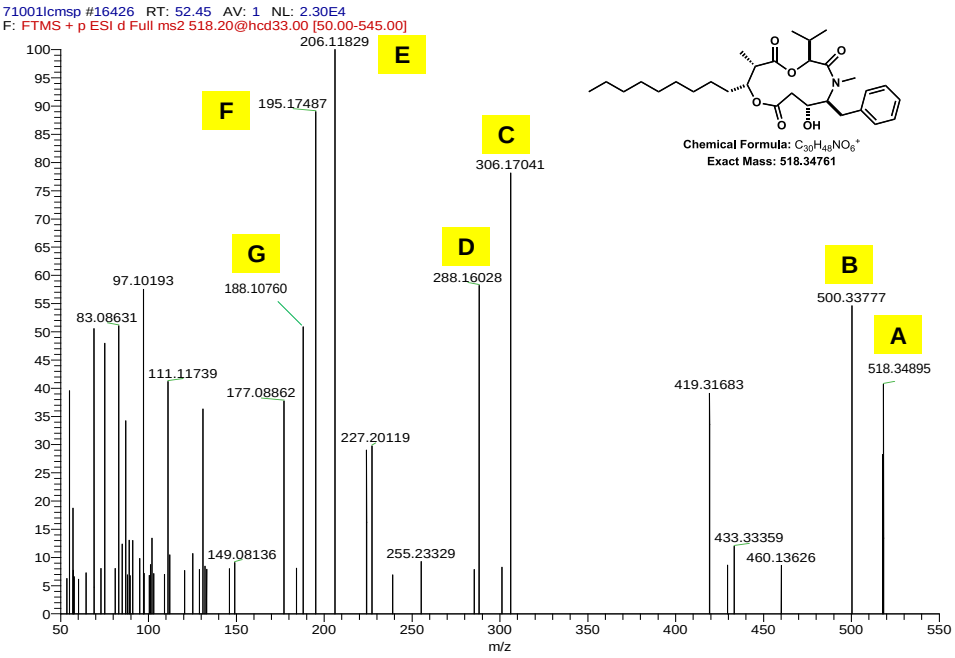
Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{27}H_{42}NO_6^+$	476.30066	476.29483	12.2
B	$C_{27}H_{40}NO_5^+$	458.29010	458.29141	-2.9
C	$C_{18}H_{26}NO_4^+$	320.18563	320.18628	-2.0
D	$C_{18}H_{24}NO_3^+$	302.17507	302.17572	-2.2
E	$C_{12}H_{18}NO_3^+$	224.12812	224.12878	-2.9
F	$C_{12}H_{16}NO_2^+$	206.11756	206.11819	-3.1
G	$C_{12}H_{14}NO^+$	188.10699	188.10753	-2.9
H	$C_9H_{15}O^+$	139.11174	139.11215	-2.9

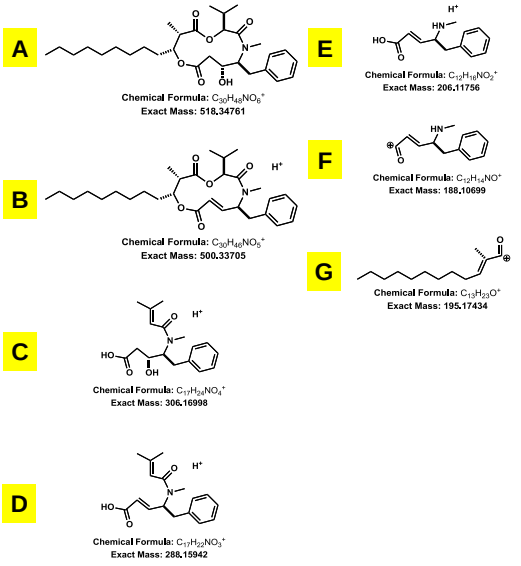
Figure D23 HR-MS-MS of compound **1d.**



Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{27}H_{44}NO_7^+$	494.31123	494.31237	-2.3
B	$C_{27}H_{42}NO_6^+$	476.30066	476.30203	-2.9
C	$C_{27}H_{40}NO_5^+$	458.29010	458.28940	1.5
D	$C_{18}H_{26}NO_4^+$	320.18563	320.18646	-2.6
E	$C_{18}H_{24}NO_3^+$	302.17507	302.17542	-1.2
F	$C_{12}H_{18}NO_3^+$	224.12812	224.12871	-2.6
G	$C_{12}H_{16}NO_2^+$	206.11756	206.11848	-4.5
H	$C_{12}H_{14}NO^+$	188.10699	188.10770	-3.8
I	$C_9H_{15}O^+$	139.11174	139.11221	-3.4

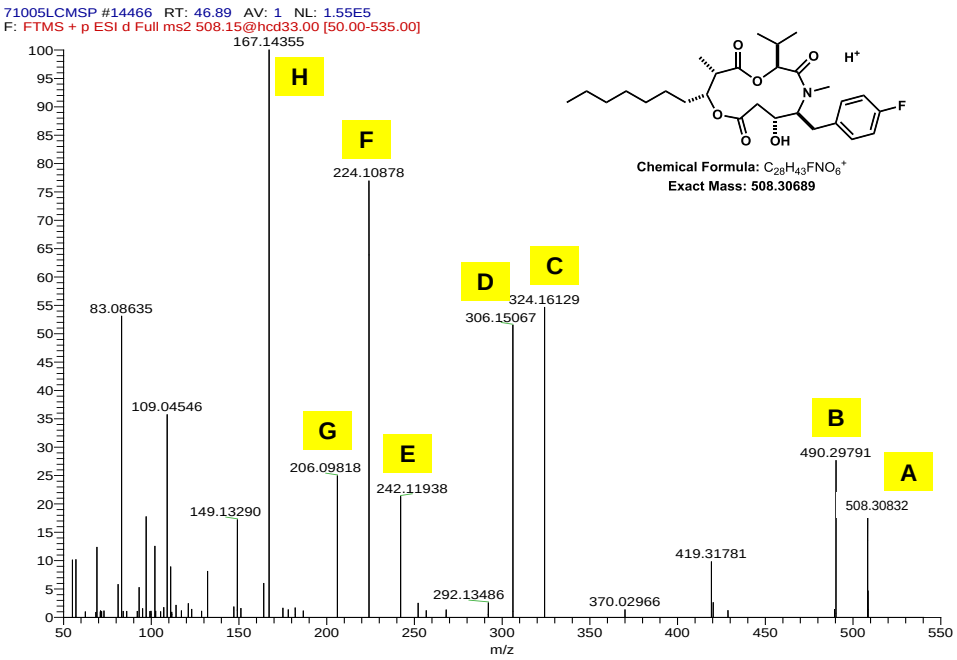
Figure D24 HR-MS-MS of compound **2d**.

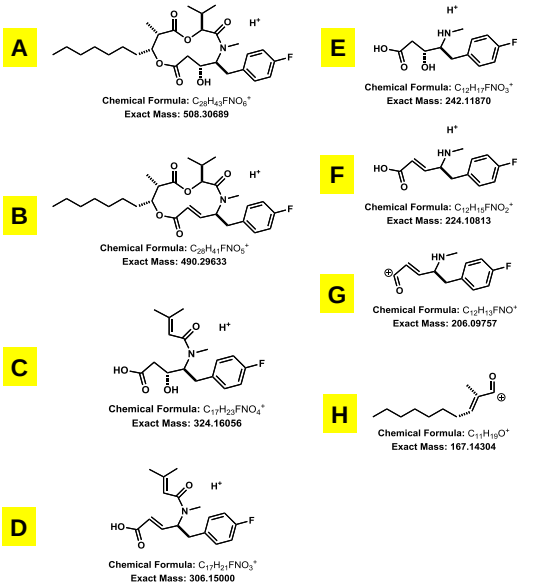




Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{30}H_{48}NO_6^+$	518.34761	518.34895	-2.6
B	$C_{30}H_{46}NO_5^+$	500.33705	500.33777	-1.4
C	$C_{17}H_{24}NO_4^+$	306.16998	306.17041	-1.4
D	$C_{17}H_{22}NO_3^+$	288.15942	288.16028	-3.0
E	$C_{12}H_{16}NO_2^+$	206.11756	206.11829	-3.5
F	$C_{12}H_{14}NO^+$	195.17343	195.17487	-7.4
G	$C_{13}H_{23}O^+$	188.10699	188.1076	-3.2

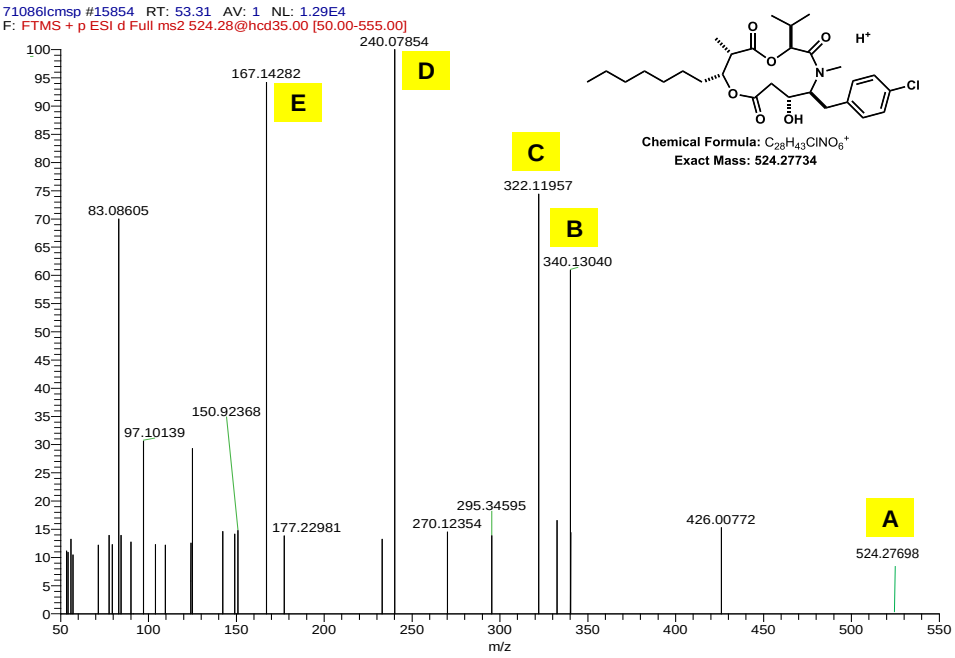
Figure D25 HR-MS-MS of compound **1e**.

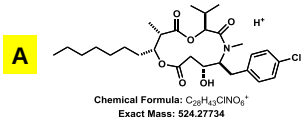


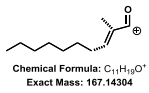


Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{28}H_{43}FNO_6^+$	508.30689	508.30832	-2.8
B	$C_{28}H_{41}FNO_5^+$	490.29633	490.29791	-2.3
C	$C_{17}H_{23}FNO_4^+$	324.16056	324.16129	-2.2
D	$C_{17}H_{21}FNO_3^+$	306.15000	306.15067	-2.2
E	$C_{12}H_{17}FNO_3^+$	242.1187	242.11938	-2.8
F	$C_{12}H_{15}FNO_2^+$	224.10813	224.10878	-2.9
G	$C_{12}H_{13}FNO^+$	206.09757	206.09818	-3.0
G	$C_{11}H_{19}N^+$	167.14304	167.14355	-3.1

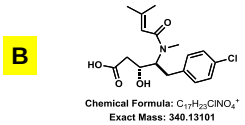
Figure D26 HR-MS-MS of compound **1f**.

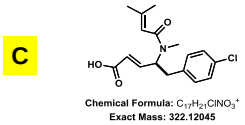






Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{28}H_{43}ClNO_6^+$	524.27734	524.27698	0.7
B	$C_{17}H_{23}ClNO_4^+$	340.13101	340.13040	1.8
C	$C_{17}H_{21}ClNO_3^+$	322.12045	322.11957	2.7
D	$C_{12}H_{15}ClNO_2^+$	240.07858	240.07854	0.2
E	$C_{11}H_{19}O^+$	167.14304	167.14282	1.3





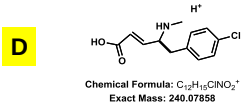
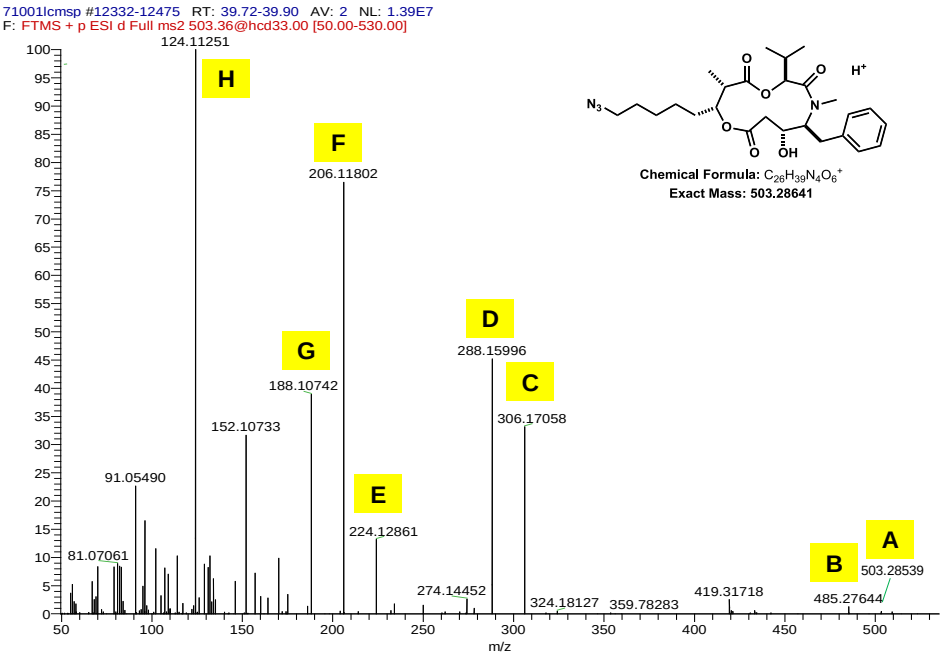
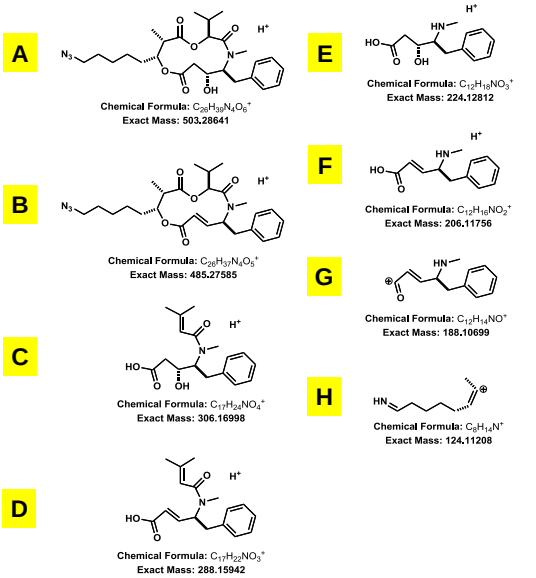


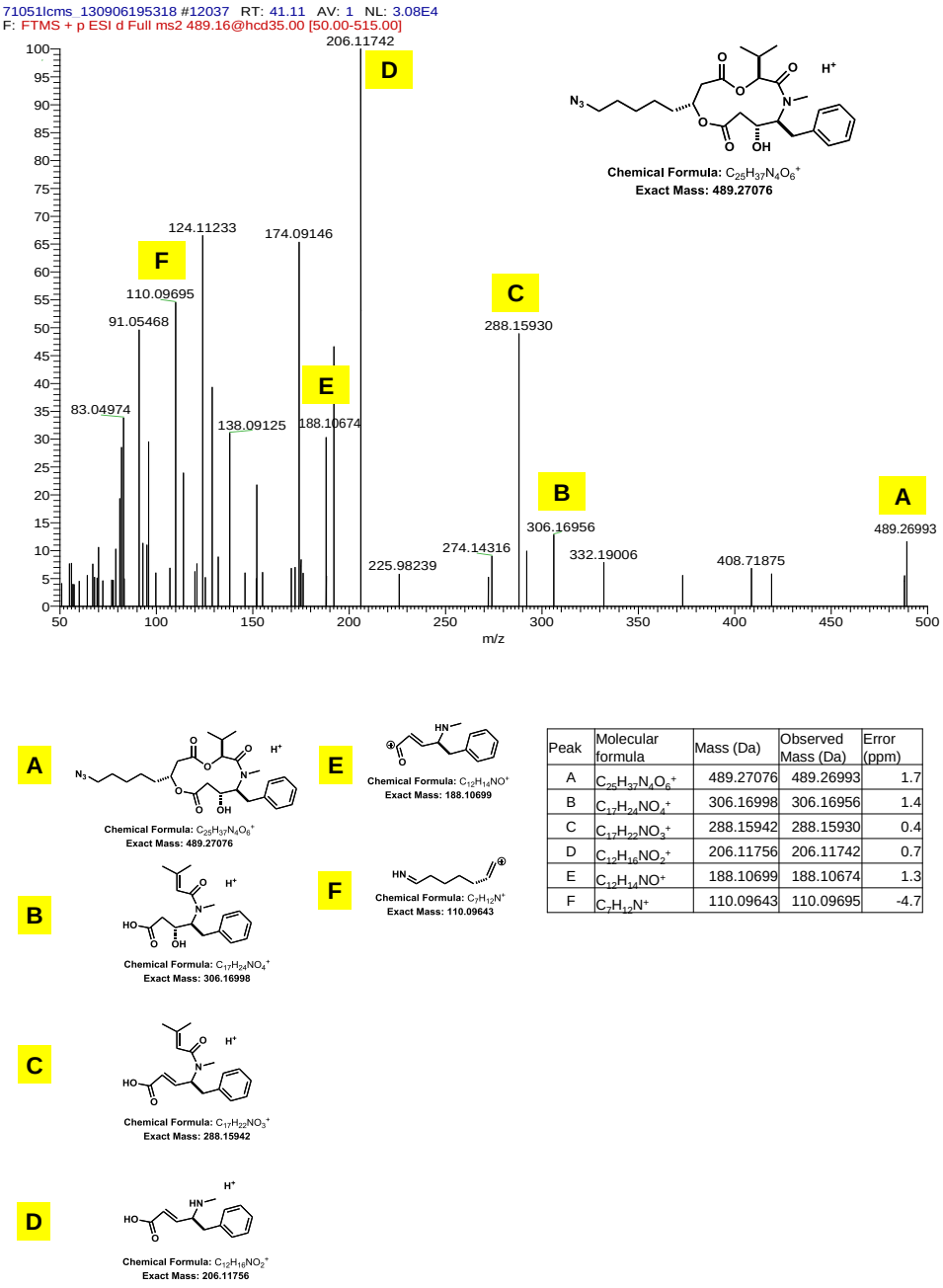
Figure D27 HR-MS-MS of compound **1g**.





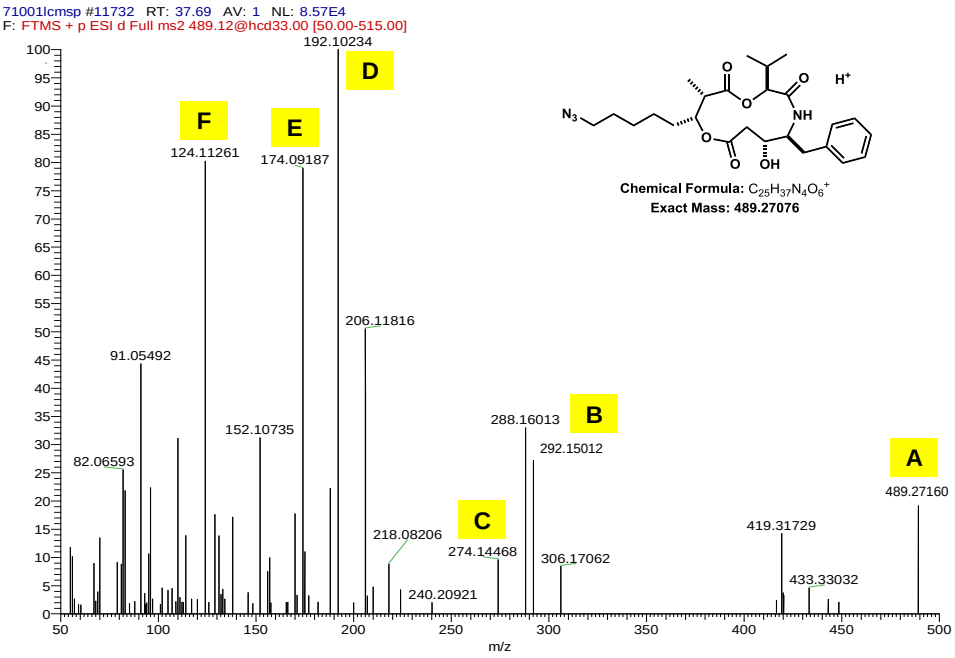
Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{26}H_{39}N_4O_6^+$	503.28641	503.28539	2.0
B	$C_{26}H_{37}N_4O_5^+$	485.27585	485.27644	-1.2
C	$C_{17}H_{24}NO_4^+$	306.16998	306.17058	-2.0
D	$C_{17}H_{22}NO_3^+$	288.15942	288.15996	-1.9
E	$C_{12}H_{18}NO_3^+$	224.12812	224.12861	-2.2
F	$C_{12}H_{16}NO_2^+$	206.11756	206.11802	-2.2
G	$C_{12}H_{14}NO^+$	188.10699	188.10742	-2.3
H	$C_8H_{14}N^+$	124.11208	124.11251	-3.5

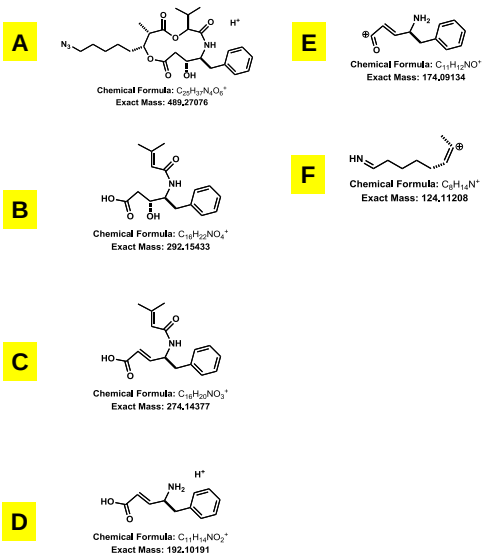
Figure D28 HR-MS-MS of compound **1h**.



Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{25}H_{37}N_4O_6^+$	489.27076	489.26993	1.7
B	$C_{17}H_{24}NO_4^+$	306.16998	306.16956	1.4
C	$C_{17}H_{22}NO_3^+$	288.15942	288.15930	0.4
D	$C_{12}H_{16}NO_2^+$	206.11756	206.11742	0.7
E	$C_{12}H_{14}NO^+$	188.10699	188.10674	1.3
F	$C_7H_{12}N^+$	110.09643	110.09695	-4.7

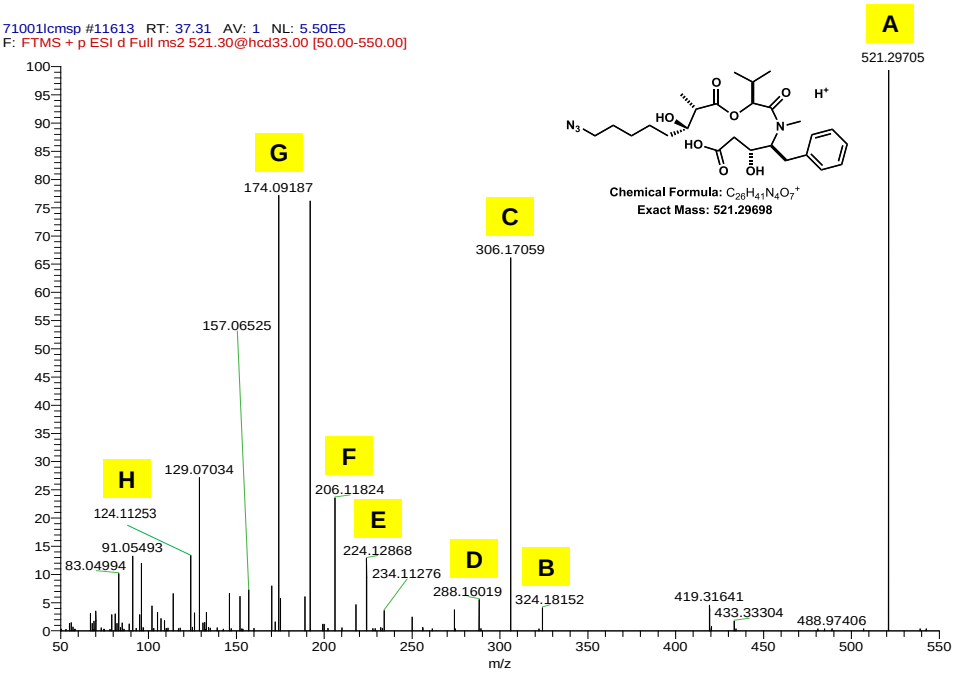
Figure D29 HR-MS-MS of compound **1ha**.

