# EXPLORING MICROBIAL SECONDARY METABOLITES THROUGH GENOME MINING

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

Master of Science

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

Kenneth P. Dietrich School of Arts and Sciences

This thesis was presented

by

Daniel H. Kwak

It was defended on

November 5$^{th}$, 2013

and approved by

Dr. Stephen Weber, Professor, Department of Chemistry

Dr. Alexander Deiters, Professor, Department of Chemistry

Thesis Director: Dr. Xinyu Liu, Assistant Professor, Department of Chemistry

**EXPLORING MICROBIAL SECONDARY METABOLITES**
**THROUGH GENOME MINING**

Daniel H. Kwak, M.S.

University of Pittsburgh, 2015

Microbial secondary metabolites are physiologically significant, exhibiting auxiliary functions for the producer and as scaffolds in the developments of new medicines. Advancements in genome sequencing technologies have enabled researchers to access unprecedented amounts of genomic data that can be used to discover the enzymatic machinery necessary to discover novel and biologically-active molecules. This approach has been termed "genome mining." In both of the investigations presented herein, genome mining was utilized to discover and characterize biosynthetic pathways of novel molecules. The findings in one study utilize this approach to discover a small molecule virulence factor from the opportunistic human pathogen *Acinetobacter baumannii.* This virulence factor has been found to be associated with a number of clinically significant phenotypes, and these findings suggest that this can be a target in the developments of next generation antibiotics. In another study, this approach was implemented to discover and characterize the biosynthetic pathway of anticancer compound hapalosin from the cyanobacterial species *Hapalosiphon welwitschii*. Cloning and expression of this biosynthetic pathway in the surrogate host *Escherichia coli* enabled its genetic characterization as well as the generation of a small combinatorial library consisting of analogs incorporating natural and unnatural substrates. Collectively, these investigations demonstrate the utility of genome mining to characterize novel molecules important in pathogenesis or in the biosynthesis of clinically-significant compounds.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# 1.0    INTRODUCTION

## 1.1    SIGNIFICANCE OF MICROBIAL SECONDARY METABOLITES

Although their introduction into the clinic was attributable to their compelling antibiotic activities, over the decades, microbial secondary metabolites have encompassed many other far reaching and vital roles as medicines[1] (**Figure 1.1.1**). Notably, these secondary metabolites have had important implications as medicines with antibacterial, anti-fungal, anti-parasitic, anti-viral, anti-flammatory, and anticancer properties.[2] In some cases, the discovery and implementation of a single secondary metabolite could be credited for the success of an entire field of medicine. For example, the immunosuppressant drug cyclosporine A from the fungus *Tolypocladium inflatum* had enabled unprecedented organ transplantation procedures at the time of its initial clinical use.[3]
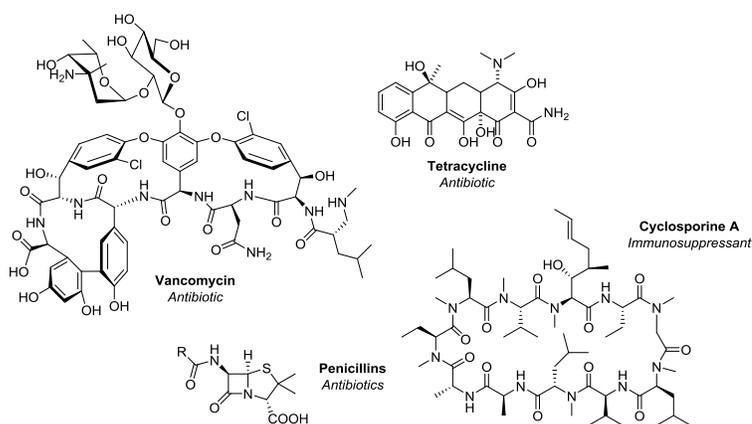
**Figure 1.1.1** Examples of microbial secondary metabolites that served as important medicines

Although the influence of microbial secondary metabolites as natural products is well established, from its prevalence as antibiotics, antiviral and immunosuppressant drugs, the intended function of these molecules by the native producing microorganism, has been much more elusive.[4] It has become increasingly clear that understanding the roles of microbial secondary metabolites outside of their medicinal value can provide insight into the development of novel therapeutics targeted at virulence factors. For example, bacteria participate in population-dependent behaviors known as quorum sensing. A class of quorum sensing molecules known as acyl-homoserine lactones have become important targets for next-generation antibiotics due to the virulence phenotypes that they often control.[5] Other physiologically significant small molecules utilized by bacteria include siderophores, which are necessary for the acquisition of iron, an important element for bacterial cellular function. Both acyl homoserine lactone and siderophore have been shown critical for the virulence of its native bacterial producer (**Figure 1.1.2**).[6] Therefore, disrupting the function of these small molecules has become a promising approach toward next generation antibiotics.
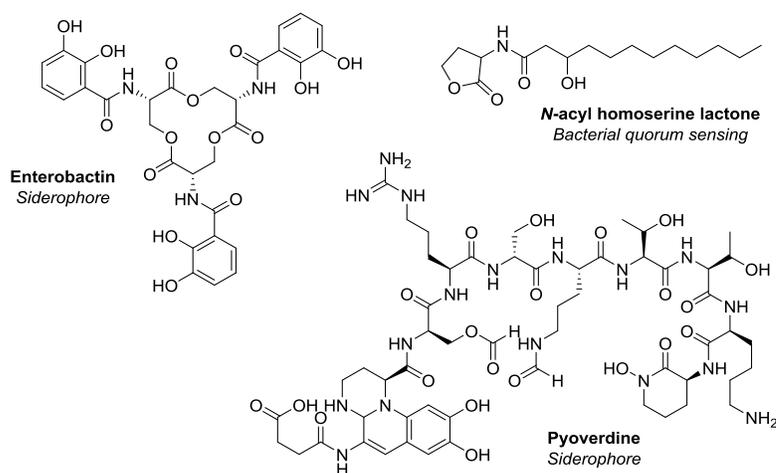


**Figure 1.1.2** Microbial secondary metabolites that have been shown to influence virulence of its bacterial producers.

## 1.2 MICROBIAL SECONDARY METABOLITE BIOSYNTHESIS

Of all the microbial natural products, which are represented from several major classes of molecules, including terpenoids and oligosaccharides, 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 1.2.1**). 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.[7]

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 1.2.2**).[8],[9] A thioesterase (TE) domain at the C-terminus of DEBS facilitates cyclization and chain release, a common feature of both PKS and NRPS assemblies.
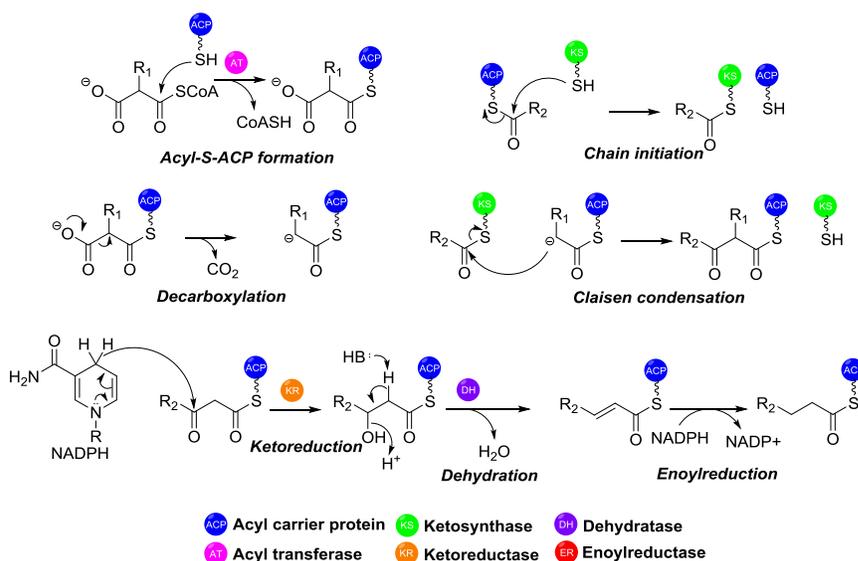


**Figure 1.2.1** Key enzymatic reactions in the biosynthesis of PK products.
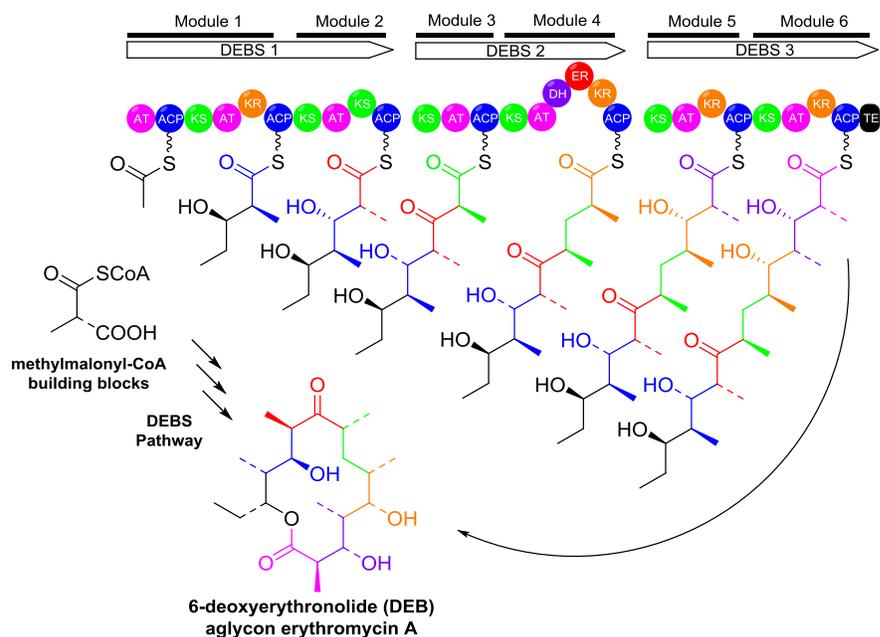
**Figure 1.2.2** The DEBS pathway is an example of how Nature uses a single ketide building block to generate a complex structure.

NRP assembly can be likened to ribosomal peptide biosynthesis in which amino acid monomers are activated as mixed esters and proceed condensation to form amide bonds between individual units (**Figure 1.2.3**). Specifically, amino acid monomers are activated to form the adenylyl-amino acid mixed ester substrate 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 assemblage of various amino acids to produce a final cyclic lipopeptide (**Figure 1.2.4**).[10],[11]
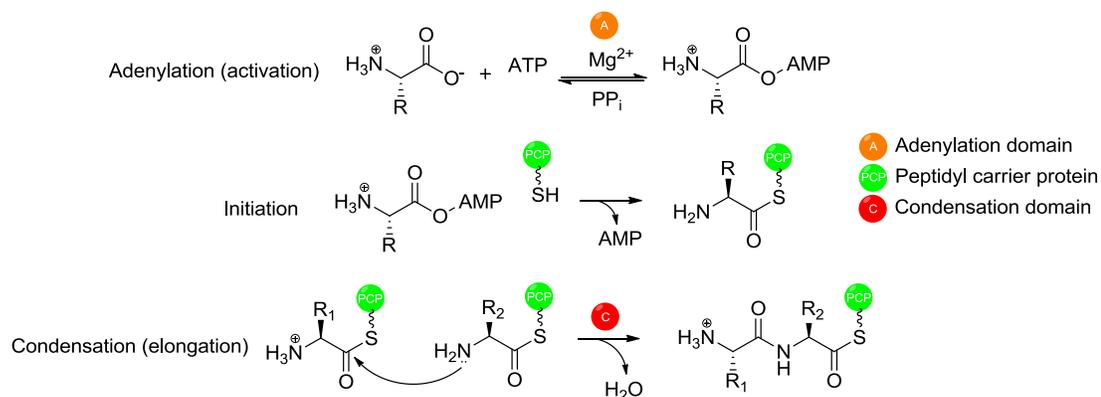
**Figure 1.2.3** Key reactions and their domains required for the formation of amide bonds and assembly of NRPs.
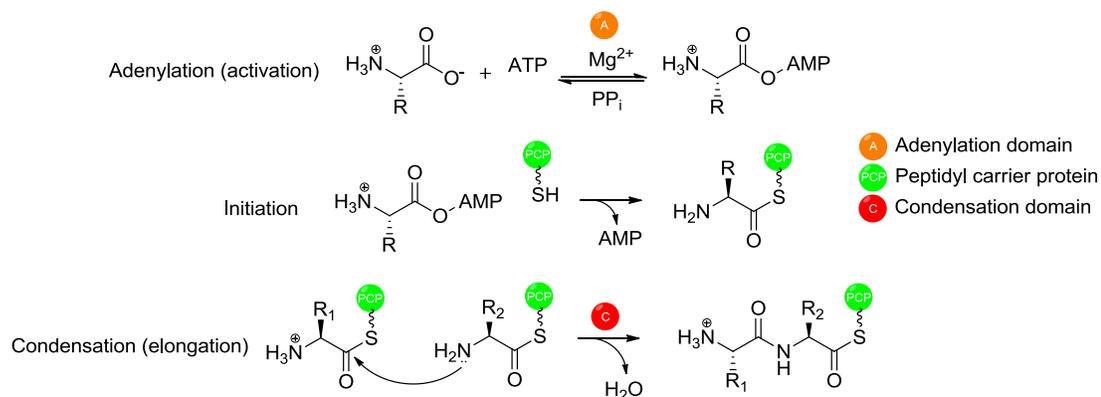