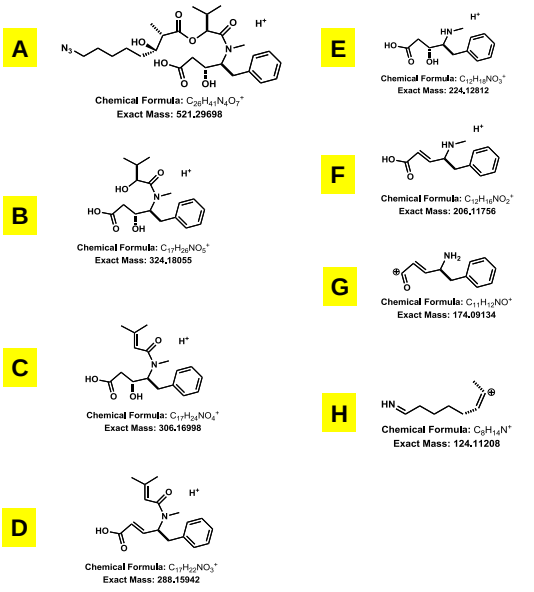




Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{25}H_{37}N_4O_6^+$	489.27076	489.2716	-1.7
B	$C_{16}H_{22}NO_4^+$	292.15433	292.15012	14.4
C	$C_{16}H_{20}NO_3^+$	274.14377	274.14468	-3.3
D	$C_{11}H_{14}NO_2^+$	192.10191	192.10234	-2.2
E	$C_{11}H_{12}NO^+$	174.09134	174.09187	-3.0
F	$C_8H_{14}N^+$	124.11208	124.11261	-4.3

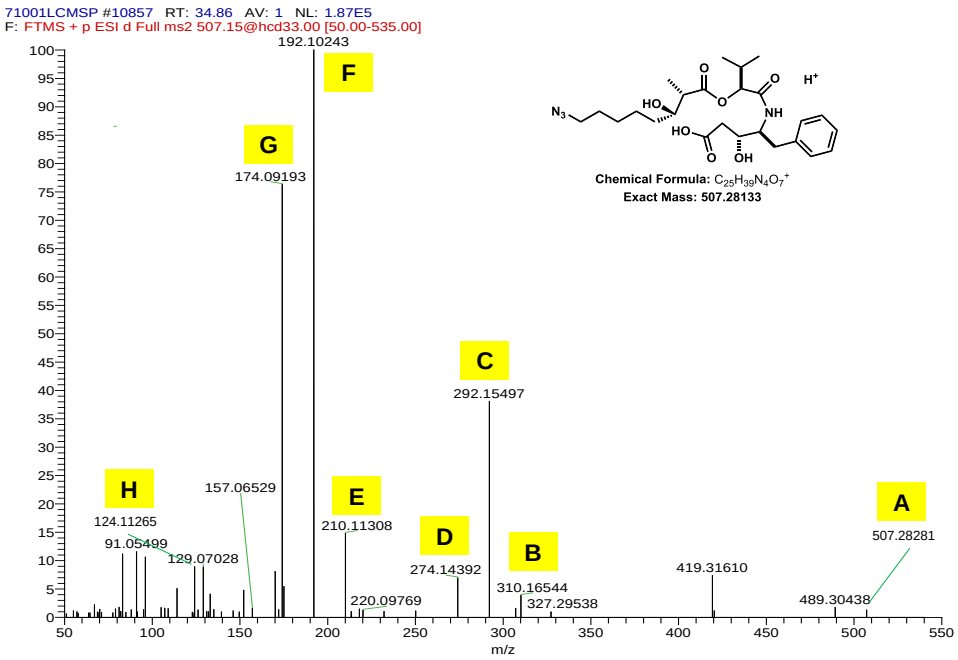
Figure D30 HR-MS-MS of compound **1hb**.

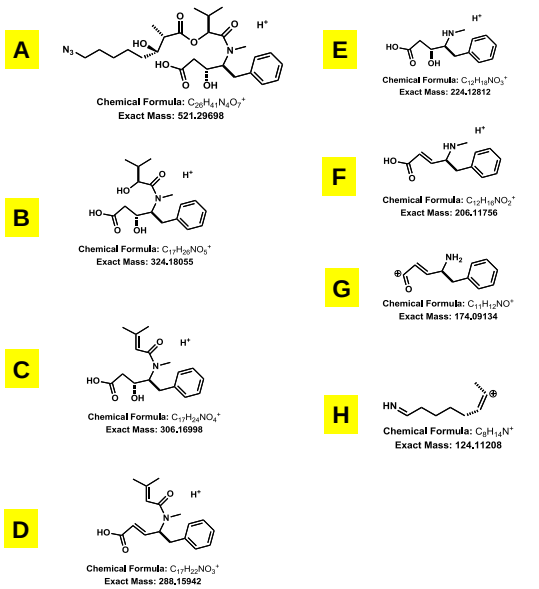




Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{26}H_{41}N_4O_7^+$	521.29698	521.29705	-0.1
B	$C_{17}H_{26}NO_5^+$	324.18055	324.18152	-3.0
C	$C_{17}H_{24}NO_4^+$	306.16998	306.17059	-2.0
D	$C_{17}H_{22}NO_3^+$	288.15942	288.16019	-2.7
E	$C_{12}H_{18}NO_3^+$	224.12812	224.12868	-2.5
F	$C_{12}H_{16}NO_2^+$	206.11756	206.11824	-3.3
G	$C_{11}H_{12}NO^+$	174.09134	174.09187	-3.0
H	$C_8H_{14}N^+$	124.11208	124.11253	-3.6

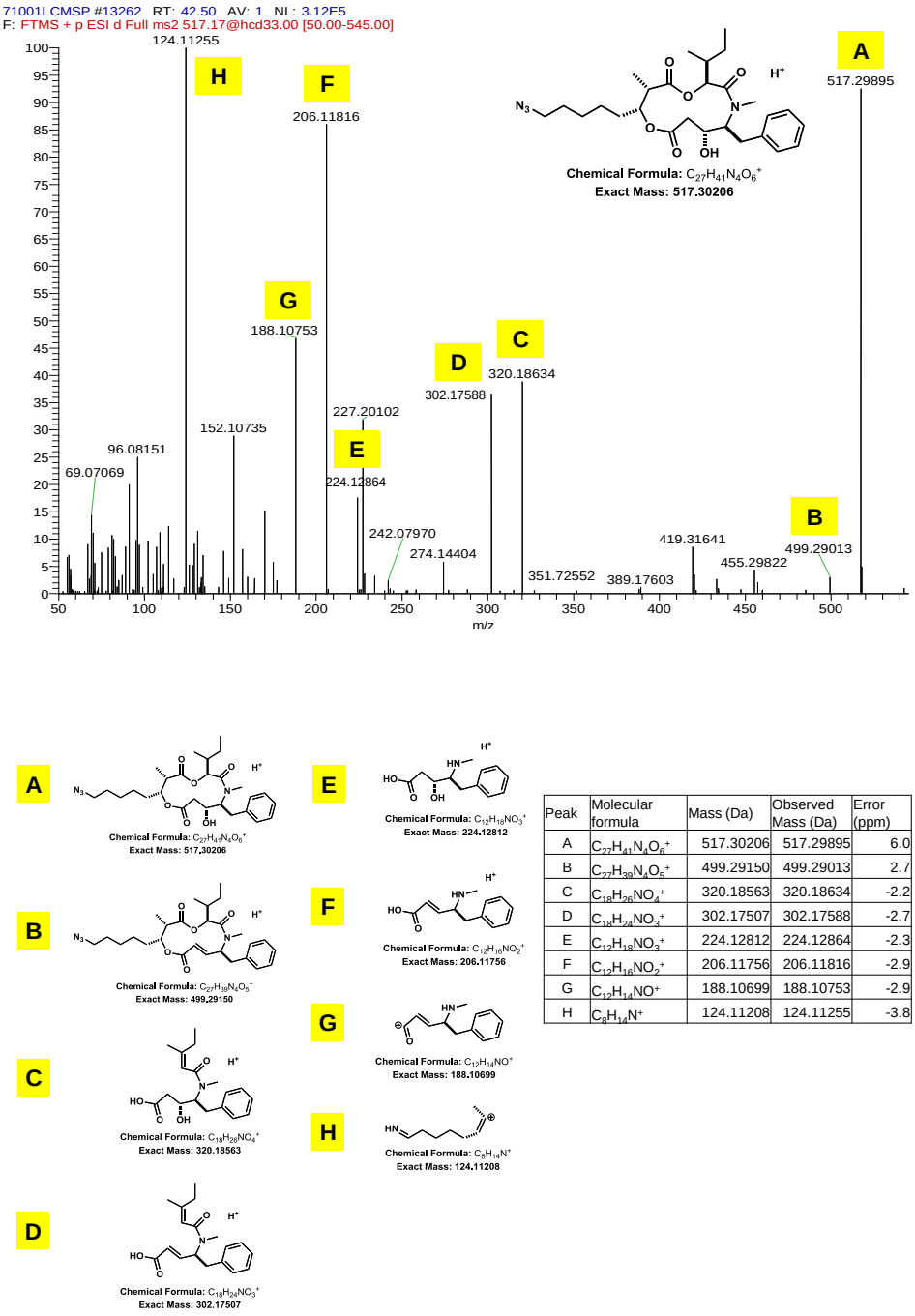
Figure D31 HR-MS-MS of compound **2h**.





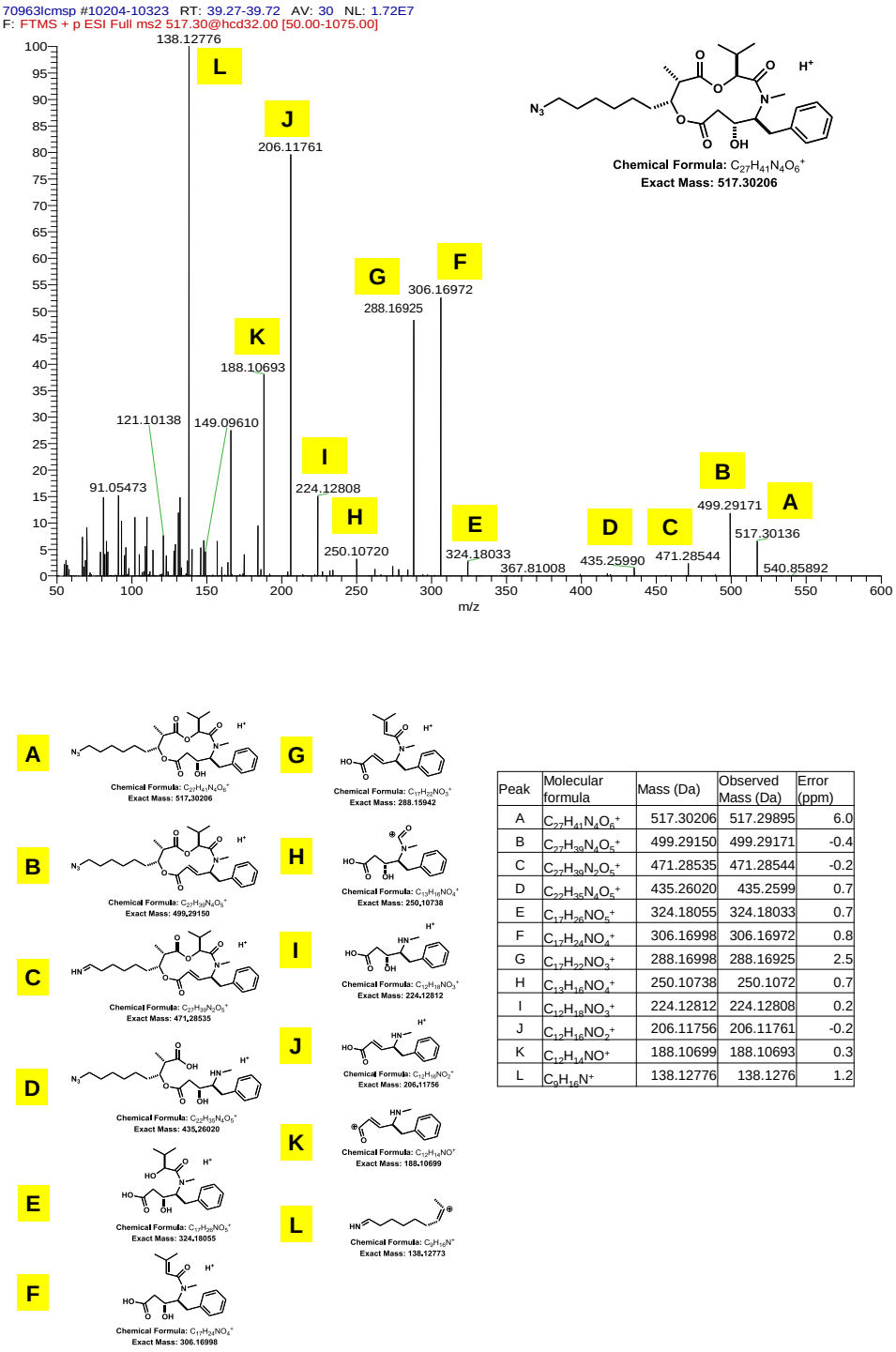
Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{26}H_{41}N_4O_7^+$	521.29698	521.29705	-0.1
B	$C_{17}H_{26}NO_5^+$	324.18055	324.18152	-3.0
C	$C_{17}H_{24}NO_4^+$	306.16998	306.17059	-2.0
D	$C_{17}H_{22}NO_3^+$	288.15942	288.16019	-2.7
E	$C_{12}H_{18}NO_3^+$	224.12812	224.12868	-2.5
F	$C_{12}H_{16}NO_2^+$	206.11756	206.11824	-3.3
G	$C_{11}H_{12}NO^+$	174.09134	174.09187	-3.0
H	$C_8H_{14}N^+$	124.11208	124.11253	-3.6

Figure D32 HR-MS-MS of compound **2hb**.



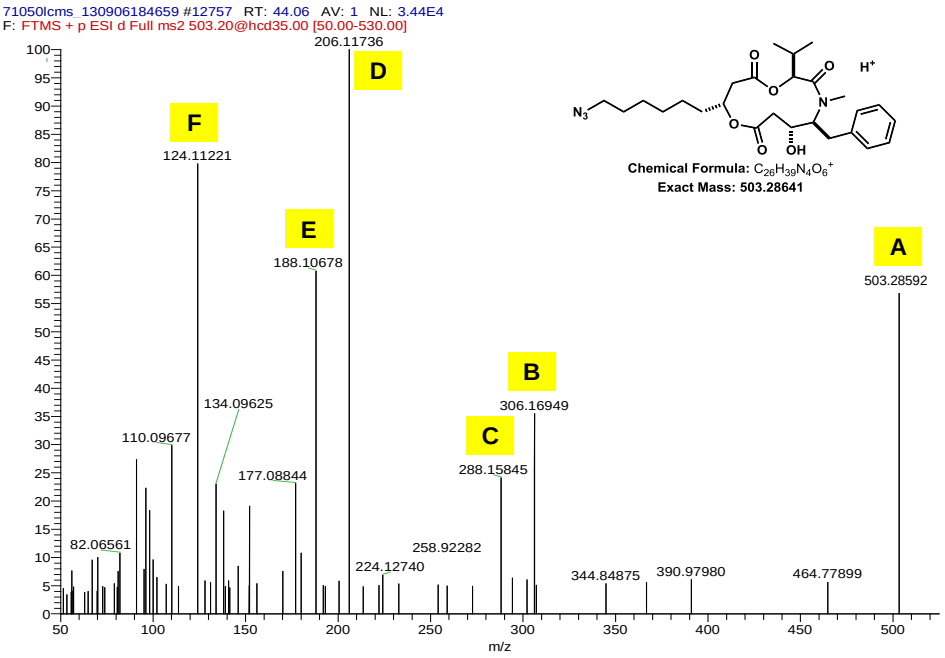
Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{27}H_{41}N_4O_6^+$	517.30206	517.29895	6.0
B	$C_{27}H_{39}N_4O_5^+$	499.29150	499.29013	2.7
C	$C_{18}H_{26}NO_4^+$	320.18563	320.18634	-2.2
D	$C_{18}H_{24}NO_3^+$	302.17507	302.17588	-2.7
E	$C_{12}H_{18}NO_3^+$	224.12812	224.12864	-2.3
F	$C_{12}H_{16}NO_2^+$	206.11756	206.11816	-2.9
G	$C_{12}H_{14}NO^+$	188.10699	188.10753	-2.9
H	$C_8H_{14}N^+$	124.11208	124.11255	-3.8

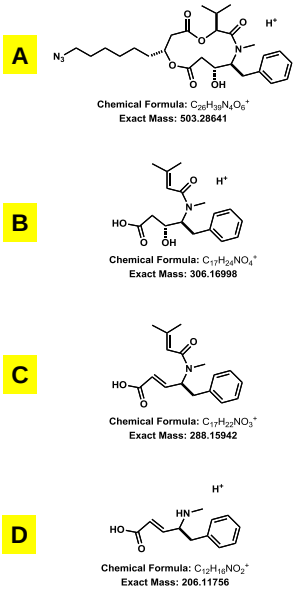
Figure D33 HR-MS-MS of compound **1i**.

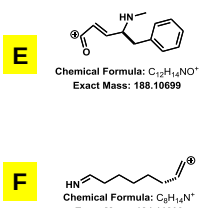


Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{27}H_{41}N_4O_6^+$	517.30206	517.29895	6.0
B	$C_{27}H_{39}N_4O_5^+$	499.29150	499.29171	-0.4
C	$C_{27}H_{39}N_2O_5^+$	471.28535	471.28544	-0.2
D	$C_{22}H_{35}N_4O_5^+$	435.26020	435.2599	0.7
E	$C_{13}H_{18}NO_3^+$	324.18055	324.18033	0.7
F	$C_{17}H_{24}NO_4^+$	306.16998	306.16972	0.8
G	$C_{17}H_{22}NO_3^+$	288.16998	288.16925	2.5
H	$C_{13}H_{16}NO_4^+$	250.10738	250.1072	0.7
I	$C_{12}H_{18}NO_3^+$	224.12812	224.12808	0.2
J	$C_{12}H_{16}NO_2^+$	206.11756	206.11761	-0.2
K	$C_{12}H_{14}NO^+$	188.10699	188.10693	0.3
L	$C_9H_{16}N^+$	138.12776	138.1276	1.2

Figure D34 HR-MS-MS of compound **1j.**

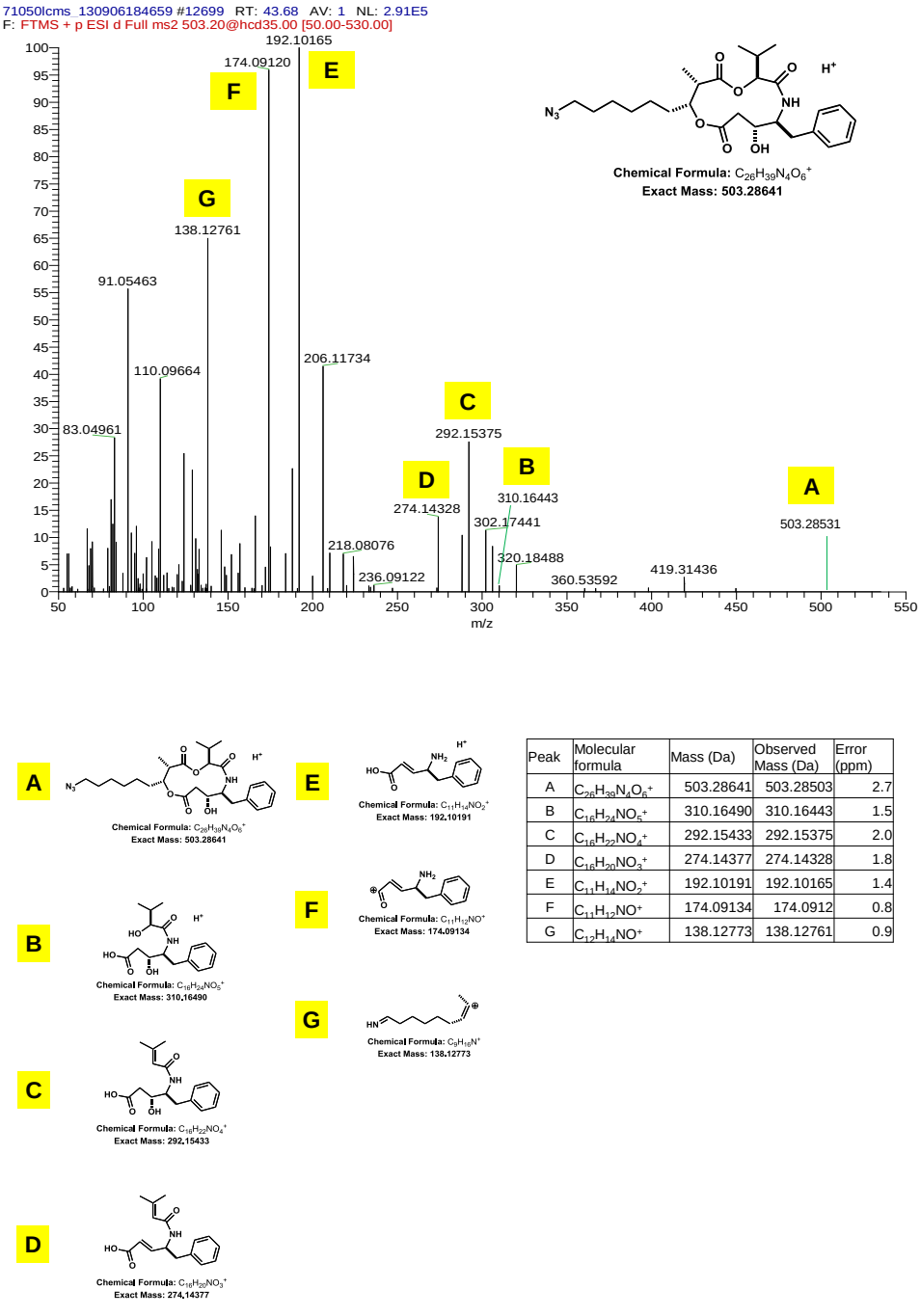






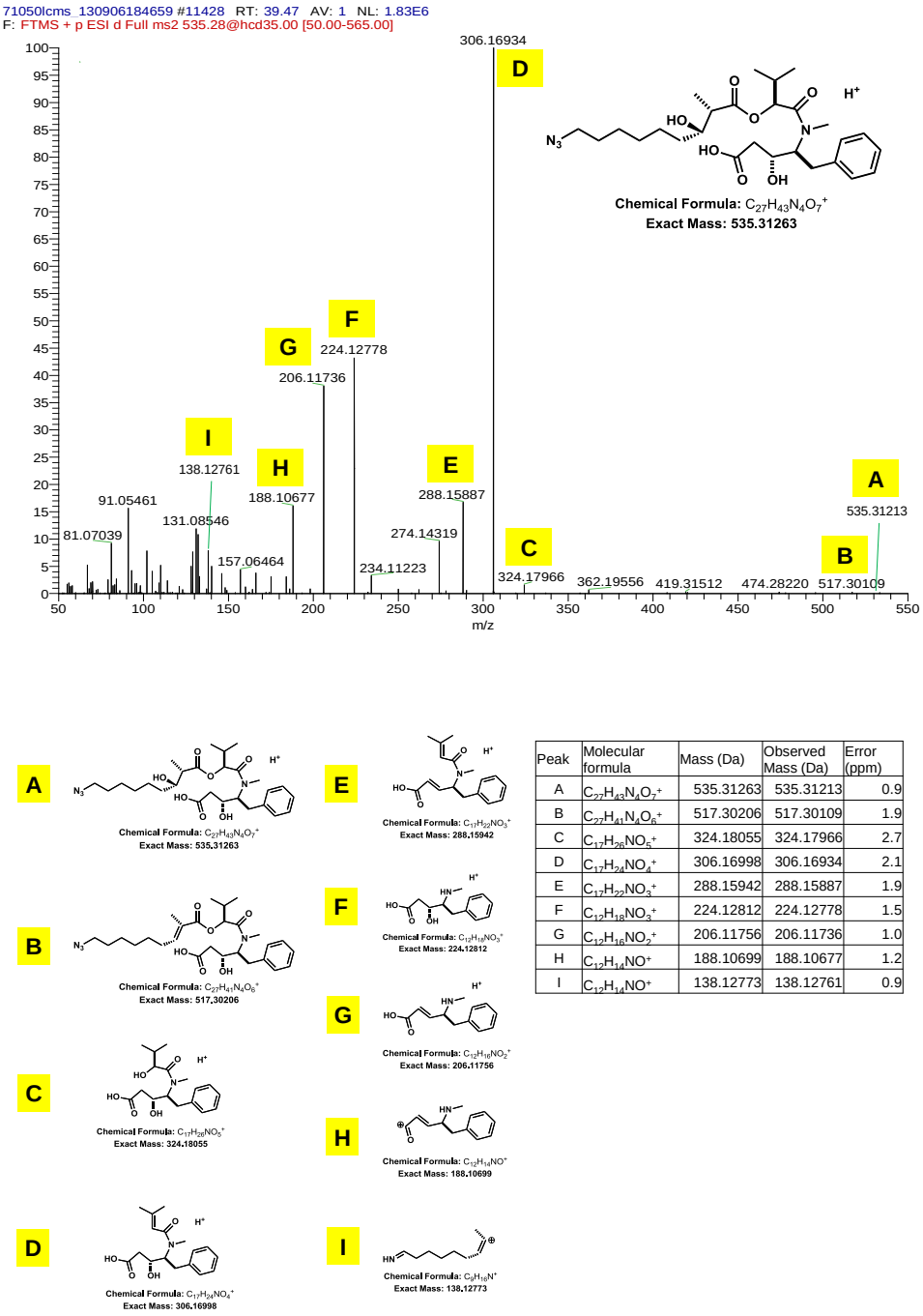
Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{26}H_{39}N_4O_6^+$	503.28641	503.28592	1.0
B	$C_{17}H_{24}NO_4^+$	306.16998	306.16949	1.6
C	$C_{17}H_{22}NO_3^+$	288.15942	288.15845	3.4
D	$C_{12}H_{16}NO_2^+$	206.11756	206.11736	1.0
E	$C_{12}H_{14}NO^+$	188.10699	188.10678	1.1
F	$C_8H_{14}N^+$	124.11208	124.11221	-1.0

Figure D35 HR-MS-MS of compound **1ja**.