**Figure 1.2.4** The biosynthesis of cyclic lipopeptide surfactin serves as a representative NRPS assembly line.

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

6

nucleophilic attack from a side chain hydroxyl moiety to the thioester tether, both mediated by the TE domain.

## 1.3    GENOMICS-GUIDED NATURAL PRODUCTS DISCOVERY

The availability of rapid and low-cost genome sequencing technologies[12] has benefited no field of research more than microbiology. Analysis of genes and biosynthetic pathways have yielded insight into bacterial pathogenesis, secondary metabolites biosynthesis, evolutionary processes, and novel antibiotic drug targets.[13] Moreover, although the knowledge of natural products of microorganisms have been known for several decades, recent advancements in genome sequencing technology have invoked even more compelling questions about their origin and function, both for humans as well as microbes. The discoveries of potent antibiotics such as penicillin and streptomycin drew attention to the obvious medicinal value of these types of compounds – including others such as anti-tumor, antibiotic, and anti-viral properties – but shied away from questions of their natural functions to microbes.[14]

Recent years have therefore seen a paradigm shift from bioactivity-based discoveries to genome-guided ones with the latter approach enabling access to traditionally cryptic molecules of silent gene clusters. Specifically, these cryptic secondary metabolites are expressed and function only under specific environmental cues unlike those found in the laboratory culture setting.[14] For example, genome mining of the soil bacteria *Streptomyces coelicolor* revealed its potential to biosynthesize approximately three dozen compounds[15], while in fact only five had been isolated and characterized in the laboratory[16]. It is hypothesized that this example is representative of many microbes and that the genome mining approach could expand the understanding of the significance of these small molecules to both bacterial and human physiology.

The availability of entire genome sequences has had two major implications in natural products characterization: first, the biosynthetic genes responsible for the production of microbial natural products can be established; second, new molecules can be discovered from the identification of novel biosynthetic pathways. For both cases, heterologous expression of these

pathways is crucial to enable the production and characterization of molecules only expressed under specific conditions or in very minor concentrations.

Heterologous expression involves cloning of the DNA sequences of interest, assembling biosynthetic pathways into expression plasmids, and introduction of the genes and plasmid into a laboratory amenable host microbe such as *Escherichia coli* or *Saccharomyces cerevisiae*. Additionally, synthetic regulatory DNA sequences, such as ribosome-binding sites and promoters, can be introduced into the biosynthetic pathway to facilitate expression. Heterologous expression hosts can then be grown quickly and to high cell density, translating to higher yields of compounds naturally found in low concentrations in the environment. For example, under laboratory growth conditions, the doubling time of *E. coli* is approximately 20 minutes, while that of many cyanobacterial strains is approximately one to two days, or approximately 100-fold slower. Therefore, the introduction of cyanobacterial biosynthetic genes into *E. coli* could yield considerably more product, an attractive option for synthetically challenging natural products.

The characterization of the salinosporamide A biosynthetic gene cluster of the marine bacterium *Salinispora tropica* demonstrates how genome sequencing data can provide the basis for the discovery of novel natural compounds with clinical relevance.[17] Salinosporamide A is a proteasome inhibitor and anticancer compound currently under human clinical trials and was first discovered by bioactivity-guided fractionation of *S. tropica* extracts.[18] When the genome sequence was available for this marine species, two copies of a proteasome subunit were discovered adjacent to the gene cluster. Gene duplication can occur within organisms during cell division and the second copy can evolve to confer novel functions. It was hypothesized that the duplicate gene for the proteasome subunit conferred self-resistance *S. tropica*. Indeed, salinosporamide A was shown to have considerably reduced inhibition for this second copy.[19] This example highlights how inspection of the genome could be implemented to discover novel molecules with desirable properties, although in this case, salinosporamide A had already been characterized previously. chain hydroxyl moiety to the thioester tether, both mediated by the TE domain.

## 1.4    GOAL OF THIS WORK

The following two chapters will highlight the characterization of PKS-NRPS gene clusters from two perspectives enabled by bioinformatic analysis of genome sequencing data, also known as the genome mining approach. One investigates an evolutionarily conserved biosynthetic pathway from the human pathogen *Acinetobacter baumannii* and the clinical significance of this gene cluster to the virulence of this gram negative bacteria, although identification of the exact molecule from this gene cluster remains elusive. The other investigates a gene cluster from the cyanobacterial species *Hapalosiphon welwitschii* that is responsible for the production of a cyclic depsi-lipopeptide with multidrug resistance reversing activity in cancer cell lines. This gene cluster was characterized heterologously in *E. coli* and a focused pre-cursor directed library was generated using this gene cluster platform.

# 2.0 CRYPTIC ACINETOBACTER TOXIN: BIOGENESIS, FUNCTION AND STRUCTURE

Multidrug resistant *Acinetobacter baumannii* (MDRAB) has emerged as the model "superbug": a pathogen difficult or impossible to treat with traditional clinical antibiotics.[20], [21] Part of its persistence as a human pathogen can be attributed to its remarkable ability to acquire antibiotic resistance phenotypes and its ability to thrive in hospital settings in a variety of environments for extended periods of time.[22]

Traditional antibiotics function to eliminate pathogens by inhibiting essential processes of bacterial metabolism (cell wall synthesis, protein synthesis, RNA transcription). However, superbugs have adapted by genetic transmissions and mutations to acquire drug resistance. The mechanisms for resistance include exportation of the antibiotic outside of the cell, drug inactivation, and drug target site alteration.[23] Because traditional antibiotics affect essential components of the cell, a substantial selective pressure exists for the consequential emergence of MDR strains.

An increasingly attractive alternative to traditional antibiotics includes the "virulence attenuation approach" in the development of next-generation antibiotics against human pathogens. The cornerstone of this approach rests on targeting genes essential to virulence, not to the overall survival of the organism.[24] Therefore, the identification of virulence-associated phenotypes and the genes that encode for them is key. Some examples of virulence phenotypes include motility (movement across surfaces), hemolysis (destruction of red blood cells), desiccation resistance (ability to survive in dry conditions), and heat shock resistance (ability to survive periods of elevated temperatures).

Important targets for the development of virulence attenuation drugs include small molecule secondary metabolites that enable the survival of pathogens. Two well-established

classes of molecules common to most bacteria include siderophores (necessary for acquisition of the micronutrient iron) and signaling molecules (specifically, quorum-sensing molecules that are responsible for population-dependent coordinated behaviors). Two small molecules from these classes were characterized previously in *A. baumannii*: acinetobactin[25] (siderophore) and *N*-(3-hydroxydodecanoyl)-L-homoserine lactone[26] (quorum sensing molecule). In line with the knowledge of how critical these small molecules can be for pathogen survival, the aim of this work is to identify a physiologically significant small molecule specific to the *A. baumannii* species and to investigate its roles in the persistence of this pathogen.

## 2.1      CRYPTIC ACINETOBACTER TOXIN GENE CLUSTER

Bioinformatic analysis of the genome sequence for the ATCC17978 *A. baumannii* strain revealed a gene cluster predicted to encode for enzymes in the biosynthesis of a hybrid polyketide-nonribosomal peptide (PKS-NRPS) small molecule (**Figure 2.1.1**). Further sequence homology analysis via Basic Local Alignment Search Tool (BLAST) of all available genome sequencing data for strains of *A. baumannii* also revealed the presence of this PKS-NRPS gene cluster, indicating its evolutionarily-conserved origin and suggestive of its importance to the persistence of *A. baumannii*. Furthermore, this gene cluster was absent from other *Acinetobacter* sp. and gram-negative bacteria based on bioinformatic analysis and is therefore specific for *A. baumannii*. Aside from another characterized NRPS siderophore, acinetobactin, this is the only other PKS-NRPS secondary metabolite found in the genome of *A. baumannii*. Its prevalence throughout the species and the dedication of several energy-intensive enzymes in its biosynthesis implies that this secondary metabolite plays an important role in the survival of *A. baumannii*. Due to the cryptic nature of this conserved PKS-NRPS pathway in *A. baumannii* and its potential contribution to *A. baumannii* virulence, we named this pathway as the *cryptic Acinetobacter toxin* (*cat*) operon.

| 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* | Putative efflux pump |
| *catF* | Porin-like protein |
| *catG* | Esterase-lipase |
| *catH* | Phosphopantetheinyl transferase |
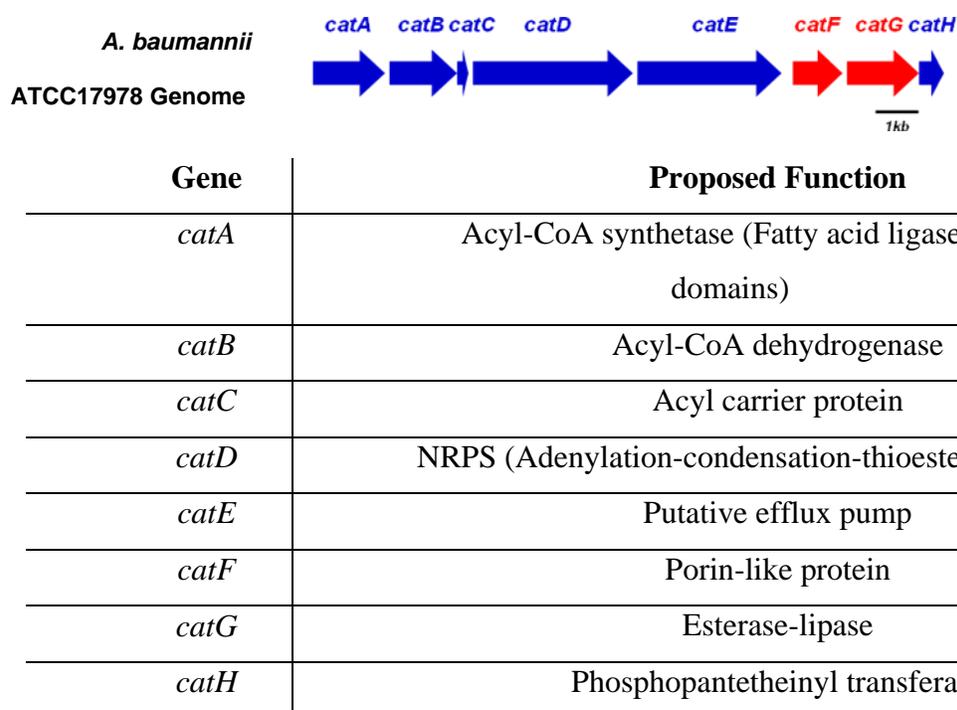
**Figure 2.1.1** cat gene cluster organization from strain ATCC17978 A. baumannii.