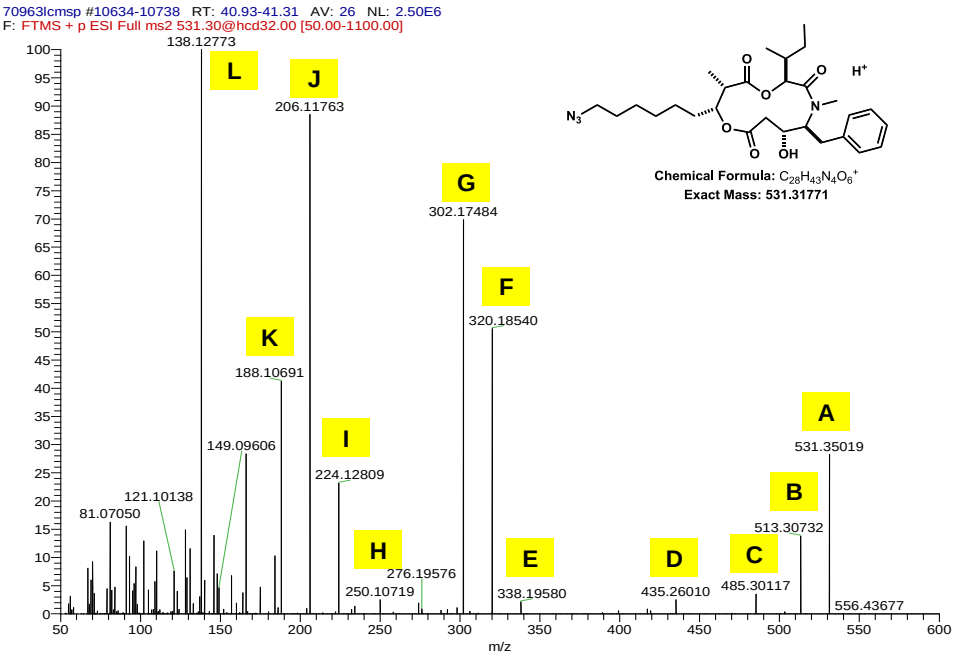
Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{26}H_{39}N_4O_6^+$	503.28641	503.28503	2.7
B	$C_{16}H_{24}NO_5^+$	310.16490	310.16443	1.5
C	$C_{16}H_{22}NO_4^+$	292.15433	292.15375	2.0
D	$C_{16}H_{20}NO_3^+$	274.14377	274.14328	1.8
E	$C_{11}H_{14}NO_2^+$	192.10191	192.10165	1.4
F	$C_{11}H_{12}NO^+$	174.09134	174.0912	0.8
G	$C_{12}H_{14}NO^+$	138.12773	138.12761	0.9

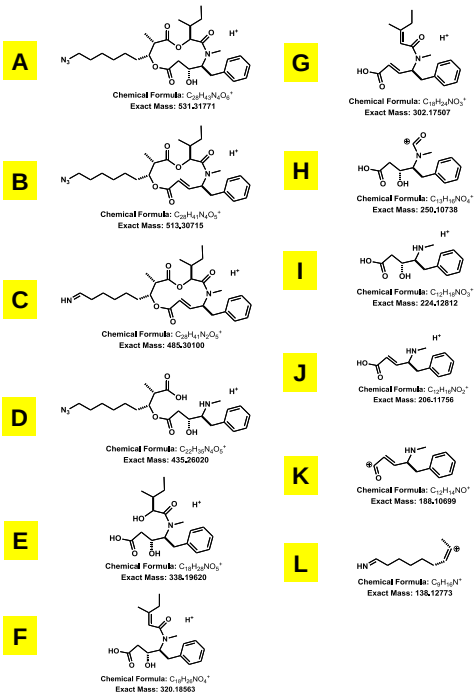
Figure D36 HR-MS-MS of compound **1jb**.



Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{27}H_{43}N_4O_7^+$	535.31263	535.31213	0.9
B	$C_{27}H_{41}N_4O_6^+$	517.30206	517.30109	1.9
C	$C_{17}H_{26}NO_5^+$	324.18055	324.17966	2.7
D	$C_{17}H_{24}NO_4^+$	306.16998	306.16934	2.1
E	$C_{17}H_{22}NO_3^+$	288.15942	288.15887	1.9
F	$C_{12}H_{18}NO_3^+$	224.12812	224.12778	1.5
G	$C_{12}H_{16}NO_2^+$	206.11756	206.11736	1.0
H	$C_{12}H_{14}NO^+$	188.10699	188.10677	1.2
I	$C_{12}H_{14}NO^+$	138.12773	138.12761	0.9

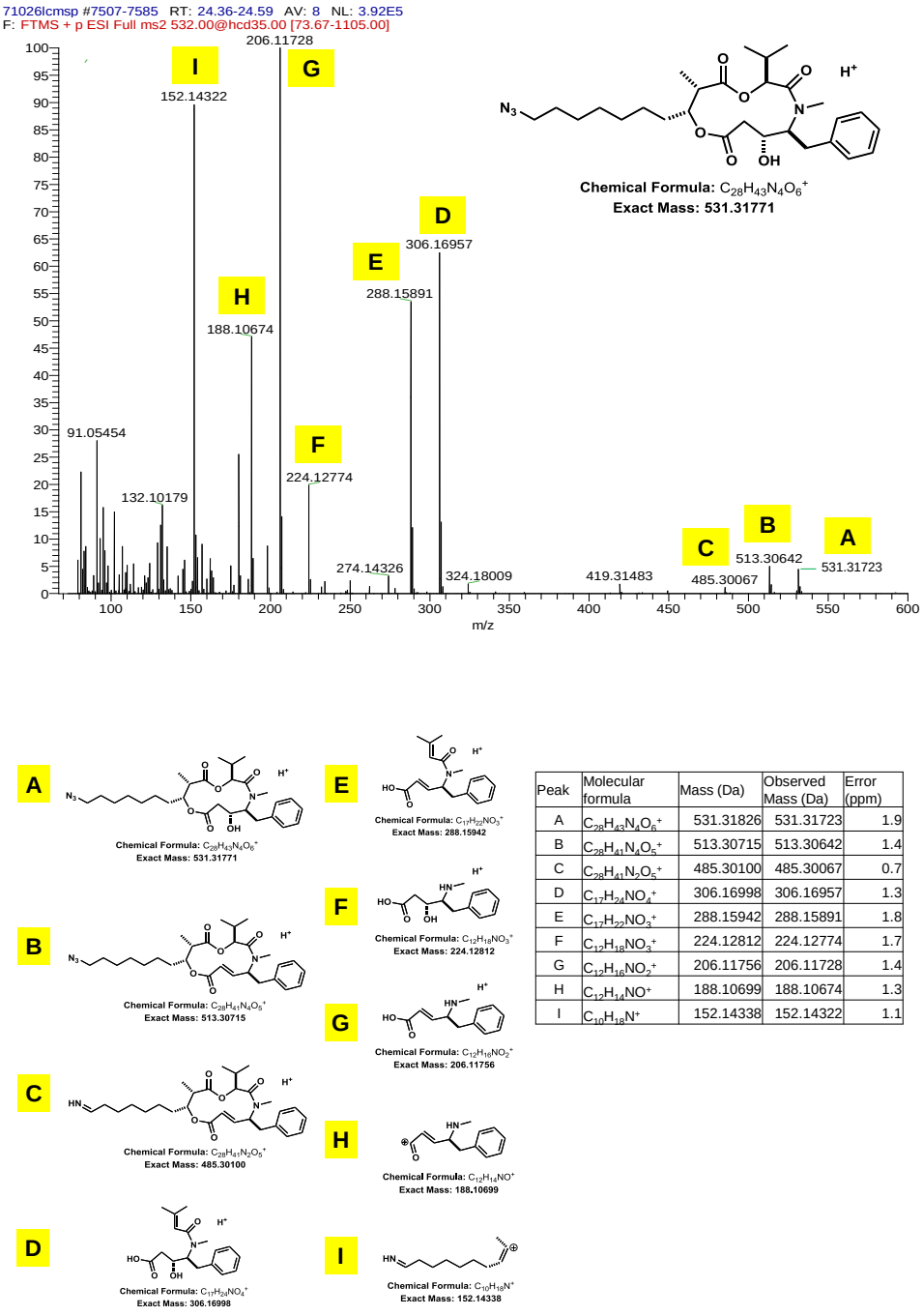
Figure D37 HR-MS-MS of compound **2j.**





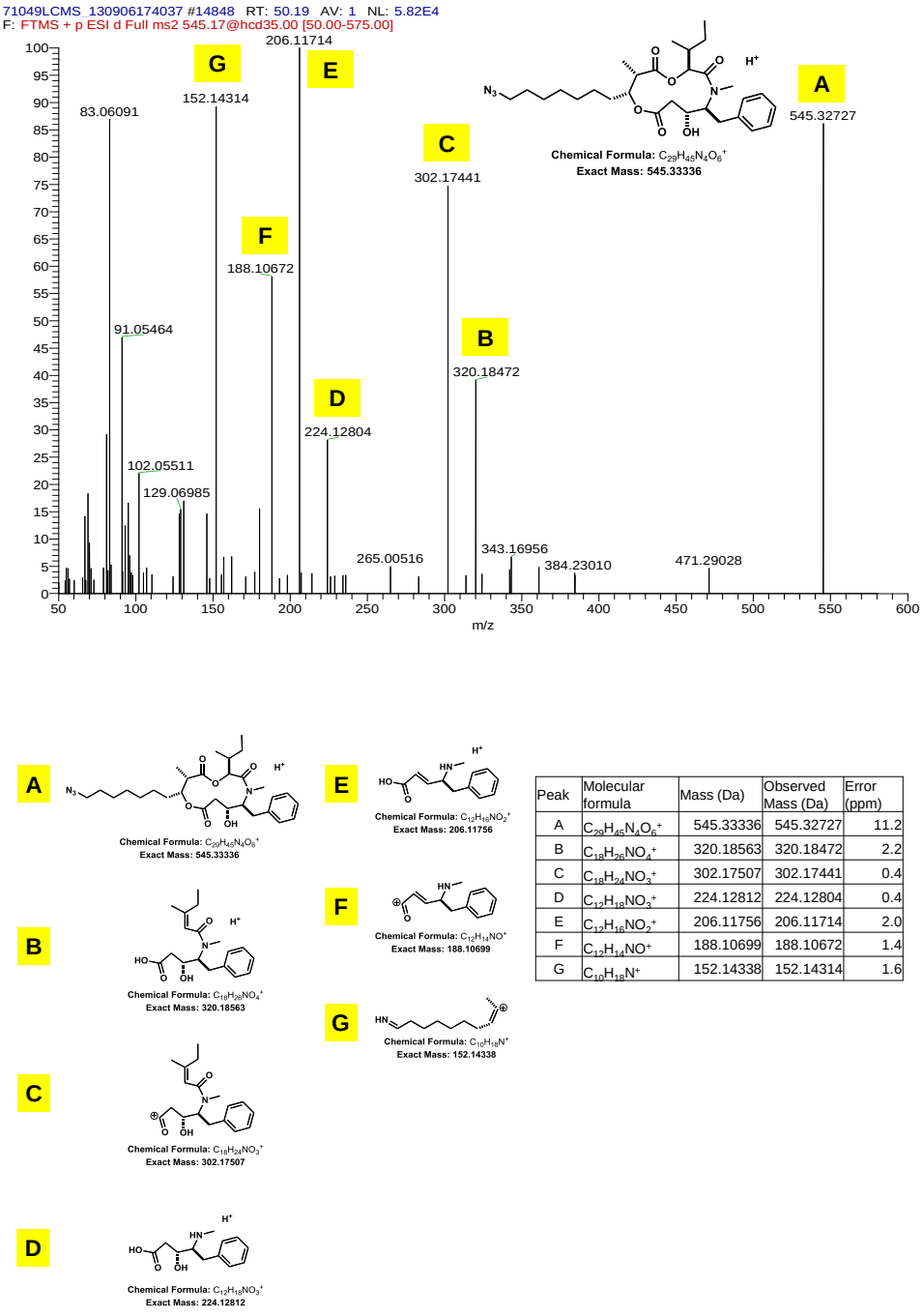
Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{28}H_{43}N_4O_6^+$	531.31826	531.31793	-0.6
B	$C_{28}H_{41}N_4O_5^+$	513.30715	513.30732	-0.3
C	$C_{28}H_{41}N_2O_5^+$	485.30100	485.30117	-0.4
D	$C_{22}H_{35}N_4O_5^+$	435.26020	435.2601	0.2
E	$C_{18}H_{28}NO_5^+$	338.19620	338.19580	0.7
F	$C_{18}H_{26}NO_4^+$	320.18563	320.18540	0.8
G	$C_{18}H_{24}NO_3^+$	302.17507	302.17484	0.8
H	$C_{13}H_{16}NO_4^+$	250.10738	250.10719	0.1
I	$C_{12}H_{18}NO_3^+$	224.12812	224.12809	-0.3
J	$C_{12}H_{16}NO_2^+$	206.11756	206.11763	0.4
K	$C_{12}H_{14}NO^+$	188.10699	188.10691	0.4
L	$C_9H_{16}N^+$	138.12776	138.12773	0.2

Figure D38 HR-MS-MS of compound **1k**.



Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{28}H_{43}N_4O_6^+$	531.31826	531.31723	1.9
B	$C_{28}H_{41}N_4O_5^+$	513.30715	513.30642	1.4
C	$C_{28}H_{41}N_2O_5^+$	485.30100	485.30067	0.7
D	$C_{17}H_{24}NO_4^+$	306.16998	306.16957	1.3
E	$C_{17}H_{22}NO_3^+$	288.15942	288.15891	1.8
F	$C_{12}H_{18}NO_3^+$	224.12812	224.12774	1.7
G	$C_{12}H_{16}NO_2^+$	206.11756	206.11728	1.4
H	$C_{12}H_{14}NO^+$	188.10699	188.10674	1.3
I	$C_{10}H_{18}N^+$	152.14338	152.14322	1.1

Figure D39 HR-MS-MS of compound **1l**.



Peak	Molecular formula	Mass (Da)	Observed Mass (Da)	Error (ppm)
A	$C_{29}H_{45}N_4O_6^+$	545.33336	545.32727	11.2
B	$C_{18}H_{26}NO_4^+$	320.18563	320.18472	2.2
C	$C_{18}H_{24}NO_3^+$	302.17507	302.17441	0.4
D	$C_{12}H_{18}NO_3^+$	224.12812	224.12804	0.4
E	$C_{12}H_{16}NO_2^+$	206.11756	206.11714	2.0
F	$C_{12}H_{14}NO^+$	188.10699	188.10672	1.4
G	$C_{10}H_{18}N^+$	152.14338	152.14314	1.6

Figure D40 HR-MS-MS of compound **1m**.

APPENDIX E

SUPPORTING INFORMATION FOR CHAPTER 7

E.1.1 General methods and materials

PCRs were executed using a C1000 thermal cycler (Bio-Rad). DNA sequencing was performed by Elim BioPharm Inc. Low resolution MS analysis was conducted using a Shimadzu LCMS-2020 instrument equipped with an Acclaim® C18 column (Dionex, 120Å, 3µm, 2.1 × 150mm). All MS experiments utilized electrospray ionization modes for analysis.

Synthetic oligonucleotides for gene amplification by PCR were purchased from Life Technologies or Integrated DNA Technology. Kapa HiFi DNA polymerase was obtained from Kapa Biosystems. Restriction endonucleases, T4 DNA ligase and Antarctic phosphatase were purchased from New England BioLabs. LB broth and agar used for culturing E. coli were obtained from Teknova.

E.1.2 Protein cloning, expression and purification

HalC adenylation domains A1 and A2 were independently cloned into the NotI and BamHI sites of the expression plasmid pQTev to yield pQTev-HalCA1 and pQTev-HalCA2, respectively. An overnight culture of the expression strain *E. coli* BL21(DE) star containing either plasmid was inoculated by 1000-fold dilution into fresh media (1.0L scale, LB media, 37°C, 200rpm). The culture was then grown to an OD600nm of approximately 0.6, cooled (16°C), and induced with isopropyl β-D-1-thiogalactopyranoside (IPTG, 1mM final concentration, overnight). The following day, protein expression cultures were centrifuged (5000g, 30min, 4°C), and the resulting cell pellet was resuspended in lysis buffer (10% (v/v) glycerol, 0.1% (v/v) Tween-20, 500mM NaCl, 50mM Tris-HCl, pH 7.5). The resuspended cells were placed on ice and then sonicated (Fisher Scientific Sonic Dismembrator 550) in 5 second pulses, followed by 20 second cool down periods, for a time totaling 5 minutes. The disrupted cell pellet was then centrifuged (20000g, 30min, 4°C), and the soluble lysate was applied to NTA nickel beads (Ni-NTA Agarose, MCLAB) and allowed to mix gently (1hr, 4°C). The protein-bound NTA nickel beads were then washed (20mL of 20mM imidazole, 10% (v/v) glycerol, 500mM NaCl, 50mM Tris-HCl, 10mM β-mercaptoethanol, pH 7.5). Following the washing step, protein was eluted from the NTA nickel beads with elution buffer (wash buffer containing 125mM imidazole). Eluted fractions were analyzed via SDS-PAGE on a 12.5% (w/v) polyacrylamide gel, using a constant voltage of 100V (Bio-Rad Mini-PROTEAN® Tetra Cell and PowerPac Basic). Protein-containing fractions were combined together and dialyzed overnight (2×500mL, 100mM Tris-HCl, 1mM $MgCl_2$, 10% (v/v) glycerol, pH 7.5, 4°C). The dialyzed protein solution was then concentrated via centrifugal filtration (Spin-X® UF Concentrator 30k MWCO PES, Corning®, 4696g, 5min, 4°C).

E.1.3 Enzyme kinetics and substrate activity assays

A pyrophosphate detection assay reported previously [249] was modified to investigate adenylation domain enzyme kinetics and substrate selectivity. Briefly, pyrophosphate reactions (100µL) contained 100mM HEPES pH 8.0, 10mM $MgCl_2$, 0.04 U pyrophosphatase, 0.1 U purine nucleoside phosphorylase, 250µM 7-methylthioguanosine, 1mM tris(2-carboxyethyl)phosphine, 150mM hydroxylamine, 2.5mM ATP, and varying concentrations of substrate. Adenylation enzyme (5µM final concentration) was added to initiate the reaction, which was monitored at 360nm at room temperature over 5 minutes. GraphPad Prism® 5 was used for curve fitting to extrapolate Michaelis-Menten enzyme kinetics parameters. To determine substrate activity, the absorbance at 360nm of the pyrophosphate reactions was measured after 5 minutes of incubation at room temperature. For these assays, 250µM of substrate or disodium pyrophosphate (Na_2PPi) was tested. Substrate activities of various substrates were normalized to Na_2PPi (positive control) or sterile dH_2O (negative control) responses.

E.1.4 Hapalosin Gene Cluster Plasmid construction

For details for the assembly of the complete hapalosin biosynthetic gene cluster, refer to **Chapters 4 and 5**. A brief summary of the procedures is outlined here. To reconstruct the entire hapalosin biosynthetic pathway, genes *halA* and *sfp* were cloned by PCR (using genomic DNA isolated from *Hapalosiphon welwitschii* UTEX1830B) and yeast recombineered into parent vector pMQ131 to yield daughter plasmid pXL2. Primers used to amplify the *sfp* and *halA* genes are as follows, respectively: sfpRBSFW/sfpRBSRV and HapAFW/HapARV. These genes were located downstream of the isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible Lac

promoter. Included within the design of the PCR primers were sequences to synthetically insert an *E. coli* ribosome binding site (RBS) immediately upstream of gene *halA*. This genetically engineered feature, therefore, enabled the enhanced translation of *halA* mRNA transcripts to yield the functional enzyme *in vivo*.

A sequential cloning procedure was implemented to yeast recombineer genes *halBCDE* into parent vector pMQ123i to yield the final daughter plasmid pXL4. Sequential construction of the final expression plasmid enabled the genetic engineering of an *E. coli* RBS sequence immediately upstream of genes *halDE* as well as a site-directed mutation of the native TTG start codon of *halD* to the high frequency start codon ATG of the *E. coli*, ATG. This was accomplished by first recombineering genes *halDE* into pMQ123i, yielding the partial expression plasmid pXL3. Similar to the construction of plasmid pXL2, PCR primers were designed to introduce the RBS sequences upstream of the *halDE* genes. The *halDE* genes were cloned as three separate fragments using the following primers: HalDE_Frag1_FW / HalDE_Frag1_RV, HalDE_Frag2_FW / HalDE_Frag2_RV, and HalDE_Frag3_FW / HalDE_Frag3_RV. Following successful construction of the *halDE* initial plasmid, known as pXL3, *halBC* were introduced to afford the *halBCDE* expression plasmid, pXL4. The *halBC* genes were cloned as four separate fragments and introduced into the Eco53kI site of pXL3; primers used for *halBC* cloning are as follows: HalBC_Frag1_FW / HalBC_Frag1_RV, HalBC_Frag2_FW / HalBC_Frag2_RV, HalBC_Frag3_FW / HalBC_Frag3_RV, and HalBC_ Frag4_FW / HalBC_Frag4_RV. These genes were situated downstream of the IPTG-inducible Tac promoter. A summary of the plasmids used in this study are listed in **Table E1**.

To introduce the mutant *halC_KR* sequence into the *hal* pathway, two amplicons were generated using primers HalC-KRmut1_FW/HalC-KRmut1_RV and HalC-KRmut2_FW/HalC-KRmut2_RV (**Table E2**; point mutations underlined). Primers were designed to introduce S2071A and Y2086F point mutations within the active site of HalC-KR. Once this mutant locus was generated and sequenced, it was used as a template to PCR amplify and recombine into the complete Hal pathway containing the mutant *halC_KR* sequence using the yeast recombineering procedures described above.