The CAT operon consists of eight genes (*cat*A-*H*). Aside from *cat*E and *cat*F that are predicted to encode proteins related to efflux, all six other genes encode proteins related to the PKS and NRPS superfamily. CatA is an acyl-CoA synthetase composed of fatty acid ligase-thiolation domains, responsible for the incorporation of a fatty acid substrate. CatB is an acyl-CoA dehydrogenase, responsible for dehydration. CatC is an acyl carrier protein, which shuttles intermediates from one catalytic domain to the next. CatD is an NRPS enzyme composed of adenylation, condensation and thioesterase domains, which select a specific amino acid, incorporates this into the growing intermediate by condensation, and then releases the product, respectively. CatG is an esterase-lipase enzyme and is predicted to be involved in the hydrolysis of an ester bond. CatH is the phosphopantetheinyl transferase, responsible for posttranslational modification of the acyl carrier protein with 4' phosphopantetheine prosthetic moiety.

## 2.2 GENERATION OF CAT PATHWAY MUTANTS

To understand the role of the CAT pathway, *A. baumannii* mutants of this pathway were generated. We first resorted to a PCR-based approach gene disruption technique, reported by Aranda *et al.* [27], to replace a portion of *catA* gene with a kanamycin resistance marker. We were initially attracted to this approach due to its operational simplicity and envision it might constitute a general approach for us to systematically disrupt genes in the *cat* operon. Briefly, a linear, chimeric DNA fragment was generated by overlap-extension polymerase chain reaction (OE-PCR). This fragment contained 5' and 3' DNA sequences homologous to the flanking regions of the *catA* genes with internal sequences for the kanamycin resistance cassette (**Figure 2.2.1**). Following introduction of this designed DNA molecule to *A. baumannii*, select kanamycin-resistant colonies were further validated by DNA sequencing of the locus of interest.
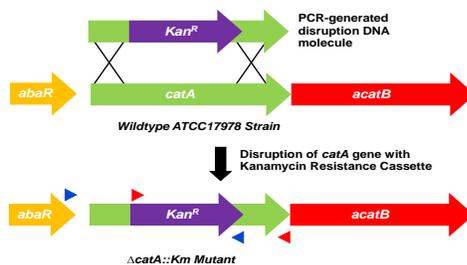


**Figure 2.2.1** Disruption mutant ΔcatA::Km generated by introduction of kanamycin resistance cassette to replace internal sequences of catA gene.

In addition to the PCR-based gene disruption technique, we also explored a suicide plasmid vector to generate an in-frame deletion. The suicide vector was designed from plasmid pMQ30 with DNA sequences flanking the gene of interest, facilitating its incorporation into the chromosome by homologous recombination.[28] Because the suicide vector does not contain an origin of replication for *A. baumannii*, only suicide vector plasmids that recombine into the chromosome of its host can propagate its antibiotic resistance cassette. A counterselectable *sacB*

13

gene was used to select for clones capable of excising the plasmid to yield a markerless in-frame deletion mutant (**Figure 2.2.2**). The resultant mutant did not contain any antibiotic-selectable markers, i.e., it was markerless.
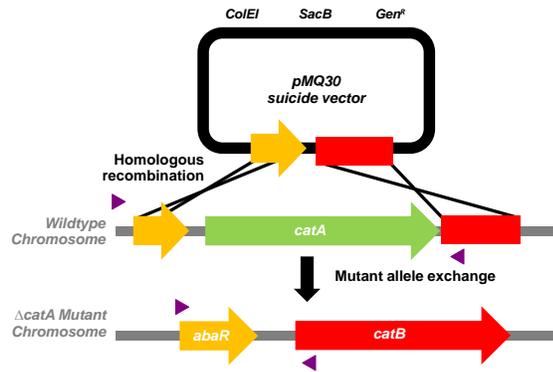


**Figure 2.2.2** Construction of markerless in-frame deletion mutant ΔcatA.

To verify the correct generation of mutants, colony PCR was implemented to yield PCR products consistent with the expected mutant region (**Figure 2.2.3**). To screen for correct Δ*catA*::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 2.2.1**. 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 2.2.2**. Therefore, the wildtype locus yields a larger PCR product of approximately 2.5kb, while the in-frame deleted gene yields a PCR product of approximately 1.0kb.

| Lane | Genomic DNA template | Expected PCR product (bp) |
|---|---|---|
| 1 | ATCC17978 WT | None |
| 2 | ATCC17978 WT | None |
| 3 | ΔcatA::Km | 1548 |
| 4 | ΔcatA::Km | 1376 |
| 5 | ATCC17978 WT | 2525 |
| 6 | ΔcatA | 914 |
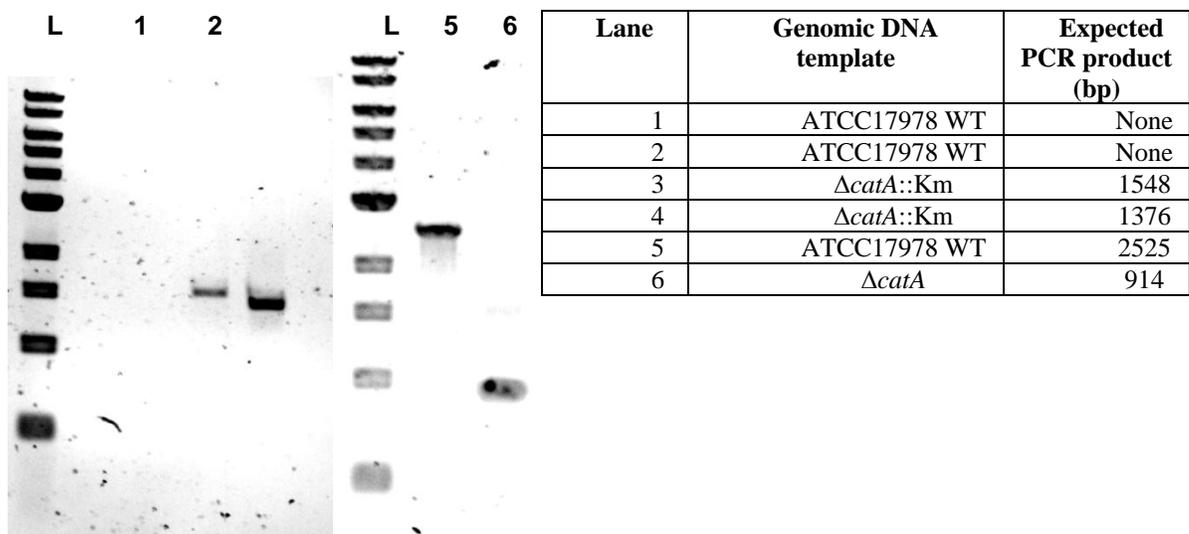
**Figure 2.2.3** PCR products obtained from colony PCR of ΔcatA::Km and ΔcatA mutants.