Construction of the stand-alone adenylation domains HalC_A1 and HalC_A2 were constructed via ligation-based cloning methods into the NotI and BamHI restriction enzyme sites of the pQTev expression plasmid. Primers HalCA1_BamHI_5'/HalCA1_BamHI_3' and HalCA2_BamHI_5'/HalCA2_BamHI_3' were used for cloning of the HalC_A1 and HalC_A2 adenylation domains (respectively) into pQTev.

E.1.5 Heterologous expression of the hapalosin biosynthetic pathway

Electrocompetent *E. coli* XL1-Blue cells were transformed with pHal2 and pHal3 or pHal4 constructs and allowed to recover in SOC media for 1hr at 37°C before plating on LB agar plates containing kanamycin (50 µg/mL) and gentamicin (10 µg/mL). A single colony of the transformant was used to inoculate 1mL of LB containing the appropriate antibiotics and allowed to grow at 37°C overnight. 500µL of the seed culture was subsequently used to inoculate fresh LB media (50mL) containing the appropriate antibiotics and allowed to grow at 37°C with shaking (200 rpm). When the culture reached mid-log phase exponential growth (OD600 = 0.6), the flask was cooled to 25°C and IPTG (1mM) was introduced. After 48hr of induction at 25°C,

the culture was centrifuged (5000 g, 30min) and the supernatant was extracted with ethyl acetate (2×50mL). The ethyl acetate layers were combined and dried over Na_2SO_4 and the organic solvent was evaporated in vacuo under gentle heating (37°C). When necessary, 2-hydroxy-3-methylbutyric acid (1mM) was added to the *E. coli* culture at the time of induction.

E.1.6 LC-MS analysis of hapalosin expression in *E. coli*

The ethyl acetate-supernatant crude extract was dissolved in methanol to a concentration of approximately 10 mg/mL and centrifuged (21.1×10^3 g, 5min). The organic supernatant was then used for subsequent LR-liquid chromatography-mass spectrometric (LC-MS) analysis. 2µL of MeOH-dissolved supernatant extract was separated by reverse phase high performance liquid chromatography, R-HPLC prior to LR-LCMS analysis, (Dionex Acclaim® C18, 120Å, 3µm, 2.1 × 150mm) and analyzed on a Shimadzu LCMS-2020 instrument using both ESI positive and negative and APCI ionizations. LC method used for analysis (0.2 mL/min flow rate): 5% (v/v) MeCN in H_2O (including 0.1% (v/v) formic acid) for 5 min, ramp to 95% (v/v) MeCN in H_2O (including 0.1% (v/v) formic acid) over 45 min, and then a final wash with 95% MeCN in H_2O (including 0.1% (v/v) formic acid) for 5 min.

E.1.7 Cultivation of *Hapalosiphon welwitschii* UTEX 1830 and metabolite analysis

UTEX B1830 was cultured as described previously. Briefly, 1L volume cultures were grown in BG-11 media adjusted to pH = 7.8. Cyanobacteria were separated via filtration and the cells were lyophilized to yield a dry, filamentous green mass. To 5.0g of lyophilized cells, 400mL of 1:1 MeOH/DCM was added and stirred overnight at room temperature. The organic solvent was

separated from the cell mass via filtration and evaporated *in vacuo* under gentle heating (37°C) to yield the dried crude extract. The crude extract was redissolved in MeOH and analyzed via LC-MS, using the protocol described above.

HalC-A1 domain homologs	GenBank Accession #	Full Length Size (aa)	Organism	Proposed function
gi 111019388	YP_702360	4903	Rhodococcus jostii RHA1	non-ribosomal peptide synthetase
gi 18034623	AAL57600	3316	Xenorhabdus bovienii	peptide synthetase XpsB
gi 83768440	BAE58579	1021	Aspergillus oryzae RIB40	unnamed protein product
gi 163939760	YP_001644644	1476	Bacillus weihenstephanensis KBAB4	amino acid adenylation domain-containing protein
gi 87121795	ZP_01077681 NZ_AANE01000000	1558	Marinomonas sp. MED121	Amino acid adenylation, partial
gi 26248307	NP_754347	1120	Escherichia coli CFT073	peptide synthetase
gi 160944390	ZP_02091618 NZ_ABED02000000	491	Faecalibacterium prausnitzii M21/2	hypothetical protein FAEPRAM212_01900
gi 68845342	P0C063	4451	Aneurinibacillus migulanus	Gramicidin S synthase II

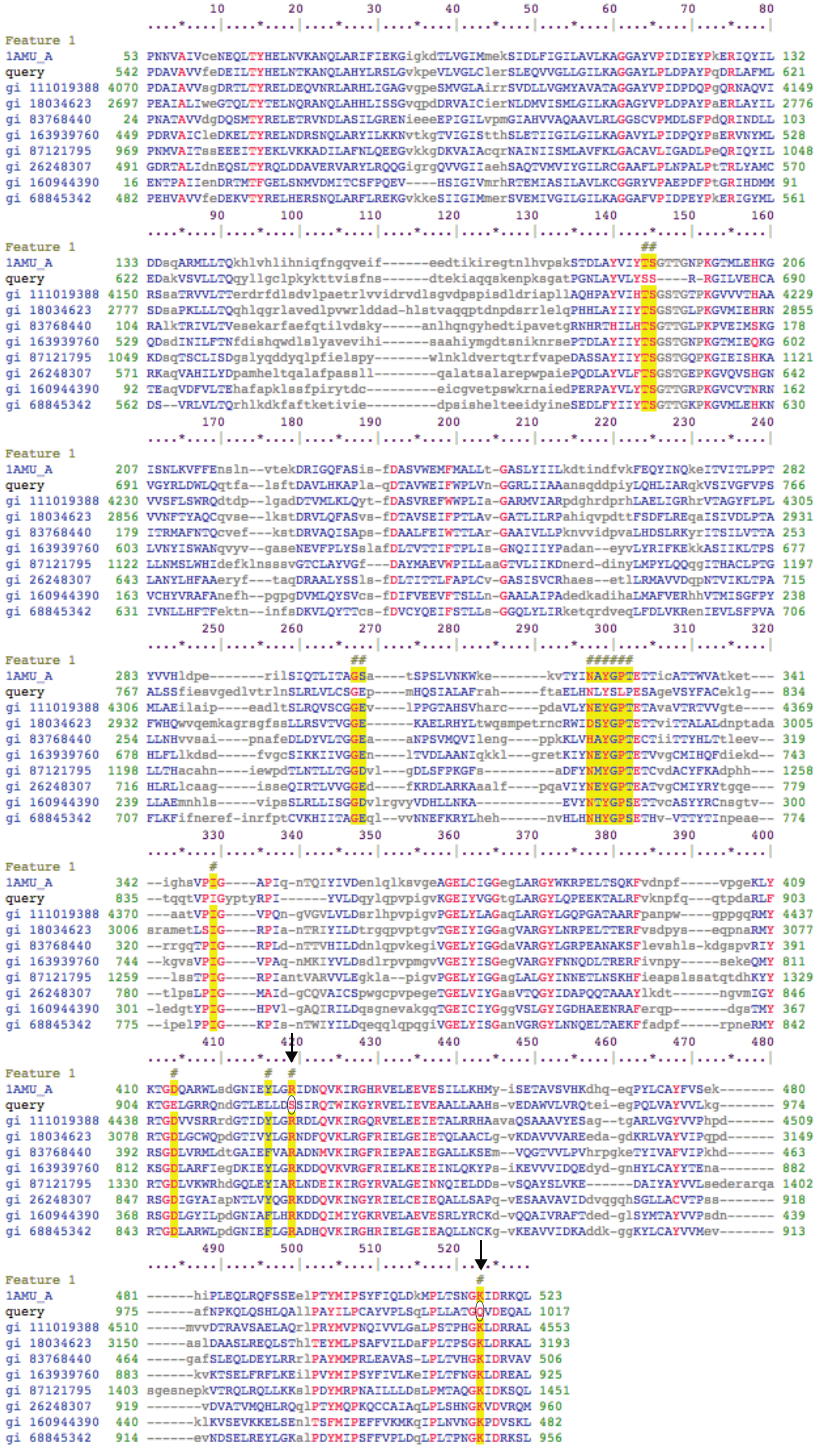


Figure E1 BLAST analysis and sequence alignments for Hal-A2 domain.

HalC-A2 domain homologs	GenBank Accession #	Full Length Size (aa)	Organism	Proposed function
gi 254763406	P40806	5043	Bacillus subtilis subsp. subtilis str. 168	Polyketide synthase PksJ
gi 17549638	NP_522978	937	Ralstonia solanacearum GMI1000	peptide synthetase
gi 116584634	ABK00751	3374	Bacillus cereus	cereulide synthetase A, CesA
gi 157691118	YP_001485580	2731	Bacillus pumilus SAFR-032	nonribosomal peptide synthetase subunit
gi 126347649	CAJ89363	2873	Streptomyces ambofaciens ATCC 23877	putative peptide synthetase
gi 133917244	CAM59601	1416	Planktothrix agardhii NIVA-CYA 126	non-ribosomal peptide synthetase/polyketide synthase hybrid enzyme
gi 156141035	ABU51602	970	Aspergillus nidulans	nonribosomal peptide synthetase TdiA
gi 157429065	ABV56587	1273	Kutzneria sp. 744	nonribosomal peptide synthetases KtzG
gi 159898935	YP_001545182	1818	Herpetosiphon aurantiacus DSM 785	AMP-dependent synthetase/ligase

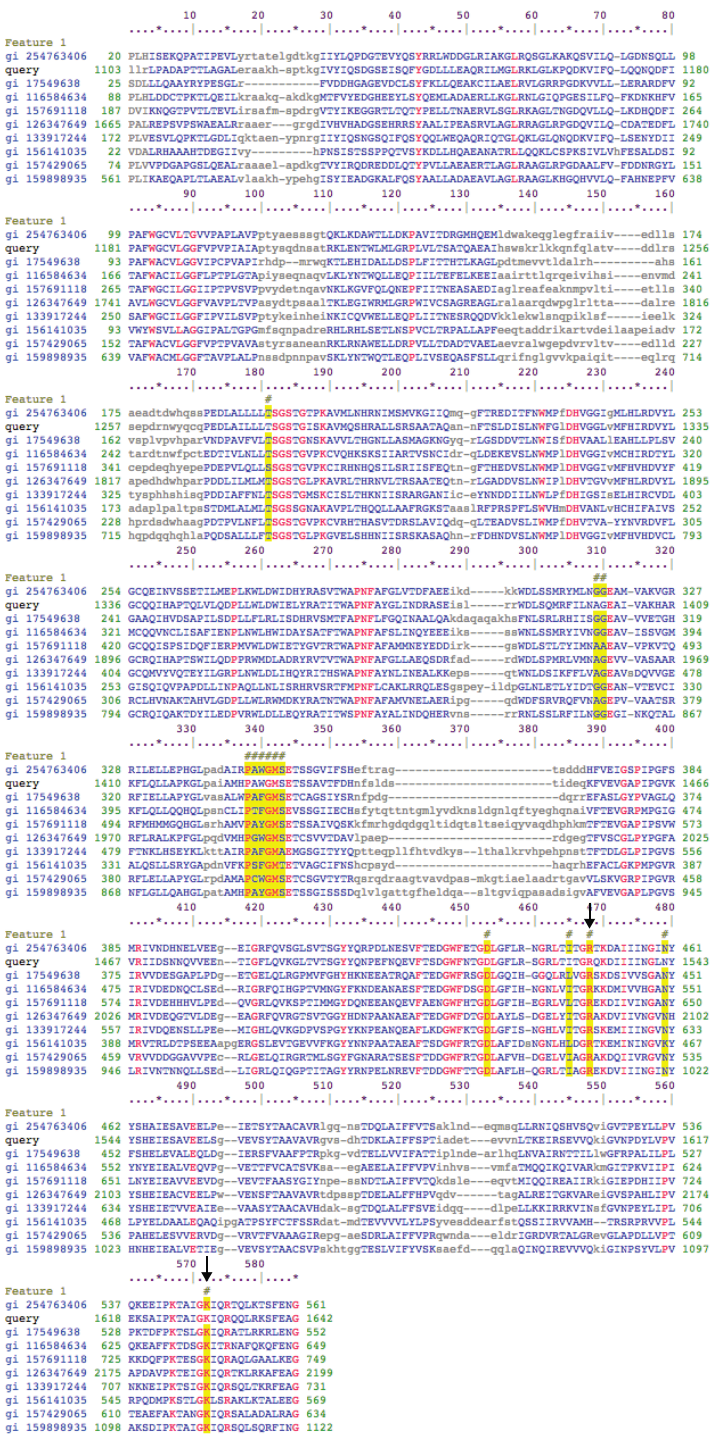


Figure E2 BLAST analysis and sequence alignments for Hal-A1 domain.

Table E1 Plasmids used in this study

Plasmid name	Features	References
pMQ131	$_{ori}$ColE1/pC194, *aphA-3*, P_{lac}-*lacZα, CEN6*	Reference [74]
pHal2	Derived from pMQ131; contains *halA* and *sfp*	This work
pMQ123i	$_{ori}$ColE1/pRK2, *aacC1*, P_{tac}-*lacZα, 2μm*	Reference [74]
pHal3	Derived from pMQ123i; contains *halDE*	This work
pHal4	Derived from pHal3; contains *halBC*	This work
pHal4mutant	Derived from pHal4; contains mutant halC_KR	This work
pQTev-HalCA1	Sequences for halC_A1 domain cloned into NotI & BamHI sites of pQTev	This work
pQTev-HalCA2	Sequences for halC_A2 domain cloned into NotI & BamHI sites of pQTev	This work

Table E2 Primers used in this study

Primer name	Sequence	Application
HalC-KRmut1_FW	gaattggatcctctagattctccatacggtcaaagagattgaagaaacaaaac	Construction of *HalC* mutant KR template
HalC-KRmut1_RV	AATGCACTAAAGGTAGCTCCTCCAAATACACTATTTACTGCAGAGAAAG	Construction of *HalC* mutant KR template
HalC-KRmut2_FW	GcagtaaatagtgtatttggaggagctacctttagtgcatTttctgctg	Construction of *HalC* mutant KR template
HalC-KRmut2_RV	catatgtatatctccttccccgggtaccgagtcaagtattttcggcttttagc	Construction of *HalC* mutant KR template
HalB-KSmut-seqFW	GTTGATATATGCAACCTTGATG	Sequencing of *HalC* mutant KR template
HalB-KSmut-seqRV	TAGAATAAGACTTGCTGTCCG	Sequencing of *HalC* mutant KR template
sfpRBSFW	TGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGAAGATTTACGGAATTTATATG	5’ primer to clone *sfp* and RBS
sfpRBSRV	TTATATCCTCCTACGGGTATGGAGAATTATAAAAGCTCTTCGTACGAG	3’ primer to clone *sfp* and RBS
HapAFW	TTCTCCATACCCGTAGGAGGATATAAATGAACAACAATTTATTTTCAG	5’ primer to clone *halA*
HapARV	CTGTTTTATCAGACCGCTTCTGCGTTCTGATGGGCCCTTAAACACTAACTGAAGAGG	3’ primer to clone *halA*

HalDE_Frag1_FW	CGAATTCGAGCTCGGTACCCGGGGAAGGAGA TATACATATGAGCAATATCTCTAAAAAAACC	5’ primer to clone fragment 1 of *halDE*
HalDE_Frag1_RV	TTCTTCTTCGGCGGCTACACTAGAG	3’ primer to clone fragment 1 of *halDE*
HalDE_Frag2_FW	GCGTTGAGTATTTGTTGCAGG	5’ primer to clone fragment 2 of *halDE*
HalDE_Frag2_RV	CCTTGTTCTAGTAGCTTCGAG	3’ primer to clone fragment 2 of *halDE*
HalDE_Frag3_FW	GTCAATTGATGCAGCAATTGC	5’ primer to clone fragment 3 of *halDE*
HalDE_Frag3_RV	TATCAGACCGCTTCTGCGTTCTGATTTAATCT GTATCATCAATCATCTGCCTGTGACCG	3’ primer to clone fragment 3 of *halDE*
HalBC_Frag1_FW	GAATTGGATCCTCTAGATTCTCCATACAGGAG GAATAATATGAATAGAGAACCAATCGCC	5’ primer to clone fragment 1 of *halBC*
HalBC_Frag1_RV	CTTGTGCTAAATTTGACCCAC	3’ primer to clone fragment 1 of *halBC*
HalBC_Frag2_FW	CACTCGCGGCTAATGCATCGTC	5’ primer to clone fragment 2 of *halBC*
HalBC_Frag2_RV	GTTTGGGGTGGGAGTGCTACC	3’ primer to clone fragment 2 of *halBC*
HalBC_Frag3_FW	gagcaacgaacccgtaatacgc	5’ primer to clone fragment 3 of *halBC*
HalBC_Frag3_RV	gtaaaattattggcttgagcggtagcag	3’ primer to clone fragment 3 of *halBC*
HalBC_Frag4_FW	ggtcaaagagattgaagaaacaaaac	5’ primer to clone fragment 4 of *halBC* containing mutant halC KR
HalBC_Frag4_RV	catatgtatatctccttccccgggtaccgagtcaagtattttcggcttttagc	3’ primer to clone fragment 4 of *halBC* containing mutant halC KR
HalCA1_BamHI_5’	GATCGGATCCcccgacgcggtagcagtggtg	Clone HalC_A1 into pQTEV
HalCA1_BamHI_3’	GATCGCGGCCGCctccttttaaaactacgtaag	Clone HalC_A1 into pQTEV
HalCA2_BamHI_5’	GATCGGATCCctacttcgacttcctgctg	Clone HalC_A2 into pQTEV
HalCA2_BamHI_3’	GATCGCGGCCGCcagcttcaaacgactttctc	Clone HalC_A2 into pQTEV

APPENDIX F

SUPPORTING INFORMATION FOR CHAPTER 8

F.1.1 General methods and materials

Refer to **Chapter 5, Plasmid Tools Expression** for detailed information regarding general methods and materials utilized for the following experiments. Heterologous gene expression and MS analysis procedures are also detailed there.

Table F1 Primers used in this study

Primer	Sequence	Purpose
mca_frag1_5'_toMQ123i	cgagctcggtacccggggaaggagatatacatATGCTAGATAGATTACATACCATC	Assembly of *mca* WT operon
mca_frag1_3'	TCTCTAGTGCCTGTTTATTAATC	Assembly of *mca* WT operon
mca_frag2_5'	GCAGCTTTACGGCAGCCTG	Assembly of *mca* WT operon
mca_frag2_3'	CATATTGTTCTCCCGGAGCAAG	Assembly of *mca* WT operon
mca_frag3_5'	GTATGTCTGCATCTTGGCTCAG	Assembly of *mca* WT operon
mca_frag3_3'	AACCAATCTTCTGCGGAAAG	Assembly of *mca* WT operon

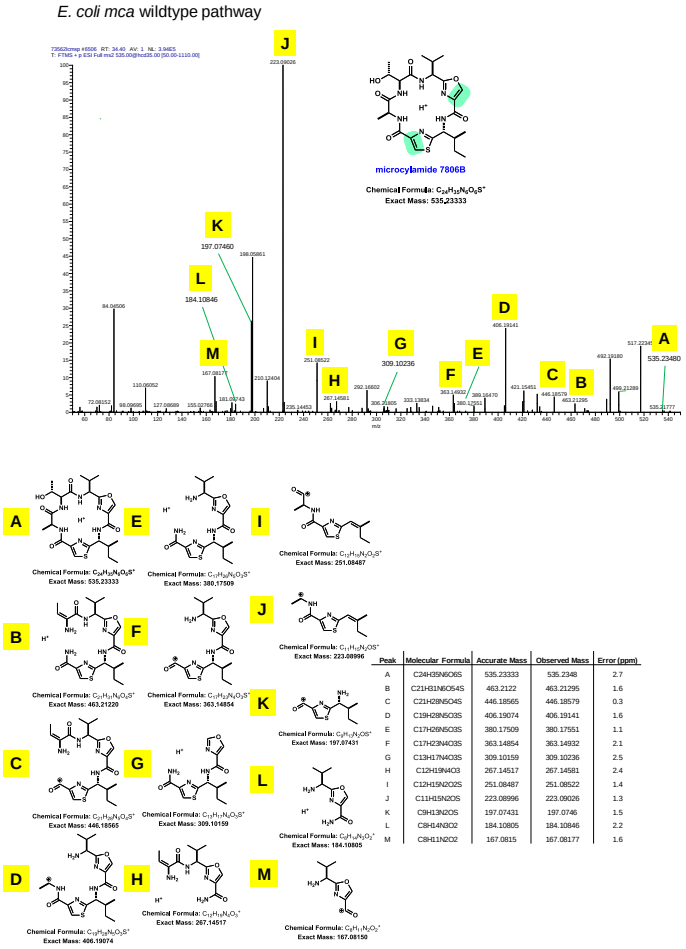
mca_frag4_5'	CTCCCATGAGTGACCAGATTGTG	Assembly of *mca* WT operon
mca_frag4_3'_toMQ123i	TTTATCAGACCGCTTCTGCGTTCTGATTTATCAGG TTTGTTCAGAACGCGAAG	Assembly of *mca* WT operon
mcaA-123i-3'	ttatatcctcctacgggtatggagaaTTAGTAGGGAGAAGACCA AGAAC	Assembly of re-organized *mca* pathway
mcaD-123i-5'	ttctccatacccgtaggaggatataaATGCAATCCACCCCACTGC TC	Assembly of re-organized *mca* pathway
mcaD-123i-3'	agttaatttctcctctttaatgaattTTAAAAGGGGATGTTGGTCT GGTTC	Assembly of re-organized *mca* pathway
mcaG-123i-5'	aattcattaaagaggagaaattaactATGAACAATTGCCCTTTTG CTATTTC	Assembly of re-organized *mca* pathway
McaE-5'-insert-to-123iMcaADG	gcggataacaatttcacacaggaaacagaattatgaggattactccaatggataag	Assembly of re-organized *mca* pathway
McaE-3'-insert-to-123iMcaADG	gtaatctatctagcatatgtatatctccttctttaaTTAGGCTTCGTCACC ATCG	Assembly of re-organized *mca* pathway
mcaG_into_mcaF_RV	agttagggtcatatgtatatctccttccccgggtaccgagtcaggtttgttcagaacg cg	Assembly of re-organized *mca* pathway
mcaF_into_mcaG_FW	tgaacaaacctgactcggtacccggggaaggagatatacatatgaccctaacttct atgc	Assembly of re-organized *mca* pathway
mcaF_into_123i_RV	ggcaaattctgttttatcagaccgcttctgcgttctgatttactaactaaaatcttctcg	Assembly of re-organized *mca* pathway

F.1.2 Microcyclamide biosynthetic pathway plasmid construction

The wildtype *mca* pathway was assembled as four separate fragments that were amplified using the following primers, respectively: mca_frag1_5'_toMQ123i / mca_frag1_3', mca_frag2_5' / mca_frag2_3', mca_frag3_5'/mca_frag3_3', mca_frag4_5'/mca_frag4_3'_toMQ123i. In order to construct the *mcaADEG*, the *mcaADG* expression plasmid was first cloned at the SmaI site of pMQ123i using three fragments using primers: mca_frag1_5'_toMQ123i/mcaA-123i-3', mcaD-123i-5'/mcaD-123i-3', mcaG-123i-5'/mca_frag4_3'_toMQ123i. The *mcaE* gene was then cloned

into the *mcaADG* construct at the Eco53kI restriction site. Primers McaE-5'-insert-to-123iMcaADG/McaE-3'-insert-to-123iMcaADG were used to clone the *mcaE* gene as a single fragment to yield the *mcaADEG* construct. Primers (McaE-5'-insert-to-123iMcaADG/mcaA-123i-3', mcaD-123i-5' / mcaD-123i-3', mcaG-123i-5' / mcaG_into_mcaF_RV, mcaF_into_ mcaG_FW / mcaF_into_123i_RV) were used to generate four fragments to assemble the *mcaADEFG* construct at the SmaI site of pMQ123i in a single cloning step. The *mcaADEG* plasmid was used as a PCR template to amplify the *mcaEA* fragments necessary to assemble the *mcaADEFG* pathway. The WT *mca* plasmid was used as PCR template to amplify the remaining fragments for assembly of the *mcaADEFG* pathway.

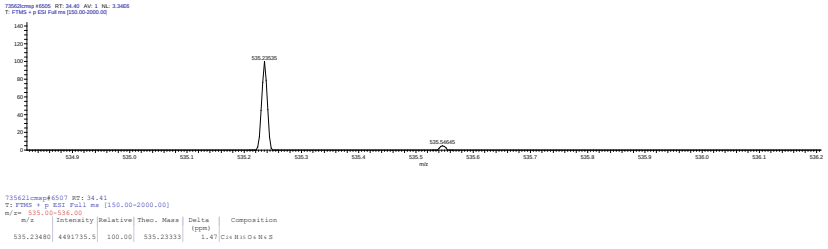
F.1.3 Mass Spectrometry Analysis of Microcyclamide Structures from Heterologous Expression Experiments



Peak	Molecular Formula	Accurate Mass	Observed Mass	Error (ppm)
A	C24H35N6O6S	535.23333	535.2348	2.7
B	C21H31N5O5S4S	463.2122	463.21295	1.6
C	C21H28N5O4S	446.18565	446.18579	0.3
D	C19H28N5O3S	406.19074	406.19141	1.6
E	C17H26N5O3S	380.17509	380.17551	1.1
F	C17H23N4O3S	363.14854	363.14932	2.1
G	C13H17N4O3S	309.10159	309.10236	2.5
H	C12H19N4O3	267.14517	267.14581	2.4
I	C12H19N2O2S	251.08487	251.08522	1.4
J	C11H15N2O2S	223.08996	223.09026	1.3
K	C9H13N2OS	197.07431	197.0746	1.5
L	C8H14N3O2	184.10805	184.10846	2.2
M	C8H11N2O2	167.0815	167.08177	1.6

E. coli mca wildtype pathway

Accurate mass for microcyclamide 7806B

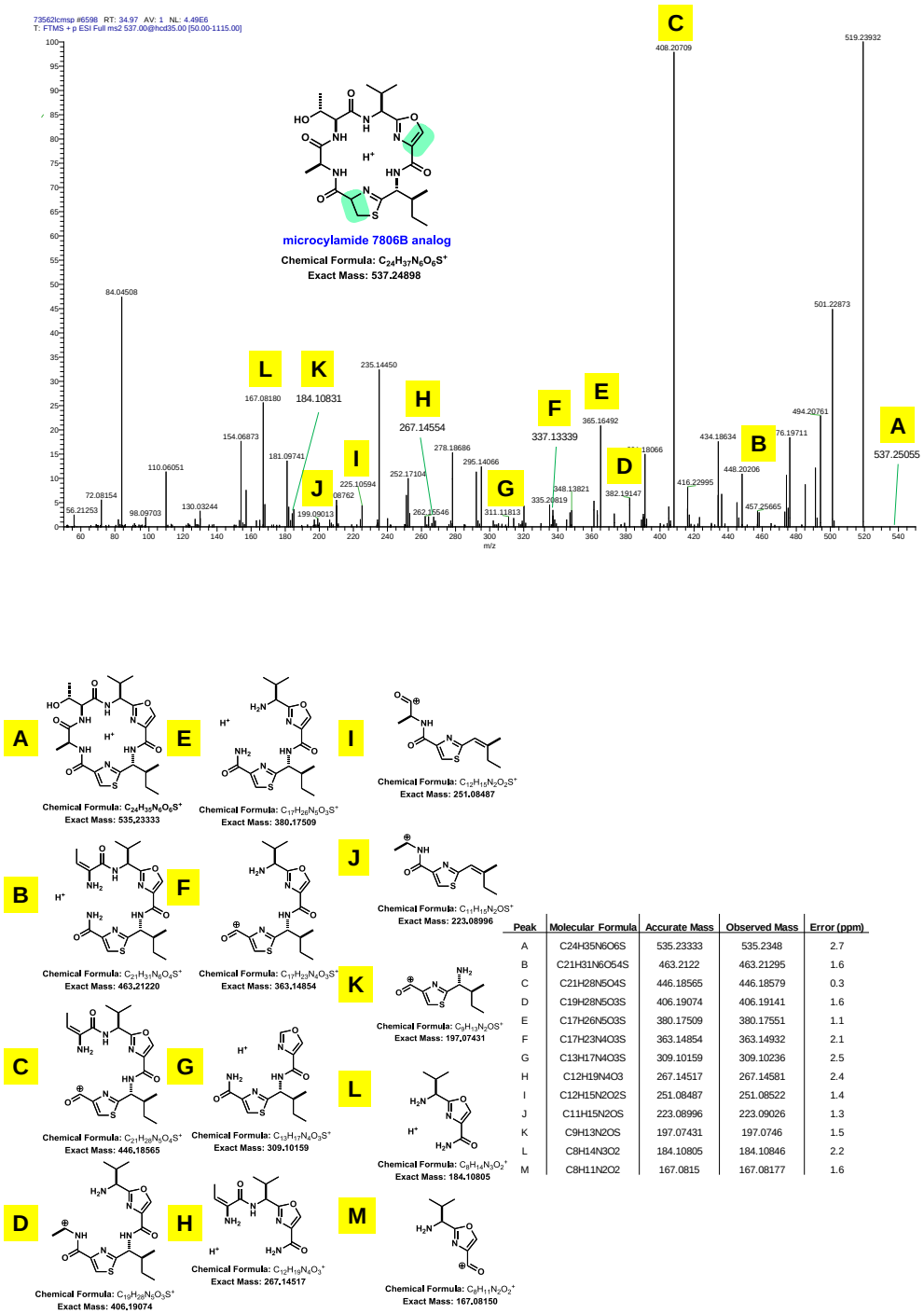


735621cmap#6507 RT: 34.41
T: FTMS + p ESI Full ms [150.00-2000.00]
m/z= 535.00-536.00

m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
535.23480	4491735.5	100.00	535.23333	1.47	C24 H35 O6 N6 S

Figure F1 MS and MSMS analysis of microcyclamide 7806B from *E. coli* heterologous expression.

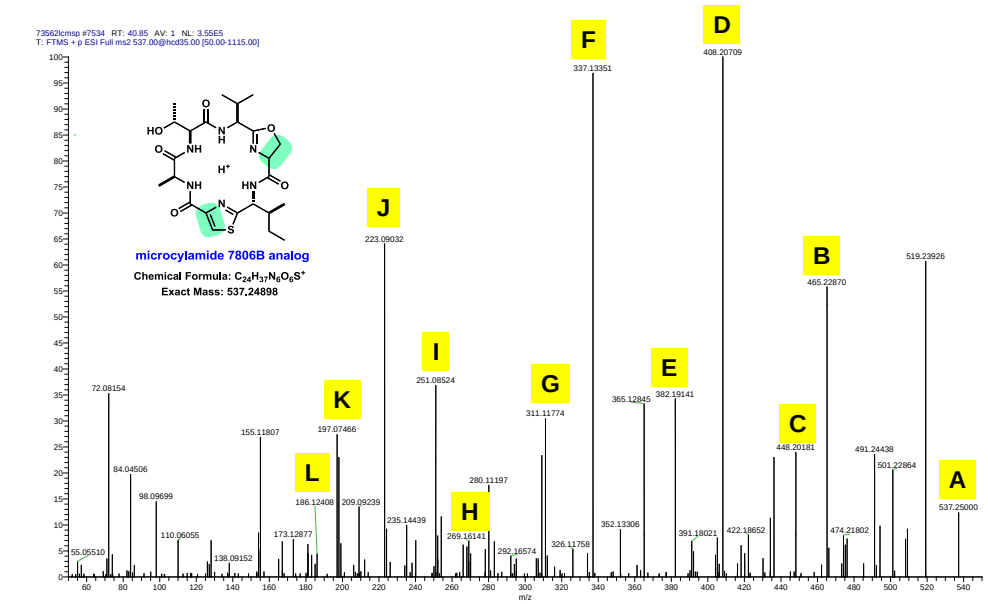
E. coli mca wildtype pathway

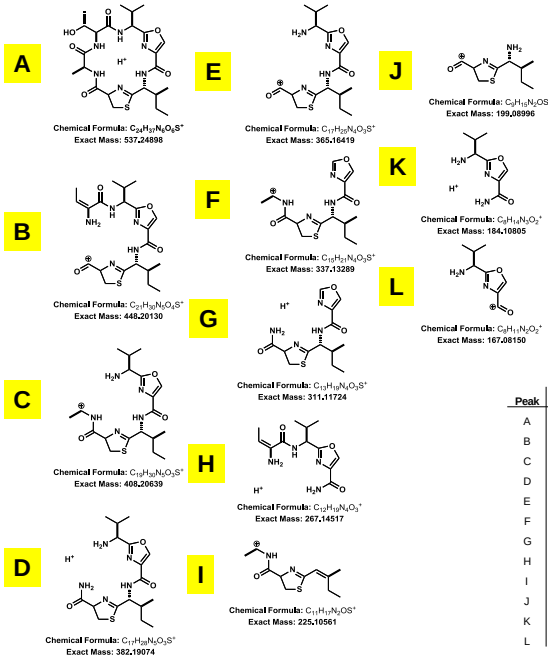


Peak	Molecular Formula	Accurate Mass	Observed Mass	Error (ppm)
A	C24H35N6O6S	535.23333	535.2348	2.7
B	C21H31N6O54S	463.2122	463.21295	1.6
C	C21H28N5O4S	446.18565	446.18579	0.3
D	C19H28N5O3S	406.19074	406.19141	1.6
E	C17H26N5O3S	380.17509	380.17551	1.1
F	C17H23N4O3S	363.14854	363.14932	2.1
G	C13H17N4O3S	309.10159	309.10236	2.5
H	C12H19N4O3	267.14517	267.14581	2.4
I	C12H15N2O2S	251.08487	251.08522	1.4
J	C11H15N2OS	223.08996	223.09026	1.3
K	C9H13N2OS	197.07431	197.0746	1.5
L	C8H14N3O2	184.10805	184.10846	2.2
M	C8H11N2O2	167.0815	167.08177	1.6

Figure F2 MS and MSMS analysis of microcyclamide 7806B analog from *E. coli* heterologous expression.

E. coli mca wildtype pathway

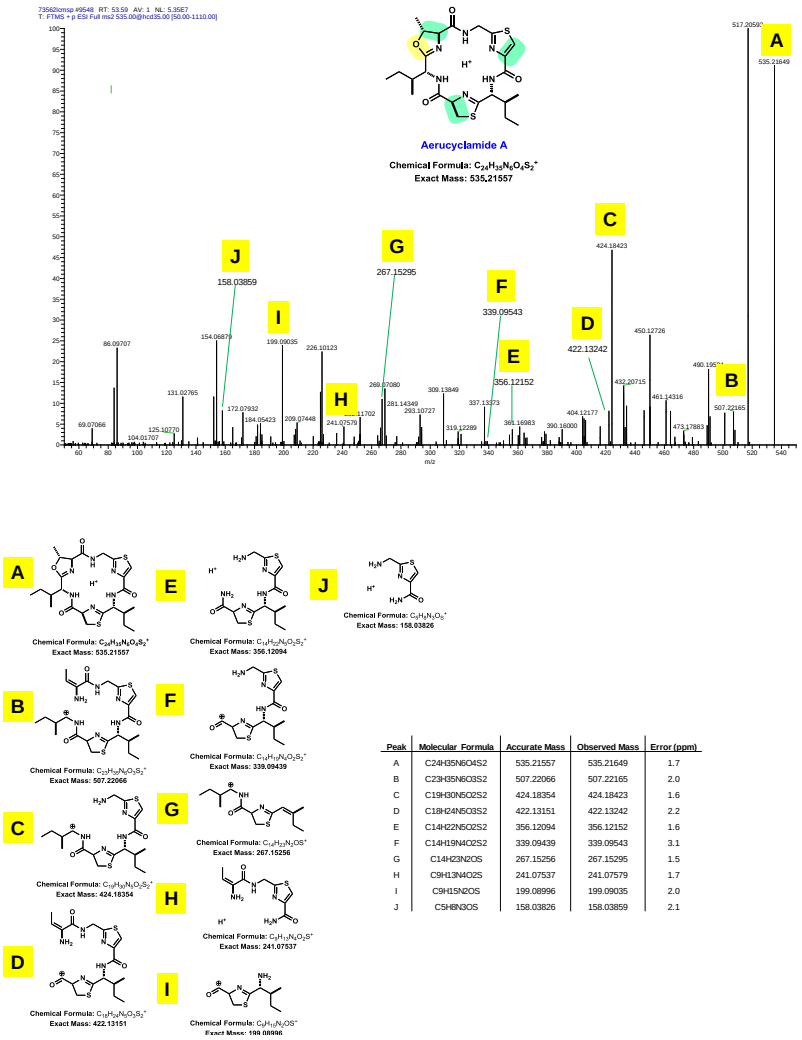




Peak	Molecular Formula	Accurate Mass	Observed Mass	Error (ppm)
A	C24H37N6O6S	537.24898	537.25	1.9
B	C21H33N6O4S	465.22785	465.2287	1.8
C	C21H30N5O4S	448.2013	448.20181	1.1
D	C19H30N5O3S	408.20639	408.20709	1.7
E	C17H28N5O3S	382.19074	382.19141	1.8
F	C15H21N4O3S	337.13289	337.13351	1.8
G	C13H19N4O3S	311.11724	311.11774	1.6
H	C12H21N4O3	269.16082	269.16141	2.2
I	C12H15N2O2S	251.08487	251.08524	1.5
J	C11H15N2OS	223.08996	223.09032	1.6
K	C9H13N2OS	197.07431	197.07466	1.8
L	C8H16N3OS	186.1237	186.12408	2.0

Figure F3 MS and MSMS analysis of microcyclamide 7806B analog from *E. coli* heterologous expression.

E. coli mca wildtype pathway



Peak	Molecular Formula	Accurate Mass	Observed Mass	Error (ppm)
A	C24H35N6O4S2	535.21557	535.21649	1.7
B	C23H35N6O3S2	507.22066	507.22165	2.0
C	C19H30N5O2S2	424.18354	424.18423	1.6
D	C18H24N5O3S2	422.13151	422.13242	2.2
E	C14H22N5O2S2	356.12094	356.12152	1.6
F	C14H19N4O2S2	339.09439	339.09543	3.1
G	C14H23N2OS	267.15256	267.15295	1.5
H	C9H13N4O2S	241.07537	241.07579	1.7
I	C9H15N2OS	199.08996	199.09035	2.0
J	C5H8N3OS	158.03826	158.03859	2.1

E. coli mca wildtype pathway

Accurate mass for aerucyclamide A

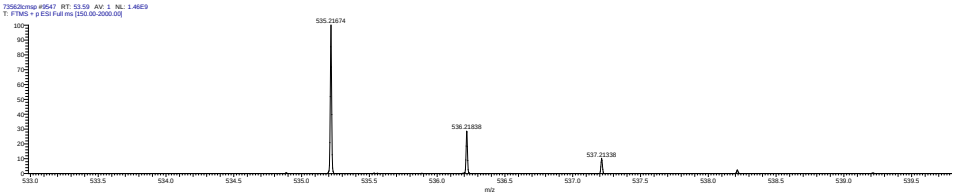


Figure F4 MS and MSMS analysis of aerucyclamide A from *E. coli* heterologous expression.

E. coli mca wildtype pathway

Accurate mass for aerucyclamide B

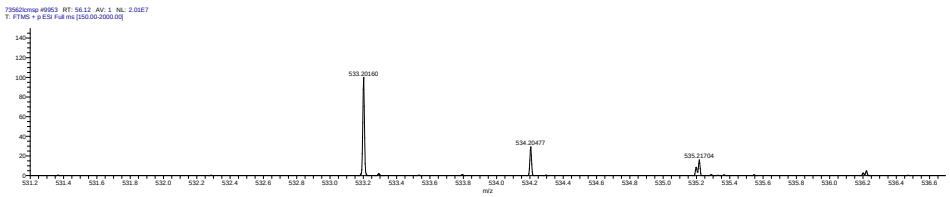
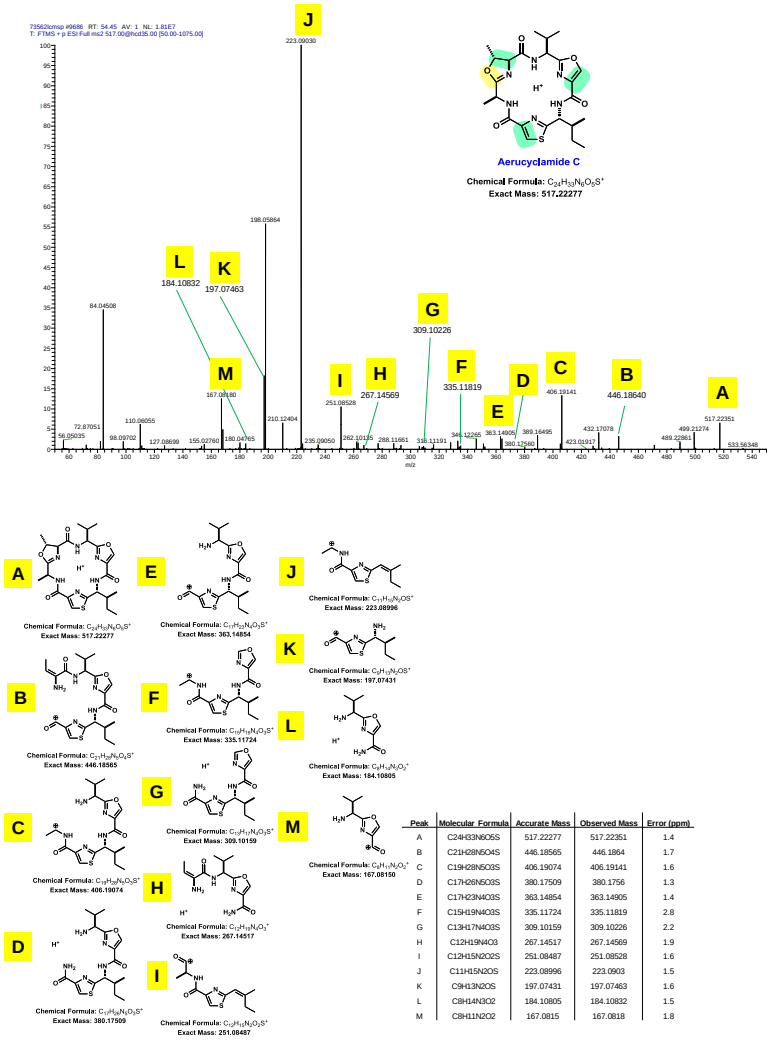


Figure F5 MS analysis of aerucyclamide B from *E. coli* heterologous expression.

E. coli mca wildtype pathway



Peak	Molecular Formula	Accurate Mass	Observed Mass	Error (ppm)
A	C24H33N6O5S	517.22277	517.22351	1.4
B	C21H28N5O4S	446.18565	446.1864	1.7
C	C19H28N5O3S	406.19074	406.19141	1.6
D	C17H26N5O3S	380.17509	380.1756	1.3
E	C17H23N4O3S	363.14854	363.14905	1.4
F	C15H19N4O3S	335.11724	335.11819	2.8
G	C13H17N4O3S	309.10159	309.10226	2.2
H	C12H19N4O3	267.14517	267.14569	1.9
I	C12H15N2O2S	251.08487	251.08528	1.6
J	C11H15N2OS	223.08996	223.0903	1.5
K	C9H13N2OS	197.07431	197.07463	1.6
L	C8H14N3O2	184.10805	184.10832	1.5
M	C8H11N2O2	167.0815	167.0818	1.8

E. coli mca wildtype pathway

Accurate mass for aerucyclamide C

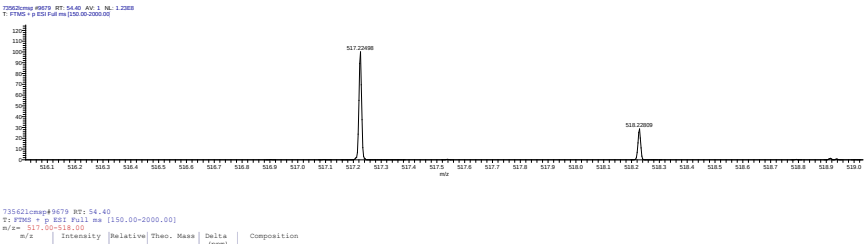
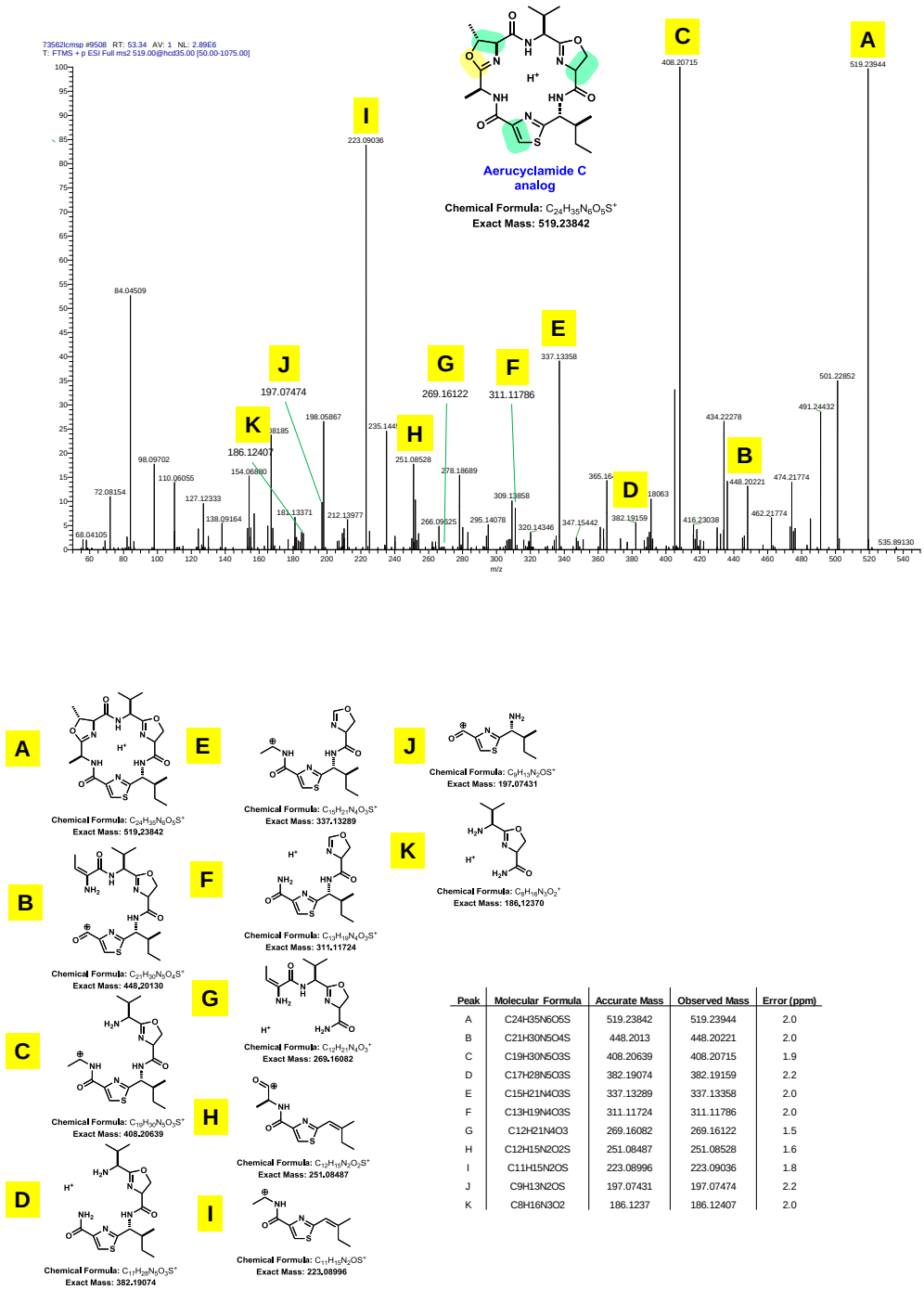


Figure F6 MS and MSMS analysis of aerucyclamide C from *E. coli* heterologous expression.