## 2.3     PHENOTYPIC ANALYSIS OF CAT PATHWAY MUTANTS

Because the *cat* gene cluster can be identified by bioinformatic analysis of every available genome sequence of different *A. baumannii* strains, it is hypothesized that this pathway plays an important role in the survival of *A. baumannii*. To investigate these roles, a number of different phenotypes associated with virulence and survival were implemented to observe differences between the ATCC17978 wildtype and *catA* deletion mutant

Several abiotic and biotic assays, including motility[29],[30], hemolysis[31], desiccation[32], wax worm survival[33], have been explored for assessing the virulence phenotypes of *A. baumannii.*. In some cases, a single gene or pathway exhibits pleiotropic effects and is therefore attributable to a number of different phenotypes. For example, quorum sensing in *A. baumannii* has been shown to modulate both motility and biofilm formation.[29] Others have shown that the RecA DNA repair enzyme has even broader implications in virulence attenuation of *A. baumannii*, demonstrating roles in desiccation, macrophage survival, chemical and

environmental stress (by increased resistance to heat, UV, antibiotic, and oxidation), as well as virulence against mice.[32]

To understand the importance of the *cat* gene cluster to *A. baumannii*, we compared the phenotypic difference between ATCC17978 wildtype and *catA* deletion mutant in the context of desiccation resistance, motility, hemolysis, colony morphology, cytotoxicity against an invertebrate wax worm model *Galleria mellonella*, and heat shock viability.

### 2.3.1   Dessication

Desiccation involves the drying of microbial samples under gentle heating conditions. In these experiments, filter paper is inoculated with *A. baumannii* strains and placed under dry conditions at 37°C. The bacteria are then rescued by placing the filter paper on nutrient-rich (LB) agar plates and desiccation resistance is measured by the growth of *A. baumannii* after incubation on LB agar. It was demonstrated that the ATCC17978 wildtype strain exhibited more robust desiccation resistance compared to the Δ*catA*::Km mutant (**Figure 2.3.1**). In this assay, 1-, 10-, and 100-fold dilutions of overnight culture were spotted onto the filter paper and recovered.
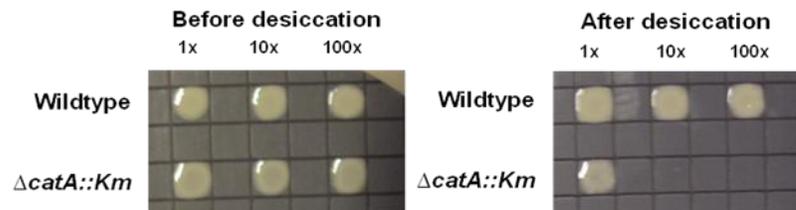


**Figure 2.3.1** Growth recovery of wildtype and mutant with or without desiccation.

## 2.3.2   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. To assess its motility, liquid culture of the Δ*catA*::Km mutant was inoculated on highly purified, low percentage agar (Nobel agar at 0.3% concentrations). The bacteria were allowed to grow and the distance traveled across the surface of the agar plate was used as measure of its motility. Disruption of the *catA* gene resulted in the inability of *A. baumannii* to travel across the surface of the agar (**Figure 2.3.2**).

**Wildtype**                                                    **Δ*catA::Km***
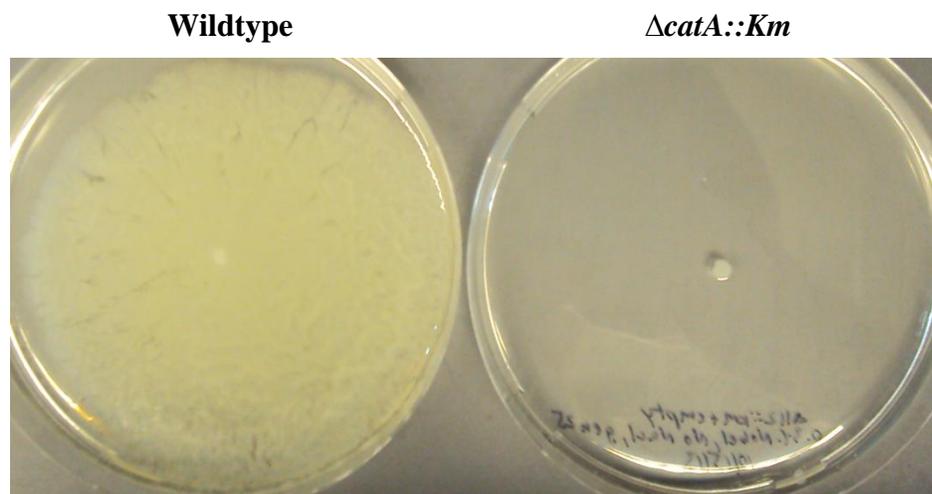


**Figure 2.3.2** Motility deficiency was observed for ΔcatA::Km mutant on low percentage agar.

This elimination of the motility phenotype in the Δ*catA*::Km mutant suggests that the product of the *cat* gene cluster functions as a surfactant, enabling motility of the bacteria across agar, similar to the role of the lipopeptide surfactin in the motility of *Bacillus subtilis*.[15]

## 2.3.3   Hemolysis

Hemolysis involves the destruction of red blood cells and can be categorized as α, β, or γ, depending on whether the hemolysis is incomplete, complete, or absent, respectively. On blood

agar plates, α hemolysis is identified by green discoloration of red blood cells and is attributable to the oxidation byproduct of heme, biliverdin. β hemolysis arises due to the complete destruction of red blood cells and results in a transparent region around the bacteria colony. Lastly, γ hemolysis is designated for the absence of apparent hemolytic activity.