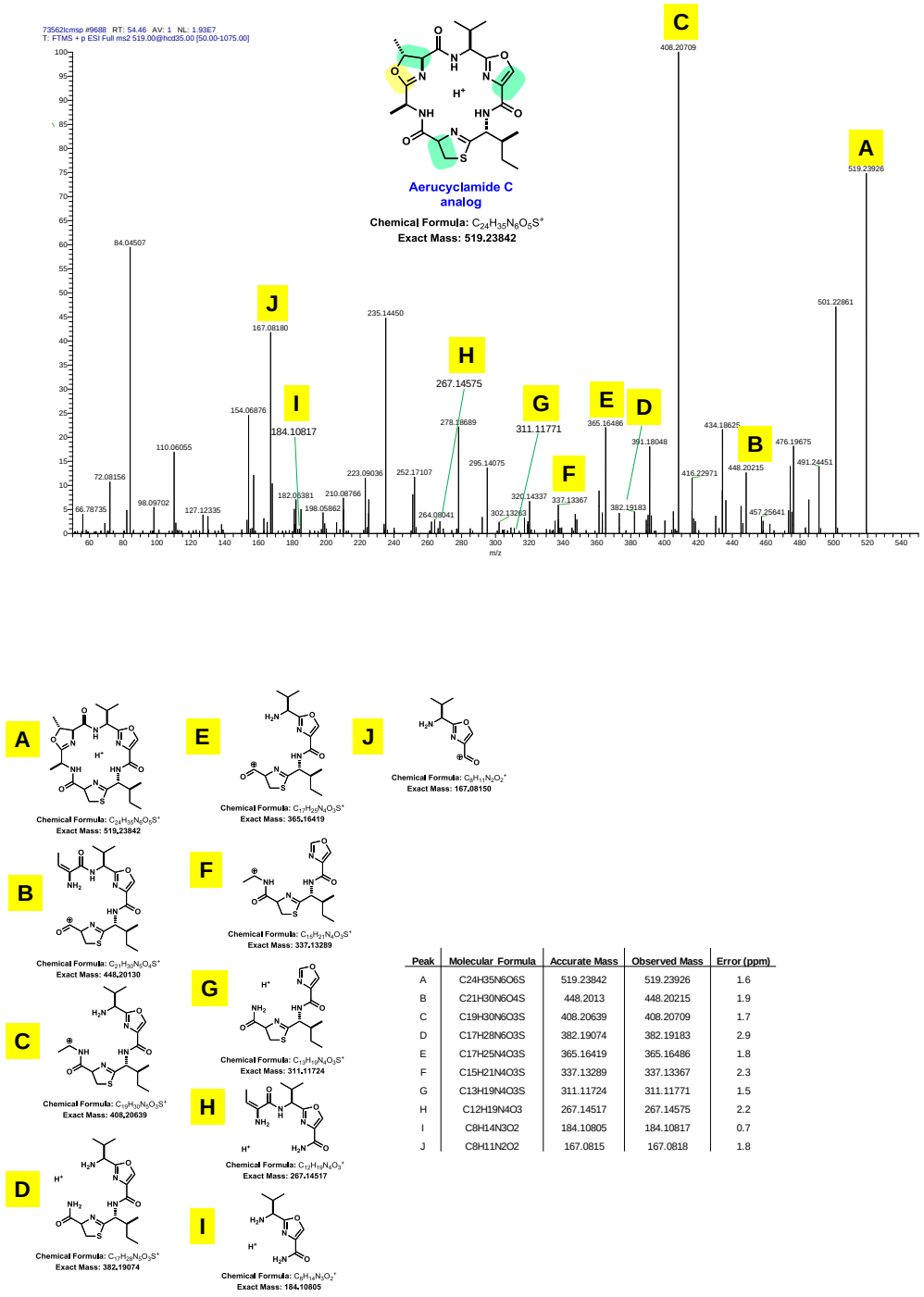
E. coli mca wildtype pathway



Peak	Molecular Formula	Accurate Mass	Observed Mass	Error (ppm)
A	C24H35N6O5S	519.23842	519.23944	2.0
B	C21H30N5O4S	448.2013	448.20221	2.0
C	C19H30N5O3S	408.20639	408.20715	1.9
D	C17H28N5O3S	382.19074	382.19159	2.2
E	C15H21N4O3S	337.13289	337.13358	2.0
F	C13H19N4O3S	311.11724	311.11786	2.0
G	C12H21N4O3	269.16082	269.16122	1.5
H	C12H15N2O2S	251.08487	251.08528	1.6
I	C11H15N2OS	223.08996	223.09036	1.8
J	C9H13N2OS	197.07431	197.07474	2.2
K	C8H16N3O2	186.1237	186.12407	2.0

Figure F7 MS and MSMS analysis of aerucyclamide C analog from *E. coli* heterologous expression.

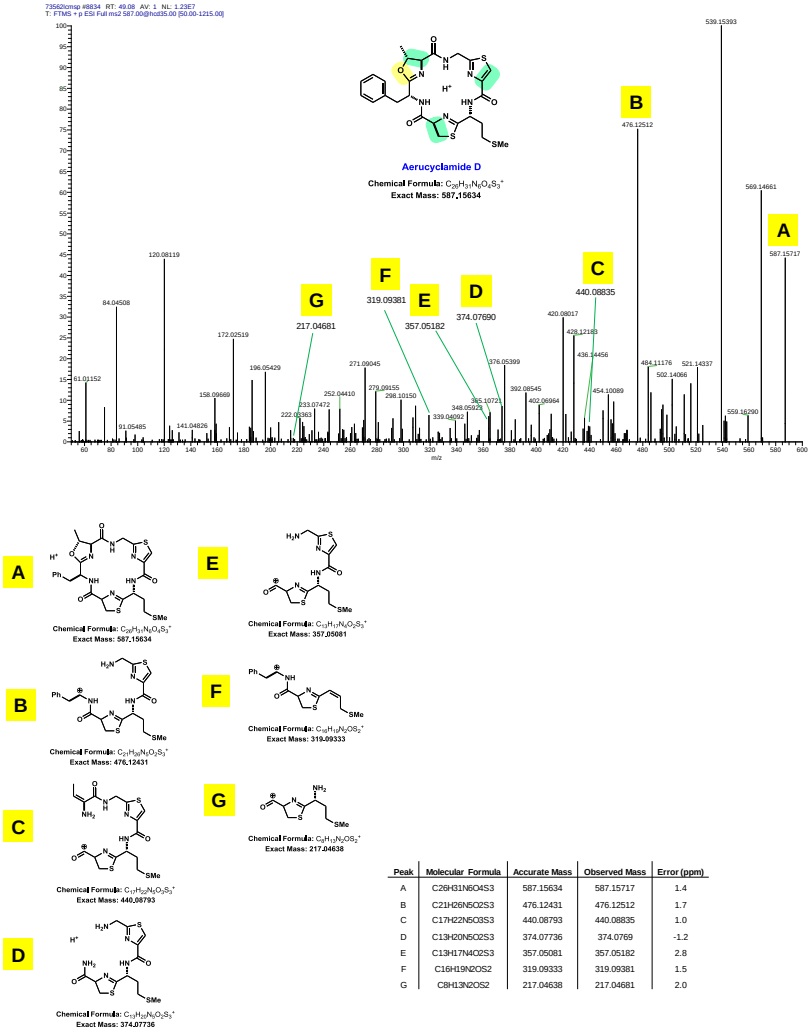
E. coli mca wildtype pathway



Peak	Molecular Formula	Accurate Mass	Observed Mass	Error (ppm)
A	C24H35N6O6S	519.23842	519.23926	1.6
B	C21H30N6O4S	448.2013	448.20215	1.9
C	C19H30N6O3S	408.20639	408.20709	1.7
D	C17H28N6O3S	382.19074	382.19183	2.9
E	C17H25N4O3S	365.16419	365.16486	1.8
F	C15H21N4O3S	337.13289	337.13367	2.3
G	C13H19N4O3S	311.11724	311.11771	1.5
H	C12H19N4O3	267.14517	267.14575	2.2
I	C8H14N3O2	184.10805	184.10817	0.7
J	C8H11N2O2	167.0815	167.0818	1.8

Figure F8 MS and MSMS analysis of aerucyclamide C analog from *E. coli* heterologous expression.

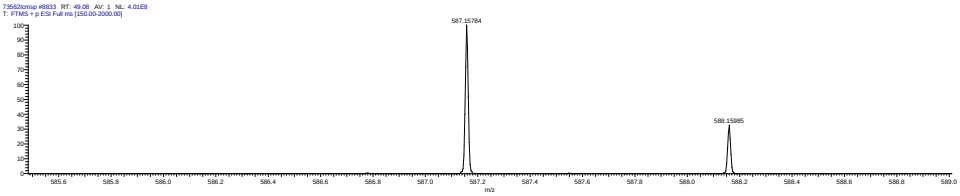
E. coli mca wildtype pathway



Peak	Molecular Formula	Accurate Mass	Observed Mass	Error (ppm)
A	C26H31N6O4S3	587.15634	587.15717	1.4
B	C21H26N5O2S3	476.12431	476.12512	1.7
C	C17H22N5O3S3	440.08793	440.08835	1.0
D	C13H20N5O2S3	374.07736	374.0769	-1.2
E	C13H17N4O2S3	357.05081	357.05182	2.8
F	C16H19N2OS2	319.09333	319.09381	1.5
G	C8H13N2OS2	217.04638	217.04681	2.0

E. coli mca wildtype pathway

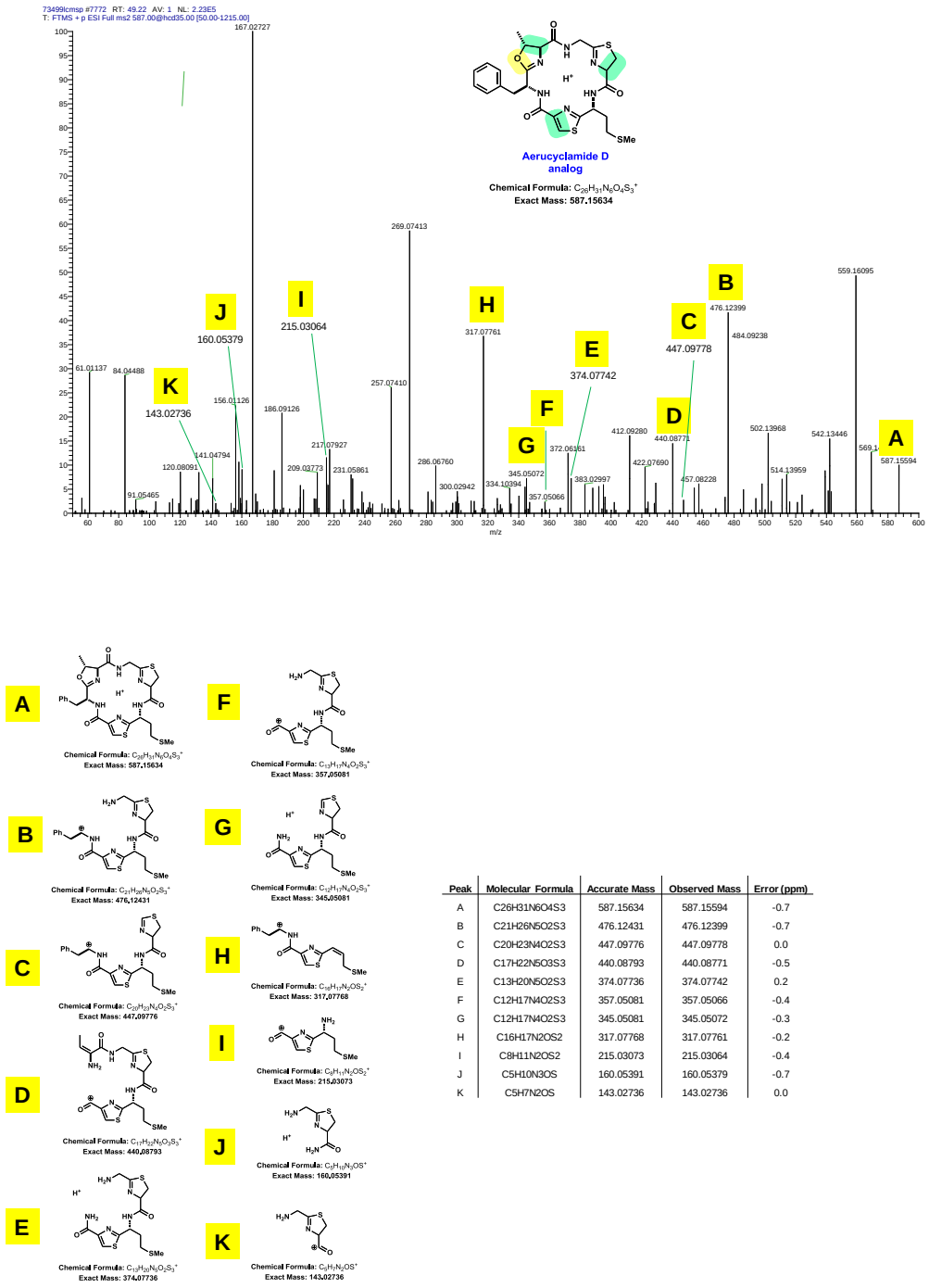
Accurate mass for aerucyclamide D



m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
587.15784	403404960.0	100.00	587.15634	1.49	C26 H31 O4 N6 S3

Figure F9 MS and MSMS analysis of aerucyclamide D from *E. coli* heterologous expression.

E. coli mcaEADGF pathway



Peak	Molecular Formula	Accurate Mass	Observed Mass	Error (ppm)
A	C26H31N6O4S3	587.15634	587.15594	-0.7
B	C21H26N5O2S3	476.12431	476.12399	-0.7
C	C20H23N4O2S3	447.09776	447.09778	0.0
D	C17H22N5O3S3	440.08793	440.08771	-0.5
E	C13H20N5O2S3	374.07736	374.07742	0.2
F	C12H17N4O2S3	357.05081	357.05066	-0.4
G	C12H17N4O2S3	345.05081	345.05072	-0.3
H	C16H17N2OS2	317.07768	317.07761	-0.2
I	C8H11N2OS2	215.03073	215.03064	-0.4
J	C5H10N3OS	160.05391	160.05379	-0.7
K	C5H7N2OS	143.02736	143.02736	0.0

Figure F10 MS and MSMS analysis of aerucyclamide D analog from *E. coli* heterologous expression.

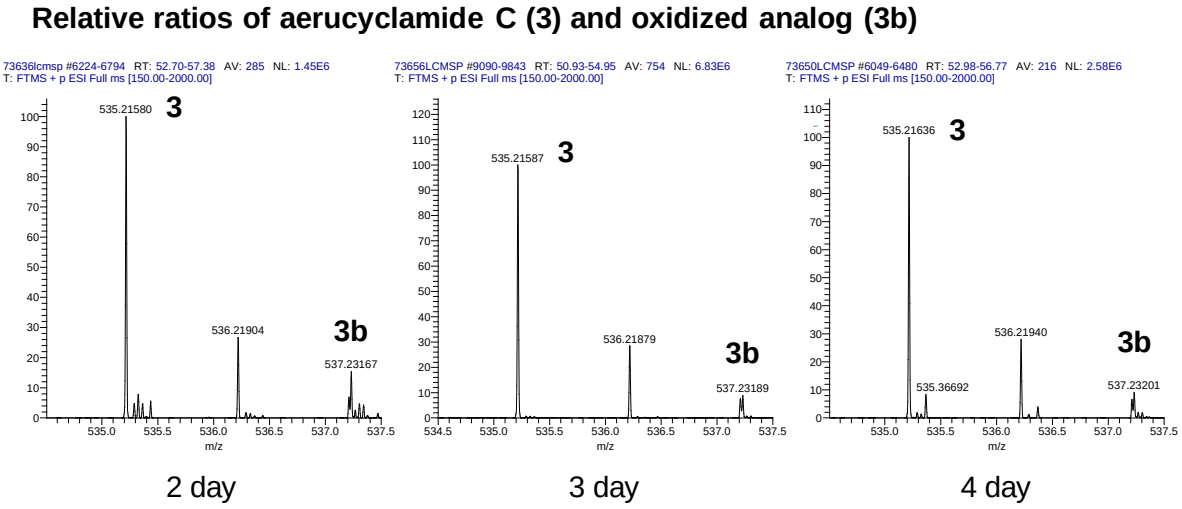


Figure F11 MS analysis to quantify relative ratios of aerucyclamide C (**3**) and its oxidized analog (**3b**) over 2 to 4 days from heterologous expression in *E. coli*.

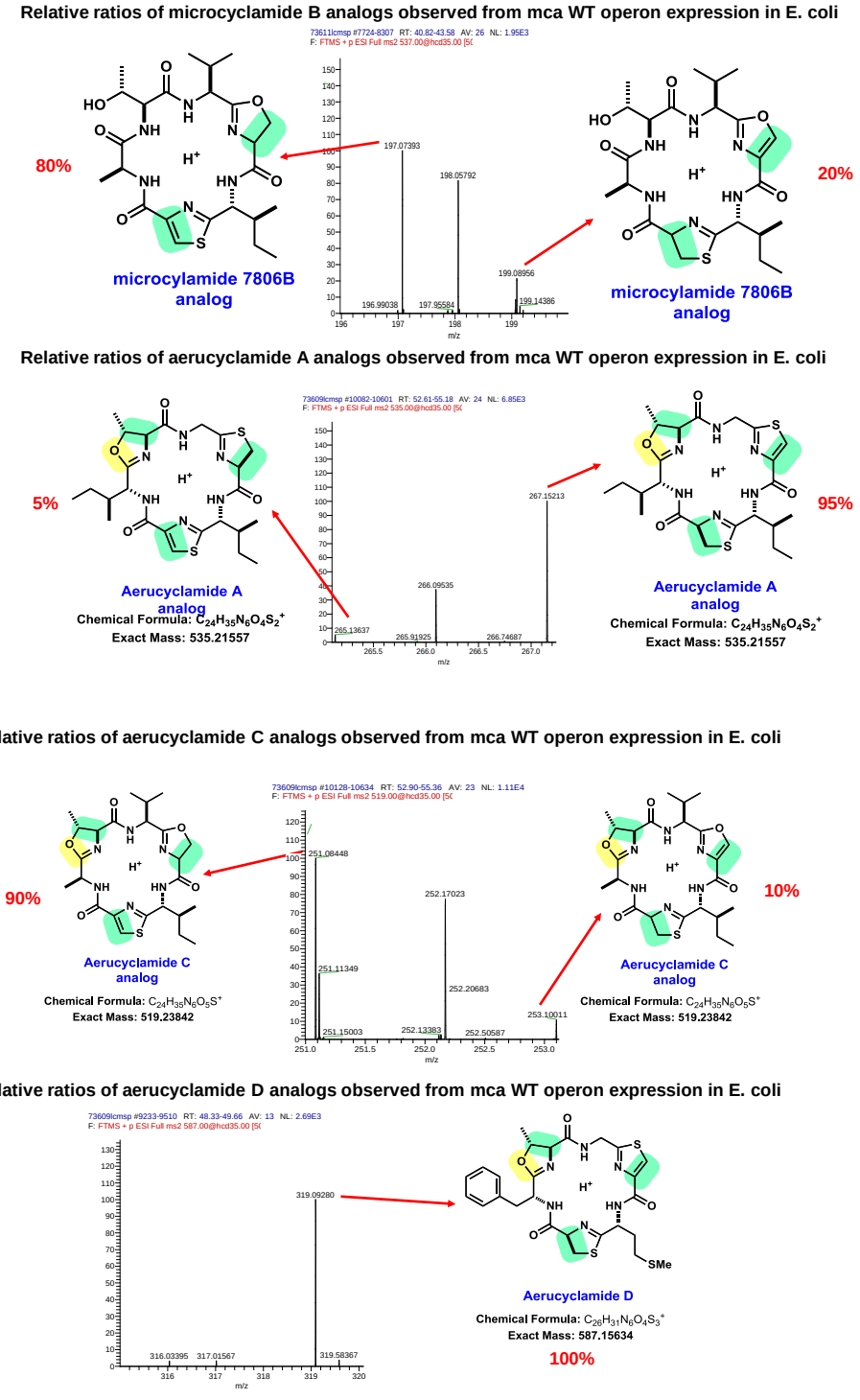


Figure F12 MSMS analysis to quantify relative ratios of microcyclamide and aerucyclamide oxidative analogs from heterologous expression in *E. coli*. Relative abundance approximates are highlighted in red.

REFERENCES

1. Livermore, D.M., *The need for new antibiotics.* Clinical Microbiology and Infection, 2004. **10**: p. 1-9.
2. Levy, S.B. and B. Marshall, *Antibacterial resistance worldwide: causes, challenges and responses.* Nat Med, 2004.
3. Wright, G.D., *Bacterial resistance to antibiotics: Enzymatic degradation and modification.* Advanced Drug Delivery Reviews, 2005. **57**(10): p. 1451-1470.
4. Roy, P.H., *Horizontal transfer of genes in bacteria.* Microbiology today, 1999. **26**: p. 168.
5. Crunkhorn, S., *Antibacterial drugs: New paths to beating bacteria.* Nat Rev Drug Discov, 2008. **7**(11): p. 891-891.
6. Travis, J. and J. Potempa, *Bacterial proteinases as targets for the development of second-generation antibiotics.* Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology, 2000. **1477**(1–2): p. 35-50.
7. Fronzes, R., H. Remaut, and G. Waksman, *Architectures and biogenesis of non-flagellar protein appendages in Gram-negative bacteria.* The EMBO Journal, 2008. **27**(17): p. 2271-2280.
8. Stamm, W.E. and S.R. Norrby, *Urinary Tract Infections: Disease Panorama and Challenges.* Journal of Infectious Diseases, 2001. **183**(Supplement 1): p. S1-S4.
9. Svensson, M., et al., *Carbohydrate Receptor Depletion as an Antimicrobial Strategy for Prevention of Urinary Tract Infection.* Journal of Infectious Diseases, 2001. **183**(Supplement 1): p. S70-S73.
10. Pinkner, J.S., et al., *Rationally designed small compounds inhibit pilus biogenesis in uropathogenic bacteria.* Proceedings of the National Academy of Sciences, 2006. **103**(47): p. 17897-17902.
11. Lo, A.W.H., et al., *Suppression of type 1 pilus assembly in uropathogenic Escherichia coli by chemical inhibition of subunit polymerization.* Journal of Antimicrobial Chemotherapy, 2013.
12. O'Brien, A.D., et al., *Shigellosis and Escherichia coli diarrhea: relative importance of invasive and toxigenic mechanisms.* The American Journal of Clinical Nutrition, 1979. **32**(1): p. 229-33.

13. Schiavo, G. and F.G. van der Goot, *The bacterial toxin toolkit.* Nat Rev Mol Cell Biol, 2001. **2**(7): p. 530-537.
14. Kaper, J.B., J.P. Nataro, and H.L.T. Mobley, *Pathogenic Escherichia coli.* Nat Rev Micro, 2004. **2**(2): p. 123-140.
15. Abrami, L., et al., *Anthrax toxin triggers endocytosis of its receptor via a lipid raft–mediated clathrin-dependent process.* The Journal of Cell Biology, 2003. **160**(3): p. 321-328.
16. Chen, Z., et al., *Efficient Neutralization of Anthrax Toxin by Chimpanzee Monoclonal Antibodies against Protective Antigen.* Journal of Infectious Diseases, 2006. **193**(5): p. 625-633.
17. Armstrong, G.D., et al., *A Phase I Study of Chemically Synthesized Verotoxin (Shiga-like Toxin) Pk-Trisaccharide Receptors Attached to Chromosorb for Preventing Hemolytic-Uremic Syndrome.* Journal of Infectious Diseases, 1995. **171**(4): p. 1042-1045.
18. Shoop, W.L., et al., *Anthrax lethal factor inhibition.* Proceedings of the National Academy of Sciences, 2005. **102**(22): p. 7958-7963.
19. Scobie, H.M., et al., *A Soluble Receptor Decoy Protects Rats against Anthrax Lethal Toxin Challenge.* Journal of Infectious Diseases, 2005. **192**(6): p. 1047-1051.
20. Hall-Stoodley, L., et al., *DIrect detection of bacterial biofilms on the middle-ear mucosa of children with chronic otitis media.* JAMA, 2006. **296**(2): p. 202-211.
21. Carron, M.A., et al., *Identification of Helicobacter pylori Biofilms in Human Gastric Mucosa.* Journal of Gastrointestinal Surgery, 2006. **10**(5): p. 712-717.
22. Lam, J., et al., *Production of mucoid microcolonies by Pseudomonas aeruginosa within infected lungs in cystic fibrosis.* Infection and Immunity, 1980. **28**(2): p. 546-556.
23. Singh, P.K., et al., *Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms.* Nature, 2000. **407**(6805): p. 762-764.
24. Benghezal, M., et al., *Inhibitors of bacterial virulence identified in a surrogate host model.* Cellular Microbiology, 2007. **9**(5): p. 1336-1342.
25. Davey, M.E., N.C. Caiazza, and G.A. O'Toole, *Rhamnolipid Surfactant Production Affects Biofilm Architecture in Pseudomonas aeruginosa PAO1.* Journal of Bacteriology, 2003. **185**(3): p. 1027-1036.
26. Hueck, C.J., *Type III Protein Secretion Systems in Bacterial Pathogens of Animals and Plants.* Microbiology and Molecular Biology Reviews, 1998. **62**(2): p. 379-433.
27. Journet, L., K.T. Hughes, and G.R. Cornelis, *Type III secretion: a secretory pathway serving both motility and virulence (Review).* Molecular Membrane Biology, 2005. **22**(1-2): p. 41-50.
28. Bailey, L., et al., *Small molecule inhibitors of type III secretion in Yersinia block the Chlamydia pneumoniae infection cycle.* FEBS Letters, 2007. **581**(4): p. 587-595.
29. Muschiol, S., et al., *A small-molecule inhibitor of type III secretion inhibits different stages of the infectious cycle of Chlamydia trachomatis.* Proceedings of the National Academy of Sciences, 2006. **103**(39): p. 14566-14571.
30. Broderick, A.H., et al., *Surface Coatings that Promote Rapid Release of Peptide-Based AgrC Inhibitors for Attenuation of Quorum Sensing in Staphylococcus aureus.* Advanced Healthcare Materials, 2014. **3**(1): p. 97-105.
31. Welsh, M.A., et al., *Small Molecule Disruption of Quorum Sensing Cross-Regulation in Pseudomonas aeruginosa Causes Major and Unexpected Alterations to Virulence Phenotypes.* Journal of the American Chemical Society, 2015.

32. Rasko, D.A., et al., *Targeting QseC Signaling and Virulence for Antibiotic Development.* Science, 2008. **321**(5892): p. 1078-1080.
33. Saha, R., et al., *Microbial siderophores: a mini review.* Journal of Basic Microbiology, 2013. **53**(4): p. 303-317.
34. Rusnak, F., et al., *Biosynthesis of the Escherichia coli siderophore enterobactin: sequence of the entF gene, expression and purification of EntF, and analysis of covalent phosphopantetheine.* Biochemistry, 1991. **30**(11): p. 2916-2927.
35. Nahlik, M.S., T.P. Fleming, and M.A. McIntosh, *Cluster of genes controlling synthesis and activation of 2,3-dihydroxybenzoic acid in production of enterobactin in Escherichia coli.* Journal of Bacteriology, 1987. **169**(9): p. 4163-4170.
36. Chenault, S.S. and C.F. Earhart, *Organization of genes encoding membrane proteins of the Escherichia coli ferrienterobactin permease.* Molecular Microbiology, 1991. **5**(6): p. 1405-1413.
37. Carrano, C.J. and K.N. Raymond, *Ferric ion sequestering agents. 2. Kinetics and mechanism of iron removal from transferrin by enterobactin and synthetic tricatechols.* Journal of the American Chemical Society, 1979. **101**(18): p. 5401-5404.
38. Raymond, K.N., E.A. Dertz, and S.S. Kim, *Enterobactin: An archetype for microbial iron transport.* Proceedings of the National Academy of Sciences, 2003. **100**(7): p. 3584-3588.
39. Martin, P., et al., *Interplay between Siderophores and Colibactin Genotoxin Biosynthetic Pathways in Escherichia coli.* PLoS Pathogens, 2013. **9**(7): p. e1003437.
40. Nielsen, A., et al., *Nigribactin, a Novel Siderophore from Vibrio nigripulchritudo, Modulates Staphylococcus aureus Virulence Gene Expression.* Marine Drugs, 2012. **10**(11): p. 2584-2595.
41. Imperi, F., et al., *Repurposing the antimycotic drug flucytosine for suppression of Pseudomonas aeruginosa pathogenicity.* Proceedings of the National Academy of Sciences, 2013. **110**(18): p. 7458-7463.
42. Fraser, C.M., J.A. Eisen, and S.L. Salzberg, *Microbial genome sequencing.* Nature, 2000. **406**(6797): p. 799-803.
43. Belda-Ferre, P., et al., *Mining Virulence Genes Using Metagenomics.* PLoS ONE, 2011. **6**(10): p. e24975.
44. Wyatt, M.A., et al., *Staphylococcus aureus Nonribosomal Peptide Secondary Metabolites Regulate Virulence.* Science, 2010. **329**(5989): p. 294-296.
45. Hentzer, M., et al., *Attenuation of Pseudomonas aeruginosa virulence by quorum sensing inhibitors*. Vol. 22. 2003. 3803-3815.
46. Biggins, J.B., M.A. Ternei, and S.F. Brady, *Malleilactone, a Polyketide Synthase-Derived Virulence Factor Encoded by the Cryptic Secondary Metabolome of Burkholderia pseudomallei Group Pathogens.* Journal of the American Chemical Society, 2012. **134**(32): p. 13192-13195.
47. Tan, M.-W., et al., *Pseudomonas aeruginosa killing of Caenorhabditis elegans used to identify P. aeruginosa virulence factors.* Proceedings of the National Academy of Sciences, 1999. **96**(5): p. 2408-2413.
48. Gordon, N.C. and D.W. Wareham, *Multidrug-resistant Acinetobacter baumannii: mechanisms of virulence and resistance.* International Journal of Antimicrobial Agents, 2010. **35**(3): p. 219-226.

49. Bergogne-Bérézin, E., D. Decreé, and M.-L. Joly-Guillou, *Opportunistic nosocomial multiply resistant bacterial infections—their treatment and prevention.* Journal of Antimicrobial Chemotherapy, 1993. **32**(suppl A): p. 39-47.
50. Vidal, R., et al., *Biofilm formation by Acinetobacter baumannii.* Microbios, 1996. **86**(346): p. 49-58.
51. Donlan, R.M. and J.W. Costerton, *Biofilms: Survival Mechanisms of Clinically Relevant Microorganisms.* Clinical Microbiology Reviews, 2002. **15**(2): p. 167-193.
52. Vidal, R., et al., *Effect of imipenem and sulbactam on sessile cells of Acinetobacter baumannii growing in biofilm.* Microbios, 1997. **91**(367): p. 79-87.
53. Vila, J. and J. Pachón, *Therapeutic options for Acinetobacter baumannii infections.* Expert Opinion on Pharmacotherapy, 2008. **9**(4): p. 587-599.
54. Michalopoulos, A. and M.E. Falagas, *Treatment of Acinetobacter infections.* Expert Opinion on Pharmacotherapy, 2010. **11**(5): p. 779-788.
55. Fournier, P.-E., et al., *Comparative Genomics of Multidrug Resistance in <italic>Acinetobacter baumannii</italic>.* PLoS Genet, 2006. **2**(1): p. e7.
56. Iacono, M., et al., *Whole-Genome Pyrosequencing of an Epidemic Multidrug-Resistant Acinetobacter baumannii Strain Belonging to the European Clone II Group.* Antimicrobial Agents and Chemotherapy, 2008. **52**(7): p. 2616-2625.
57. Choi, C.H., et al., *Outer membrane protein 38 of Acinetobacter baumannii localizes to the mitochondria and induces apoptosis of epithelial cells.* Cellular Microbiology, 2005. **7**(8): p. 1127-1138.
58. Choi, C.H., et al., *Acinetobacter baumannii outer membrane protein A targets the nucleus and induces cytotoxicity.* Cellular Microbiology, 2008. **10**(2): p. 309-319.
59. Siroy, A., et al., *Channel Formation by CarO, the Carbapenem Resistance-Associated Outer Membrane Protein of Acinetobacter baumannii.* Antimicrobial Agents and Chemotherapy, 2005. **49**(12): p. 4876-4883.
60. Clark, R.B., *Imipenem resistance among Acinetobacter baumannii: association with reduced expression of a 33–36 kDa outer membrane protein.* Journal of Antimicrobial Chemotherapy, 1996. **38**(2): p. 245-251.
61. Limansky, A.S., M.A. Mussi, and A.M. Viale, *Loss of a 29-Kilodalton Outer Membrane Protein in Acinetobacter baumannii Is Associated with Imipenem Resistance.* Journal of Clinical Microbiology, 2002. **40**(12): p. 4776-4778.
62. Choi, C., et al., *Acinetobacter baumannii invades epithelial cells and outer membrane protein A mediates interactions with epithelial cells.* BMC Microbiology, 2008. **8**(1): p. 216.
63. Gaddy, J.A., A.P. Tomaras, and L.A. Actis, *The Acinetobacter baumannii 19606 OmpA Protein Plays a Role in Biofilm Formation on Abiotic Surfaces and in the Interaction of This Pathogen with Eukaryotic Cells.* Infection and Immunity, 2009. **77**(8): p. 3150-3160.
64. Tomaras, A.P., et al., *Attachment to and biofilm formation on abiotic surfaces by Acinetobacter baumannii: involvement of a novel chaperone-usher pili assembly system.* Microbiology, 2003. **149**(12): p. 3473-3484.
65. Clemmer, K.M., R.A. Bonomo, and P.N. Rather, *Genetic analysis of surface motility in Acinetobacter baumannii.* Microbiology, 2011. **157**(Pt 9): p. 2534-2544.
66. Yamamoto, S., N. Okujo, and Y. Sakakibara, *Isolation and structure elucidation of acinetobactin., a novel siderophore from Acinetobacter baumannii.* Archives of Microbiology, 1994. **162**(4): p. 249-254.

67. Niu, C., et al., *Isolation and Characterization of an Autoinducer Synthase from Acinetobacter baumannii.* Journal of Bacteriology, 2008. **190**(9): p. 3386-3392.
68. Gaddy, J.A., et al., *Role of Acinetobactin-Mediated Iron Acquisition Functions in the Interaction of Acinetobacter baumannii Strain ATCC 19606(T) with Human Lung Epithelial Cells, Galleria mellonella Caterpillars, and Mice.* Infection and Immunity, 2012. **80**(3): p. 1015-1024.
69. Gaddy, J.A. and L.A. Actis, *Regulation of Acinetobacter baumannii biofilm formation.* Future Microbiology, 2009. **4**(3): p. 273-278.
70. Stacy, D.M., et al., *Attenuation of Quorum Sensing in the Pathogen Acinetobacter baumannii Using Non-native N-Acyl Homoserine Lactones.* ACS Chemical Biology, 2012. **7**(10): p. 1719-1728.
71. Smith, M.G., et al., *New insights into Acinetobacter baumannii pathogenesis revealed by high-density pyrosequencing and transposon mutagenesis.* Genes & Development, 2007. **21**(5): p. 601-614.
72. Rumbo-Feal, S., et al., *Whole Transcriptome Analysis of Acinetobacter baumannii Assessed by RNA-Sequencing Reveals Different mRNA Expression Profiles in Biofilm Compared to Planktonic Cells.* PLoS ONE, 2013. **8**(8): p. e72968.
73. Aranda, J., et al., *A rapid and simple method for constructing stable mutants of Acinetobacter baumannii.* BMC Microbiology, 2010. **10**: p. 279-279.
74. Shanks, R.M.Q., et al., *New yeast recombineering tools for bacteria.* Plasmid, 2009. **62**(2): p. 88-97.
75. Ugalde, J.E., et al., *Gene Organization and Transcription Analysis of the Agrobacterium tumefaciens Glycogen (glg) Operon: Two Transcripts for the Single Phosphoglucomutase Gene.* Journal of Bacteriology, 1998. **180**(24): p. 6557-6564.
76. Limberger, R.J., et al., *Insertional Inactivation of Treponema denticola tap1 Results in a Nonmotile Mutant with Elongated Flagellar Hooks.* Journal of Bacteriology, 1999. **181**(12): p. 3743-3750.
77. Goetting-Minesky, M.P. and J.C. Fenno, *A simplified erythromycin resistance cassette for Treponema denticola mutagenesis.* Journal of microbiological methods, 2010. **83**(1): p. 66-68.
78. Villegas, M.V. and A.I. Hartstein, *Acinetobacter Outbreaks, 1977–2000.* Infection Control & Hospital Epidemiology, 2003. **24**(04): p. 284-295.
79. Potts, M., *The anhydrobiotic cyanobacterial cell.* Physiologia Plantarum, 1996. **97**(4): p. 788-794.
80. Potts, M., *Desiccation tolerance of prokaryotes.* Microbiological Reviews, 1994. **58**(4): p. 755-805.
81. Dose, K., et al., *Survival in extreme dryness and DNA-single-strand breaks.* Advances in Space Research, 1992. **12**(4): p. 221-229.
82. Quadri, L.E.N., et al., *Characterization of Sfp, a Bacillus subtilis Phosphopantetheinyl Transferase for Peptidyl Carrier Protein Domains in Peptide Synthetases.* Biochemistry, 1998. **37**(6): p. 1585-1595.
83. Harshey, R.M., *Bacterial Motility on a Surface: Many Ways to a Common Goal.* Annual Review of Microbiology, 2003. **57**(1): p. 249-273.
84. Jarrell, K.F. and M.J. McBride, *The surprisingly diverse ways that prokaryotes move.* Nat Rev Micro, 2008. **6**(6): p. 466-476.

85. Kinsinger, R.F., M.C. Shirk, and R. Fall, *Rapid Surface Motility in Bacillus subtilis Is Dependent on Extracellular Surfactin and Potassium Ion.* Journal of Bacteriology, 2003. **185**(18): p. 5627-5631.
86. Caiazza, N.C., R.M.Q. Shanks, and G.A. O'Toole, *Rhamnolipids Modulate Swarming Motility Patterns of Pseudomonas aeruginosa.* Journal of Bacteriology, 2005. **187**(21): p. 7351-7361.
87. Shanks, R.M.Q., et al., *Serratamolide is a Hemolytic Factor Produced by Serratia marcescens.* PLoS ONE, 2012. **7**(5): p. e36398.
88. Spiers, A.J., et al., *Adaptive Divergence in Experimental Populations of Pseudomonas fluorescens. I. Genetic and Phenotypic Bases of Wrinkly Spreader Fitness.* Genetics, 2002. **161**(1): p. 33-46.
89. Peleg, A.Y., et al., *Galleria mellonella as a Model System To Study Acinetobacter baumannii Pathogenesis and Therapeutics.* Antimicrobial Agents and Chemotherapy, 2009. **53**(6): p. 2605-2609.
90. Gophna, U. and E.Z. Ron, *Virulence and the heat shock response.* International Journal of Medical Microbiology, 2003. **292**(7–8): p. 453-461.
91. Sokolovic, Z. and W. Goebel, *Synthesis of listeriolysin in Listeria monocytogenes under heat shock conditions.* Infection and Immunity, 1989. **57**(1): p. 295-298.
92. Clemmer, K.M., R.A. Bonomo, and P.N. Rather, *Genetic analysis of surface motility in Acinetobacter baumannii.* Microbiology, 2011. **157**(9): p. 2534-2544.
93. Kracht M, R.H., Ozel M, Kowall M, Pauli G, Vater J., *Antiviral and hemolytic activities of surfactin isoforms and their methyl ester derivatives.* J Antibiot, 1999. **52**(7): p. 613-619.
94. Johnson, M.K. and D. Boese-Marrazzo, *Production and properties of heat-stable extracellular hemolysin from Pseudomonas aeruginosa.* Infection and Immunity, 1980. **29**(3): p. 1028-1033.
95. Finn, S., et al., *Transcriptomic Responses of Salmonella Species to Desiccation and Low-Moisture Environments: Extending Our Knowledge of How Bacteria Cope with Low-Moisture Stress*, in *The Microbiological Safety of Low Water Activity Foods and Spices*, J.B. Gurtler, M.P. Doyle, and J.L. Kornacki, Editors. 2014, Springer New York. p. 49-66.
96. Balaji, B., et al., *Timing of Induction of Osmotically Controlled Genes in Salmonella enterica Serovar Typhimurium, Determined with Quantitative Real-Time Reverse Transcription-PCR.* Applied and Environmental Microbiology, 2005. **71**(12): p. 8273-8283.
97. Yildiz, F.H., et al., *Molecular analysis of rugosity in a Vibrio cholerae O1 El Tor phase variant.* Molecular microbiology, 2004. **53**(2): p. 497-515.
98. Govan, J.R. and V. Deretic, *Microbial pathogenesis in cystic fibrosis: mucoid Pseudomonas aeruginosa and Burkholderia cepacia.* Microbiological Reviews, 1996. **60**(3): p. 539-74.
99. Nickel, J.C., et al., *Tobramycin resistance of Pseudomonas aeruginosa cells growing as a biofilm on urinary catheter material.* Antimicrobial Agents and Chemotherapy, 1985. **27**(4): p. 619-624.
100. Curtis Nickel, J., J.A. Downey, and J. William Costerton, *Ultrastructural study of microbiologic colonization of urinary catheters.* Urology, 1989. **34**(5): p. 284-291.
101. Costerton, J.W., P.S. Stewart, and E.P. Greenberg, *Bacterial Biofilms: A Common Cause of Persistent Infections.* Science, 1999. **284**(5418): p. 1318-1322.

102. McClure, C.D. and N.L. Schiller, *Inhibition of Macrophage Phagocytosis by Pseudomonas aeruginosa Rhamnolipids In Vitro and In Vivo.* Current Microbiology, 1996. **33**(2): p. 109-117.
103. Read, R.C., et al., *Effect of Pseudomonas aeruginosa rhamnolipids on mucociliary transport and ciliary beating*. Vol. 72. 1992. 2271-2277.
104. Déziel, E., et al., *rhlA is required for the production of a novel biosurfactant promoting swarming motility in Pseudomonas aeruginosa: 3-(3-hydroxyalkanoyloxy)alkanoic acids (HAAs), the precursors of rhamnolipids.* Microbiology, 2003. **149**(8): p. 2005-2013.
105. Mavrodi, D.V., et al., *Functional Analysis of Genes for Biosynthesis of Pyocyanin and Phenazine-1-Carboxamide from Pseudomonas aeruginosa PAO1*. Journal of Bacteriology, 2001. **183**(21): p. 6454-6465.
106. Lau, G.W., et al., *Pseudomonas aeruginosa Pyocyanin Is Critical for Lung Infection in Mice.* Infection and Immunity, 2004. **72**(7): p. 4275-4278.
107. Muller, M., *Pyocyanin induces oxidative stress in human endothelial cells and modulates the glutathione redox cycle.* Free Radical Biology and Medicine, 2002. **33**(11): p. 1527-1533.
108. Britigan, B.E., M.A. Railsback, and C.D. Cox, *The Pseudomonas aeruginosa Secretory Product Pyocyanin Inactivates α(1) Protease Inhibitor: Implications for the Pathogenesis of Cystic Fibrosis Lung Disease.* Infection and Immunity, 1999. **67**(3): p. 1207-1212.
109. Schnare, M., M. Röllinghoff, and S. Qureshi, *Toll-Like Receptors: Sentinels of Host Defence against Bacterial Infection.* International Archives of Allergy and Immunology, 2006. **139**(1): p. 75-85.
110. Fitzgerald, K.A. and Z.J. Chen, *Sorting out Toll Signals.* Cell. **125**(5): p. 834-836.
111. Yamamoto, Y., et al., *Septic Shock Is Associated with Receptor for Advanced Glycation End Products Ligation of LPS.* The Journal of Immunology, 2011. **186**(5): p. 3248-3257.
112. MacIntyre, S., R. Lucken, and P. Owen, *Smooth lipopolysaccharide is the major protective antigen for mice in the surface extract from IATS serotype 6 contributing to the polyvalent Pseudomonas aeruginosa vaccine PEV.* Infection and Immunity, 1986. **52**(1): p. 76-84.
113. Priebe, G.P., et al., *Construction and Characterization of a Live, Attenuated aroA Deletion Mutant of Pseudomonas aeruginosa as a Candidate Intranasal Vaccine.* Infection and Immunity, 2002. **70**(3): p. 1507-1517.
114. Priebe, G.P., et al., *Protection against Fatal Pseudomonas aeruginosa Pneumonia in Mice after Nasal Immunization with a Live, Attenuated aroA Deletion Mutant.* Infection and Immunity, 2003. **71**(3): p. 1453-1461.
115. De Kievit, T.R., *Quorum sensing in Pseudomonas aeruginosa biofilms.* Environmental Microbiology, 2009. **11**(2): p. 279-288.
116. Glessner, A., et al., *Roles of Pseudomonas aeruginosa las and rhl Quorum-Sensing Systems in Control of Twitching Motility.* Journal of Bacteriology, 1999. **181**(5): p. 1623-1629.
117. Hentzer, M., et al., *Attenuation of Pseudomonas aeruginosa virulence by quorum sensing inhibitors.* The EMBO Journal, 2003. **22**(15): p. 3803-3815.
118. Pesci, E.C., et al., *Quinolone signaling in the cell-to-cell communication system of Pseudomonas aeruginosa.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(20): p. 11229-11234.