To examine the hemolytic role of the *cat* operon, horse blood agar was inoculated, bacteria were allowed to grow, and following growth, hemolysis was observed (**Figure 2.3.3**). An increased transparent region was observed around the wildtype colony, implicating a β-hemolytic role of the *cat* gene cluster.

**Wildtype**     **Δ*catA::Km***



**Decreased hemolysis was observed for mutant Δ*catA*::Km.**

**Figure 2.3.3** Decreased hemolysis observed for ΔcatA::Km mutant on 5% horse blood agar.

### 2.3.4   Colony Morphology

Bacterial colony morphology has been shown to be an important indicator of environmental stress. The earliest known example of this trait was observed in *Vibrio cholera*, the causative agent of the disease cholera. *V. cholera* is able to switch between rugose to smooth morphologies under different environmental cues, such as osmotic stress and acidity. The rugose morphology has been associated with the ability to withstand these unfavorable growth environments, while smooth colonies do not. These phenotypes as well as its rugosity has been attributed to a gene cluster responsible for the production of an exopolysaccharide in *V. cholera*.[35]

To assess the role of the *cat* gene cluster in colony morphology, each wildtype or mutant strain was inoculated on agar plates and allowed to grow. Following growth, the colony

morphologies were observed for each strain (**Figure 2.3.4**). Inspection of each colony suggests that disruption of *catA* results in a smooth morphology and therefore the *cat* gene cluster could be responsible for environmental stress adaptation.
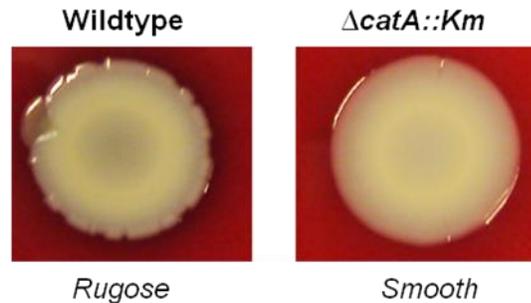


**Figure 2.3.4** Colony morphology of the wildtype and mutant.

## 2.3.5    Virulence Against Galleria mellonella

*Galleria mellonella*, also known as the wax worm, has become an important organism to investigate pathogenesis of *A. baumannii*. Experiments to determine the pathogenicity of microorganisms proceed by inoculating the model host with the pathogen of interest and measuring the rate of wax worm survival.[33] To assess the virulence of the *cat* gene cluster of *A. baumannii*, the wildtype and Δ*catA*::Km were introduced into wax worms and the number of surviving worms were measured periodically over a period of four days (**Figure 2.3.5**). The increased survival of wax worms after inoculation with the Δ*catA*::Km mutant compared to their inoculation of the wildtype *A. baumannii* suggests that the *cat* gene cluster functions to increase the pathogenicity of *A. baumannii*.

**G. mellonella Survival after A. baumannii**

**Inoculation Over Time**
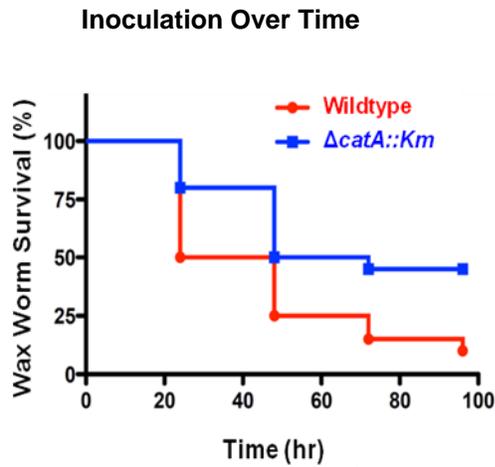


**Figure 2.3.5** Wax worm Galleria mellonella survival rate after inoculation with the wildtype and mutant.

## 2.3.6 Heat Shock Resistance

An important measure of a pathogen's virulence is represented by its ability to withstand periods of elevated heat. Heat shock involves exposure of the bacteria to periods of heat and to subsequently recover them in nutrient-rich media and to physiological temperatures. Under this stress, heat shock proteins are upregulated to ameliorate protein damage.[36] To this end, heat shock resistance was assessed between the wildtype and Δ*acat::Km* mutant (**Figure 2.3.6**), which revealed that the mutant was more sensitive to heat shock, as evidence by the decreased recovery after exposure to 10 minutes of elevated heating (55°C).
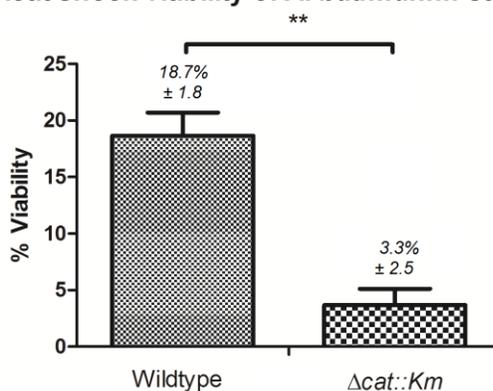
**Figure 2.3.6** Heat shock viability was analyzed between the wildtype and mutant.

## 2.4      COMPLEMENTATION OF CAT PATHWAY MUTANTS

In addition to understanding the roles of these genes by comparative genetic analysis between wildtype and mutant strains, it is important to complement the observed mutant phenotypes. To accomplish this, the genes of interest are introduced to the mutant *in trans* (on a plasmid vector) and restoration of the phenotype should be observed. However, few plasmids are available for use in *A. baumannii* and none have been designed to enable the facile and rapid construction of entire biosynthetic pathways. Of the few reported, *A. baumannii* plasmids have been designed to function using a variety of promoters for transcription, including the *Tet*[37], β-lactamase CTX-M14 gene[27], and the arabinose[38] promoters; however, all of these depend on traditional cloning techniques based on restriction enzymes and DNA ligases.