119. Mashburn, L.M. and M. Whiteley, *Membrane vesicles traffic signals and facilitate group activities in a prokaryote.* Nature, 2005. **437**(7057): p. 422-425.
120. McBroom, A.J. and M.J. Kuehn, *Release of outer membrane vesicles by Gram-negative bacteria is a novel envelope stress response.* Molecular Microbiology, 2007. **63**(2): p. 545-558.
121. Schalk, I.J. and L. Guillon, *Pyoverdine biosynthesis and secretion in Pseudomonas aeruginosa: implications for metal homeostasis.* Environmental Microbiology, 2013. **15**(6): p. 1661-1673.
122. Ankenbauer, R., S. Sriyosachati, and C.D. Cox, *Effects of siderophores on the growth of Pseudomonas aeruginosa in human serum and transferrin.* Infection and Immunity, 1985. **49**(1): p. 132-140.
123. Martin, L., et al., *Pseudomonas siderophores in the sputum of patients with cystic fibrosis.* BioMetals, 2011. **24**(6): p. 1059-1067.
124. Meyer, J.M., et al., *Pyoverdin is essential for virulence of Pseudomonas aeruginosa.* Infection and Immunity, 1996. **64**(2): p. 518-523.
125. Lamont, I.L., et al., *Siderophore-mediated signaling regulates virulence factor production in Pseudomonas aeruginosa.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(10): p. 7072-7077.
126. Budzikiewicz, H., *Siderophore-Antibiotic Conjugates Used as Trojan Horses Against Pseudomonas aeruginosa.* Current Topics in Medicinal Chemistry, 2001. **1**(1): p. 73-82.
127. Mislin, G.L.A. and I.J. Schalk, *Siderophore-dependent iron uptake systems as gates for antibiotic Trojan horse strategies against Pseudomonas aeruginosa.* Metallomics, 2014. **6**(3): p. 408-420.
128. Ottow, J.C.G., *Ecology, Physiology, and Genetics of Fimbriae and Pili.* Annual Review of Microbiology, 1975. **29**(1): p. 79-108.
129. Skerker, J.M. and H.C. Berg, *Direct observation of extension and retraction of type IV pili.* Proceedings of the National Academy of Sciences, 2001. **98**(12): p. 6901-6904.
130. Aldsworth, T.G., R.L. Sharman, and C.E.R. Dodd, *Bacterial suicide through stress.* Cellular and Molecular Life Sciences CMLS, 1999. **56**(5-6): p. 378-383.
131. Aertsen, A., et al., *Induction of Oxidative Stress by High Hydrostatic Pressure in Escherichia coli.* Applied and Environmental Microbiology, 2005. **71**(5): p. 2226-2231.
132. Lewis, K., *Programmed Death in Bacteria.* Microbiology and Molecular Biology Reviews, 2000. **64**(3): p. 503-514.
133. Häussler, S. and T. Becker, *The Pseudomonas Quinolone Signal (PQS) Balances Life and Death in <italic>Pseudomonas aeruginosa</italic> Populations.* PLoS Pathog, 2008. **4**(9): p. e1000166.
134. Tomasz, A., A. Albino, and E.V.E. Zanati, *Multiple Antibiotic Resistance in a Bacterium with Suppressed Autolytic System.* Nature, 1970. **227**(5254): p. 138-140.
135. Huber, W.W., B. Grasl-kraupp, and R. Schulte-hermann, *Hepatocarcinogenic Potential of Di(2-Ethylhexyl)phthalate in Rodents and its Implications on Human Risk.* Critical Reviews in Toxicology, 1996. **26**(4): p. 365-481.
136. O'Toole, G.A. and R. Kolter, *Flagellar and twitching motility are necessary for Pseudomonas aeruginosa biofilm development.* Molecular Microbiology, 1998. **30**(2): p. 295-304.
137. Cude, W.N., et al., *Production of the Antimicrobial Secondary Metabolite Indigoidine Contributes to Competitive Surface Colonization by the Marine Roseobacter*

Phaeobacter sp. Strain Y4I. Applied and Environmental Microbiology, 2012. **78**(14): p. 4771-4780.
138. Newman, D.J., G.M. Cragg, and K.M. Snader, *Natural Products as Sources of New Drugs over the Period 1981−2002.* Journal of Natural Products, 2003. **66**(7): p. 1022-1037.
139. Cragg, G.M. and D.J. Newman, *Natural products: A continuing source of novel drug leads.* Biochimica et Biophysica Acta (BBA) - General Subjects, 2013. **1830**(6): p. 3670-3695.
140. de Lima Procópio, R.E., et al., *Antibiotics produced by Streptomyces.* The Brazilian Journal of Infectious Diseases, 2012. **16**(5): p. 466-471.
141. Watve, M.G., et al., *How many antibiotics are produced by the genus Streptomyces?* Arch Microbiol, 2001. **176**(5): p. 386-90.
142. Amann, R.I., W. Ludwig, and K.H. Schleifer, *Phylogenetic identification and in situ detection of individual microbial cells without cultivation.* Microbiological Reviews, 1995. **59**(1): p. 143-69.
143. Bentley, S.D., et al., *Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2).* Nature, 2002. **417**(6885): p. 141-147.
144. Miguélez, E.M., C. Hardisson, and M.B. Manzanal, *Streptomycetes: a new model to study cell death.* International microbiology : the official journal of the Spanish Society for Microbiology, 2000. **3**(3): p. 153-158.
145. Chater, K.F., *Taking a genetic scalpel to the Streptomyces colony.* Microbiology, 1998. **144**(6): p. 1465-1478.
146. Lin, Y.-S., et al., *The chromosomal DNA of Streptomyces lividans 66 is linear.* Molecular Microbiology, 1993. **10**(5): p. 923-933.
147. Chen, C.W., et al., *Once the circle has been broken: dynamics and evolution of Streptomyces chromosomes.* Trends in Genetics, 2002. **18**(10): p. 522-529.
148. Güneş, G., B. Smith, and P. Dyson, *Genetic instability associated with insertion of IS6100 into one end of the Streptomyces lividans chromosome.* Microbiology, 1999. **145**(9): p. 2203-2208.
149. Volff, J.N. and J. Altenbuchner, *High-frequency transposition of IS1373, the insertion sequence delimiting the amplifiable element AUD2 of Streptomyces lividans.* Journal of Bacteriology, 1997. **179**(17): p. 5639-42.
150. Lin, Y.-S. and C.W. Chen, *Instability of artificially circularized chromosomes of Streptomyces lividans.* Molecular Microbiology, 1997. **26**(4): p. 709-719.
151. Eichenseer, C. and J. Altenbuchner, *The very large amplifiable element AUD2 from Streptomyces lividans 66 has insertion sequence-like repeats at its ends.* Journal of Bacteriology, 1994. **176**(22): p. 7107-7112.
152. Chater, K.F. and L.C. Wilde, *Restriction of a bacteriophage of Streptomyces albus G involving endonuclease SalI.* Journal of Bacteriology, 1976. **128**(2): p. 644-650.
153. Komatsu, M., et al., *Engineered Streptomyces avermitilis Host for Heterologous Expression of Biosynthetic Gene Cluster for Secondary Metabolites.* ACS Synthetic Biology, 2013. **2**(7): p. 384-396.
154. Gomez-Escribano, J.P. and M.J. Bibb, *Engineering Streptomyces coelicolor for heterologous expression of secondary metabolite gene clusters.* Microbial Biotechnology, 2011. **4**(2): p. 207-215.

155. Komatsu, M., et al., *Genome-minimized Streptomyces host for the heterologous expression of secondary metabolism.* Proceedings of the National Academy of Sciences, 2010. **107**(6): p. 2646-2651.
156. Hopwood, D.A. and D.H. Sherman, *Molecular Genetics of Polyketides and its Comparison to Fatty Acid Biosynthesis.* Annual Review of Genetics, 1990. **24**(1): p. 37-62.
157. Donadio, S., et al., *Modular organization of genes required for complex polyketide biosynthesis.* Science, 1991. **252**(5006): p. 675-679.
158. Cortes, J., et al., *An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea.* Nature, 1990. **348**(6297): p. 176-178.
159. Strieker, M., A. Tanović, and M.A. Marahiel, *Nonribosomal peptide synthetases: structures and dynamics.* Current Opinion in Structural Biology, 2010. **20**(2): p. 234-240.
160. Nakano, M.M., M.A. Marahiel, and P. Zuber, *Identification of a genetic locus required for biosynthesis of the lipopeptide antibiotic surfactin in Bacillus subtilis.* Journal of Bacteriology, 1988. **170**(12): p. 5662-5668.
161. Cosmina, P., et al., *Sequence and analysis of the genetic locus responsible for surfactin synthesis in Bacillus subtilis.* Molecular Microbiology, 1993. **8**(5): p. 821-831.
162. Arnison, P.G., et al., *Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature.* Natural Product Reports, 2013. **30**(1): p. 108-160.
163. Willey, J.M. and W.A. van der Donk, *Lantibiotics: Peptides of Diverse Structure and Function.* Annual Review of Microbiology, 2007. **61**(1): p. 477-501.
164. Claesen, J. and M. Bibb, *Genome mining and genetic analysis of cypemycin biosynthesis reveal an unusual class of posttranslationally modified peptides.* Proceedings of the National Academy of Sciences, 2010. **107**(37): p. 16297-16302.
165. Just-Baringo, X., F. Albericio, and M. Álvarez, *Thiopeptide Antibiotics: Retrospective and Recent Advances.* Marine Drugs, 2014. **12**(1): p. 317-351.
166. Duquesne, S., et al., *Structural and Functional Diversity of Microcins, Gene-Encoded Antibacterial Peptides from Enterobacteria.* Journal of Molecular Microbiology and Biotechnology, 2007. **13**(4): p. 200-209.
167. Hallen, H.E., et al., *Gene family encoding the major toxins of lethal Amanita mushrooms.* Proceedings of the National Academy of Sciences, 2007. **104**(48): p. 19097-19101.
168. Gillon, A.D., et al., *Biosynthesis of circular proteins in plants.* The Plant Journal, 2008. **53**(3): p. 505-515.
169. Kaas, Q., J.-C. Westermann, and D.J. Craik, *Conopeptide characterization and classifications: An analysis using ConoServer.* Toxicon, 2010. **55**(8): p. 1491-1509.
170. Sivonen, K., et al., *Cyanobactins—ribosomal cyclic peptides produced by cyanobacteria.* Applied Microbiology and Biotechnology, 2010. **86**(5): p. 1213-1225.
171. Shih, P.M., et al., *Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing.* Proceedings of the National Academy of Sciences, 2013. **110**(3): p. 1053-1058.
172. Houssen, W.E. and M. Jaspars, *Azole-Based Cyclic Peptides from the Sea Squirt Lissoclinum Patella: Old Scaffolds, New Avenues.* ChemBioChem, 2010. **11**(13): p. 1803-1815.

173. Ishida, K., et al., *Kawaguchipeptin B, an Antibacterial Cyclic Undecapeptide from the Cyanobacterium Microcystis aeruginosa.* Journal of Natural Products, 1997. **60**(7): p. 724-726.
174. Portmann, C., et al., *Aerucyclamides A and B: Isolation and Synthesis of Toxic Ribosomal Heterocyclic Peptides from the Cyanobacterium Microcystis aeruginosa PCC 7806.* Journal of Natural Products, 2008. **71**(7): p. 1193-1196.
175. Schmidt, E.W., et al., *Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(20): p. 7315-7320.
176. Long, P.F., et al., *Shotgun Cloning and Heterologous Expression of the Patellamide Gene Cluster as a Strategy to Achieving Sustained Metabolite Production.* ChemBioChem, 2005. **6**(10): p. 1760-1765.
177. Leikoski, N., et al., *Genome Mining Expands the Chemical Diversity of the Cyanobactin Family to Include Highly Modified Linear Peptides.* Chemistry & Biology, 2013. **20**(8): p. 1033-1043.
178. McIntosh, J.A. and E.W. Schmidt, *Marine Molecular Machines: Heterocyclization in Cyanobactin Biosynthesis.* ChemBioChem, 2010. **11**(10): p. 1413-1421.
179. Koehnke, J., et al., *The structural biology of patellamide biosynthesis.* Current Opinion in Structural Biology, 2014. **29**(0): p. 112-121.
180. McIntosh, J.A., et al., *Circular Logic: Nonribosomal Peptide-like Macrocyclization with a Ribosomal Peptide Catalyst.* Journal of the American Chemical Society, 2010. **132**(44): p. 15499-15501.
181. Donia, Mohamed S. and Eric W. Schmidt, *Linking Chemistry and Genetics in the Growing Cyanobactin Natural Products Family.* Chemistry & Biology, 2011. **18**(4): p. 508-519.
182. Lee, J., et al., *Using Marine Natural Products to Discover a Protease that Catalyzes Peptide Macrocyclization of Diverse Substrates.* Journal of the American Chemical Society, 2009. **131**(6): p. 2122-2124.
183. Sardar, D., et al., *Recognition Sequences and Substrate Evolution in Cyanobactin Biosynthesis.* ACS Synthetic Biology, 2014.
184. Donia, M.S., J. Ravel, and E.W. Schmidt, *A global assembly line to cyanobactins.* Nature chemical biology, 2008. **4**(6): p. 341-343.
185. Donia, M.S., et al., *Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians.* Nat Chem Biol, 2006. **2**(12): p. 729-735.
186. Dunbar, K.L. and D.A. Mitchell, *Insights into the Mechanism of Peptide Cyclodehydrations Achieved through the Chemoenzymatic Generation of Amide Derivatives.* Journal of the American Chemical Society, 2013. **135**(23): p. 8692-8701.
187. Agarwal, V., et al., *Structures of Cyanobactin Maturation Enzymes Define a Family of Transamidating Proteases.* Chemistry & Biology, 2012. **19**(11): p. 1411-1422.
188. Koehnke, J., et al., *The mechanism of patellamide macrocyclization revealed by the characterization of the PatG macrocyclase domain.* Nat Struct Mol Biol, 2012. **19**(8): p. 767-772.
189. Tianero, M.D.B., et al., *Ribosomal Route to Small-Molecule Diversity.* Journal of the American Chemical Society, 2011. **134**(1): p. 418-425.

190. Hansson, L.-A., et al., *Cyanobacterial chemical warfare affects zooplankton community composition.* Freshwater Biology, 2007. **52**(7): p. 1290-1301.
191. Herdman, M., et al., *Genome Size of Cyanobacteria.* Journal of General Microbiology, 1979. **111**(1): p. 73-85.
192. Koksharova, O. and C. Wolk, *Genetic tools for cyanobacteria.* Applied Microbiology and Biotechnology, 2002. **58**(2): p. 123-137.
193. Jones, A.C., et al., *Evaluation of Streptomyces coelicolor A3(2) as a heterologous expression host for the cyanobacterial protein kinase C activator lyngbyatoxin A.* FEBS Journal, 2012. **279**(7): p. 1243-1251.
194. Kim, E.J., et al., *Heterologous Production of 4-O-Demethylbarbamide, a Marine Cyanobacterial Natural Product.* Organic Letters, 2012. **14**(23): p. 5824-5827.
195. Ongley, S.E., et al., *High-Titer Heterologous Production in E. coli of Lyngbyatoxin, a Protein Kinase C Activator from an Uncultured Marine Cyanobacterium.* ACS Chemical Biology, 2013. **8**(9): p. 1888-1893.
196. Leikoski, N., et al., *Analysis of an Inactive Cyanobactin Biosynthetic Gene Cluster Leads to Discovery of New Natural Products from Strains of the Genus <italic>Microcystis</italic>.* PLoS ONE, 2012. **7**(8): p. e43002.
197. Ziemert, N., et al., *Microcyclamide Biosynthesis in Two Strains of Microcystis aeruginosa: from Structure to Genes and Vice Versa.* Applied and Environmental Microbiology, 2008. **74**(6): p. 1791-1797.
198. Ziemert, N., et al., *Exploiting the Natural Diversity of Microviridin Gene Clusters for Discovery of Novel Tricyclic Depsipeptides.* Applied and Environmental Microbiology, 2010. **76**(11): p. 3568-3574.
199. Leikoski, N., et al., *Highly Diverse Cyanobactins in Strains of the Genus Anabaena.* Applied and Environmental Microbiology, 2010. **76**(3): p. 701-709.
200. Oldenburg, K.R., et al., *Recombination-mediated PCR-directed plasmid construction in vivo in yeast.* Nucleic Acids Research, 1997. **25**(2): p. 451-452.
201. Mallet, L. and M. Jacquet, *Intergenic flip flop, a method for systematic gene disruption and cloning in yeast.* Yeast, 1996. **12**(13): p. 1351-1357.
202. DeMarini, D.J., et al., *Oligonucleotide-mediated, PCR-independent cloning by homologous recombination.* BioTechniques, 2001. **30**(3): p. 520-523.
203. Shanks, R.M.Q., et al., *Saccharomyces cerevisiae-Based Molecular Tool Kit for Manipulation of Genes from Gram-Negative Bacteria.* Applied and Environmental Microbiology, 2006. **72**(7): p. 5027-5036.
204. Tomitani, A., et al., *The evolutionary diversification of cyanobacteria: Molecular–phylogenetic and paleontological perspectives.* Proceedings of the National Academy of Sciences, 2006. **103**(14): p. 5442-5447.
205. Golden, J.W., S.J. Robinson, and R. Haselkorn, *Rearrangement of nitrogen fixation genes during heterocyst differentiation in the cyanobacterium Anabaena.* Nature, 1985. **314**(6010): p. 419-423.
206. Henson, B.J., et al., *Molecular phylogeny of the heterocystous cyanobacteria (subsections IV and V) based on nifD.* International Journal of Systematic and Evolutionary Microbiology, 2004. **54**(2): p. 493-497.
207. Lawton, L.A., L.A. Morris, and M. Jaspars, *A Bioactive Modified Peptide, Aeruginosamide, Isolated from the Cyanobacterium Microcystis aeruginosa.* The Journal of Organic Chemistry, 1999. **64**(14): p. 5329-5332.

208. Ishida, K., et al., *Kawaguchipeptin A, a novel cyclic undecapeptide from cyanobacterium Microcystis aeruginosa (NIES-88).* Tetrahedron, 1996. **52**(27): p. 9025-9030.
209. Ishida, K., H. Nakagawa, and M. Murakami, *Microcyclamide, a Cytotoxic Cyclic Hexapeptide from the Cyanobacterium Microcystis aeruginosa.* Journal of Natural Products, 2000. **63**(9): p. 1315-1317.
210. Christiansen, G., et al., *Genetic Variation of Adenylation Domains of the Anabaenopeptin Synthesis Operon and Evolution of Substrate Promiscuity.* Journal of Bacteriology, 2011. **193**(15): p. 3822-3831.
211. Fewer, D.P., et al., *The non-ribosomal assembly and frequent occurrence of the protease inhibitors spumigins in the bloom-forming cyanobacterium Nodularia spumigena.* Molecular Microbiology, 2009. **73**(5): p. 924-937.
212. Stratmann, K., et al., *Hapalosin, a Cyanobacterial Cyclic Depsipeptide with Multidrug-Resistance Reversing Activity.* The Journal of Organic Chemistry, 1994. **59**(24): p. 7219-7226.
213. Martin, V.J.J., et al., *Engineering a mevalonate pathway in Escherichia coli for production of terpenoids.* Nat Biotech, 2003. **21**(7): p. 796-802.
214. Rouhiainen, L., et al., *Two Alternative Starter Modules for the Non-Ribosomal Biosynthesis of Specific Anabaenopeptin Variants in Anabaena (Cyanobacteria).* Chemistry & Biology, 2010. **17**(3): p. 265-273.
215. Moore, R.E., et al., *Hapalindoles, antibacterial and antimycotic alkaloids from the cyanophyte Hapalosiphon fontinalis.* The Journal of Organic Chemistry, 1987. **52**(6): p. 1036-1043.
216. Gross, E.M., C.P. Wolk, and F. Jüttner, *FISCHERELLIN, A NEW ALLELOCHEMICAL FROM THE FRESHWATER CYANOBACTERIUM FISCHERELLA MUSCICOLA1*. Journal of Phycology, 1991. **27**(6): p. 686-692.
217. Falch, B.S., et al., *Ambigol A and B: new biologically active polychlorinated aromatic compounds from the terrestrial blue-green alga Fischerella ambigua.* The Journal of Organic Chemistry, 1993. **58**(24): p. 6570-6575.
218. Hillwig, M.L. and X. Liu, *A new family of iron-dependent halogenases acts on freestanding substrates.* Nat Chem Biol, 2014. **10**(11): p. 921-923.
219. Hillwig, M.L., et al., *Identification and Characterization of a Welwitindolinone Alkaloid Biosynthetic Gene Cluster in the Stigonematalean Cyanobacterium Hapalosiphon welwitschii*. ChemBioChem, 2014. **15**(5): p. 665-669.
220. Hillwig, M.L., Q. Zhu, and X. Liu, *Biosynthesis of Ambiguine Indole Alkaloids in Cyanobacterium Fischerella ambigua.* ACS Chemical Biology, 2014. **9**(2): p. 372-377.
221. Donia, M.S., J. Ravel, and E.W. Schmidt, *A global assembly line for cyanobactins.* Nat Chem Biol, 2008. **4**(6): p. 341-343.
222. Rausch, C., et al., *Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs).* Nucleic Acids Research, 2005. **33**(18): p. 5799-5808.
223. Magarvey, N.A., M. Ehling-Schulz, and C.T. Walsh, *Characterization of the Cereulide NRPS α-Hydroxy Acid Specifying Modules: Activation of α-Keto Acids and Chiral Reduction on the Assembly Line.* Journal of the American Chemical Society, 2006. **128**(33): p. 10698-10699.
224. Ma, H., et al., *Plasmid construction by homologous recombination in yeast.* Gene, 1987. **58**(2–3): p. 201-216.

225. Shen, Q., D.D. Lantvit, J., Zi, J. Orjala, S.M. Swanson. *The Novel Depsipeptide Hapalosin B is Potent Inhibitor of Anticancer Drug Resistance Associated with ABCB1 or ABCC1 Overexpression*. in *52nd Annual Meeting of the American Society of Pharmacognosy*. 2011. San Diego.
226. Cronan, J.E., X. Zhao, and Y. Jiang, *Function, Attachment and Synthesis of Lipoic Acid in Escherichia coli*, in *Advances in Microbial Physiology*, K.P. Robert, Editor 2005, Academic Press. p. 103-146.
227. Morris, T.W., K.E. Reed, and J.E. Cronan, *Lipoic acid metabolism in Escherichia coli: the lplA and lipB genes define redundant pathways for ligation of lipoyl groups to apoprotein.* Journal of Bacteriology, 1995. **177**(1): p. 1-10.
228. Belshaw, P.J., C.T. Walsh, and T. Stachelhaus, *Aminoacyl-CoAs as Probes of Condensation Domain Selectivity in Nonribosomal Peptide Synthesis.* Science, 1999. **284**(5413): p. 486-489.
229. Khosla, C., et al., *TOLERANCE AND SPECIFICITY OF POLYKETIDE SYNTHASES.* Annual Review of Biochemistry, 1999. **68**(1): p. 219-253.
230. Oursel, D., et al., *Identification and relative quantification of fatty acids in Escherichia coli membranes by gas chromatography/mass spectrometry*. Rapid Communications in Mass Spectrometry, 2007. **21**(20): p. 3229-3233.
231. Thiericke, R. and J. Rohr, *Biological variation of microbial metabolites by precursor-directed biosynthesis.* Natural Product Reports, 1993. **10**(3): p. 265-289.
232. Sundermann, U., et al., *Enzyme-Directed Mutasynthesis: A Combined Experimental and Theoretical Approach to Substrate Recognition of a Polyketide Synthase.* ACS Chemical Biology, 2013. **8**(2): p. 443-450.
233. Kries, H., et al., *Reprogramming Nonribosomal Peptide Synthetases for "Clickable" Amino Acids.* Angewandte Chemie International Edition, 2014. **53**(38): p. 10105-10108.
234. Fernandez-Suarez, M., et al., *Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes.* Nat Biotech, 2007. **25**(12): p. 1483-1487.
235. Bégué, J.-P. and D. Bonnet-Delpon, *Recent advances (1995–2005) in fluorinated pharmaceuticals based on natural products.* Journal of Fluorine Chemistry, 2006. **127**(8): p. 992-1012.
236. Roy, A.D., et al., *Gene Expression Enabling Synthetic Diversification of Natural Products: Chemogenetic Generation of Pacidamycin Analogs.* Journal of the American Chemical Society, 2010. **132**(35): p. 12243-12245.
237. Yanai, K., et al., *Para-position derivatives of fungal anthelmintic cyclodepsipeptides engineered with Streptomyces venezuelae antibiotic biosynthetic genes.* Nat Biotech, 2004. **22**(7): p. 848-855.
238. Feifel, S.C., et al., *In Vitro Synthesis of New Enniatins: Probing the α-D-Hydroxy Carboxylic Acid Binding Pocket of the Multienzyme Enniatin Synthetase.* ChemBioChem, 2007. **8**(15): p. 1767-1770.
239. Stachelhaus, T., H.D. Mootz, and M.A. Marahiel, *The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases.* Chemistry & Biology. **6**(8): p. 493-505.
240. Sivonen, K., et al., *Cyanobactins—ribosomal cyclic peptides produced by cyanobacteria.* Applied Microbiology and Biotechnology, 2010. **86**(5): p. 1213-1225.
241. Portmann, C., et al., *Isolation of Aerucyclamides C and D and Structure Revision of Microcyclamide 7806A: Heterocyclic Ribosomal Peptides from Microcystis aeruginosa*

PCC 7806 and Their Antiparasite Evaluation. Journal of Natural Products, 2008. **71**(11): p. 1891-1896.
242. Schneider, T.L., B. Shen, and C.T. Walsh, *Oxidase Domains in Epothilone and Bleomycin Biosynthesis: Thiazoline to Thiazole Oxidation during Chain Elongation†.* Biochemistry, 2003. **42**(32): p. 9722-9730.
243. Nyulászi, L., P. Várnai, and T. Veszprémi, *About the aromaticity of five-membered heterocycles.* Journal of Molecular Structure: THEOCHEM, 1995. **358**(1–3): p. 55-61.
244. Metcalf, W.W., et al., *Conditionally Replicative and Conjugative Plasmids CarryinglacZα for Cloning, Mutagenesis, and Allele Replacement in Bacteria.* Plasmid, 1996. **35**(1): p. 1-13.
245. Hunger, M., et al., *Analysis and nucleotide sequence of an origin of an origin of DNA replication in Acinetobacter calcoaceticus and its use for Escherichia coli shuttle plasmids.* Gene, 1990. **87**(1): p. 45-51.
246. Aranda, J., et al., *Acinetobacter baumannii RecA Protein in Repair of DNA Damage, Antimicrobial Resistance, General Stress Response, and Virulence.* Journal of Bacteriology, 2011. **193**(15): p. 3740-3747.
247. Evans, B.S., et al., *Discovery of the Antibiotic Phosacetamycin via a New Mass Spectrometry-Based Method for Phosphonic Acid Detection.* ACS Chemical Biology, 2013. **8**(5): p. 908-913.
248. Huddleston, M.J., et al., *Selective detection of phosphopeptides in complex mixtures by electrospray liquid chromatography/mass spectrometry.* Journal of the American Society for Mass Spectrometry, 1993. **4**(9): p. 710-717.
249. Wilson, D.J. and C.C. Aldrich, *A Continuous Kinetic Assay for Adenylation Enzyme Activity and Inhibition.* Analytical biochemistry, 2010. **404**(1): p. 56-63.