To address the shortcomings of previously reported *A. baumannii* plasmids, a new expression plasmid tool based on a yeast-recombineering plasmid[28] was constructed to include an arabinose-titratable promoter and an *A. baumannii* origin of replication.

In this design, a sequence containing the origin of replication for *A. baumannii*[38] was introduced into a parent plasmid pMQ124[28]. This resulted in a shuttle vector compatible with the yeast *Saccharomyces cerevisiae*, *E. coli*, and *A. baumannii*. The successful propagation of

this shuttle vector was demonstrated in various strains of *A. baumannii*, including several clinical isolates (obtained from Dr. Yohei Doi, Division of Infectious Disease, University of Pittsburgh Medical Center). Additionally, a number of individual genes as well as the entire *cat* gene cluster was assembled into this pMQ124-derived plasmid. To verify that this *A. baumannii* expression plasmid is functional and responsive to arabinose induction, the *LacZ* reporter gene was cloned downstream of the arabinose promoter, $P_{BAD}$ and expression of *LacZ* was measured from *A. baumannii*. To measure *LacZ* expression, the hydrolysis of a chromogenic substrate, *o*-nitrophenyl-D-galactoside (ONPG), was measured as a function of arabinose concentration (**Figure 2.4.1**). The results verify that the $P_{BAD}$ promoter can be finely tuned for expression of genes in this new *A. baumannii* plasmid vector.



**Figure 2.4.1** Arabinose-dependent expression of *LacZ* demonstrated for new *A. baumanni* expression plasmid tool.

To substantiate the role of the *catA* gene in desiccation resistance and motility, a plasmid containing the *catA* gene was introduced to the Δ*catA*::Km mutant to analyze whether *in trans*

genetic complementation of these phenotypes could be achieved. Attempts to complement the mutant phenotypes by introducing a plasmid containing *catA* into Δ*catA*::Km had failed (**Figure 2.4.2**). Two possible explanations for this observation include: (1) a mutation in a locus elsewhere in the genome of Δ*catA*::Km is responsible for these mutant phenotypes, or (2) disruption of the *catA* gene with the kanamycin resistance cassette had introduced a polar effect, disturbing partially or fully the expression of genes downstream in the operon. An experiment to test the former hypothesis would not be easily accomplished. On the other hand, a polar effect is plausible, given the operon organization and placement of the kanamycin cassette.

Δ*catA*::Km with empty plasmid          Δ*catA*::Km with A1S_0112
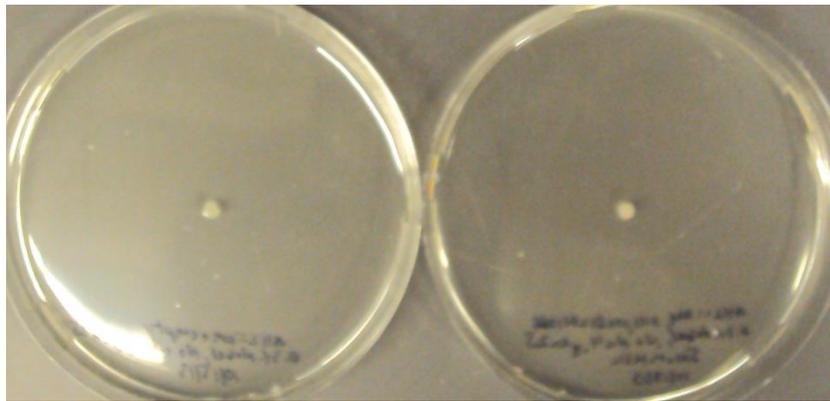


**Figure 2.4.2** Inability to restore motility of the Δ*catA*::Km mutant.

Due to the absence of nuclear membranes, bacteria possess a complex but efficient system to synchronize transcription and translation as a regulatory mechanism. For example, this synchronization has been well established for the tryptophan biosynthetic gene cluster *trp* of *E. coli* and *Serratia marcescens*, demonstrating that RNA polymerase pause during transcription until the ribosome has sufficiently traversed and translated the corresponding mRNA transcript.[39] Therefore, inefficient navigation of the ribosome across its mRNA transcript halts further transcription. Another facet of the complex gene regulatory mechanisms of bacteria is

demonstrated by translational coupling. In this mechanism, the ribosome is able to efficiently initiate translation of downstream genes only after complete translation of the preceding gene.[40]

Two hypotheses have been proposed to explain this occurrence. One argues that the ribosome for translation of a downstream gene must be recruited from the stop codon from the upstream gene.[41], [42] The other argues that ribosomal translation of an upstream gene can release RNA secondary structures that may be present in the ribosomal recognition sequence of the downstream gene, enabling efficient translational initiation.[43], [44]

A notable feature of the defined Δ*catA*::Km mutant is the presence of two stop codons upstream of *catB*, one from *catA*, the other from the kanamycin resistance cassette, situated approximately 500bp upstream of *catB*. Due to these characteristics of major gene regulatory mechanisms of bacteria, a polar effect could explain the failure to genetically complement the function of *catA* of the Δ*catA*::Km mutant. Specifically, premature dissociation of the ribosome from the mRNA transcript due to the stop codon of the kanamycin resistance cassette would disallow translation of downstream genes, e.g., *catB*, *catC*, etc. Further, discontinued translation of the downstream genes of the operon would also discontinue transcription of those genes. A major consequence, therefore, would manifest in the reduced or absent transcription of the *cat* gene cluster.

To understand whether the polar effect manifests at the transcriptional level, reverse transcriptional PCR (RT-PCR) was executed. In this experiment, the in-frame deletion mutant Δ*catA* was analyzed in parallel. The in-frame deletion mutant was generated to delete approximately 95% of the *catA* gene with 5' and 3' flanking regions intact to avoid interference with any possible regulatory sequences. In addition, this deletion was in-frame, avoiding any possible polar effects due to frameshift mutations.

Wildtype or mutant strains were grown to mid-log exponential growth phase, a complementary DNA (cDNA) library was generated, and PCR was executed to amplify sequences from genes of the *cat* operon (**Figure 2.4.3**).
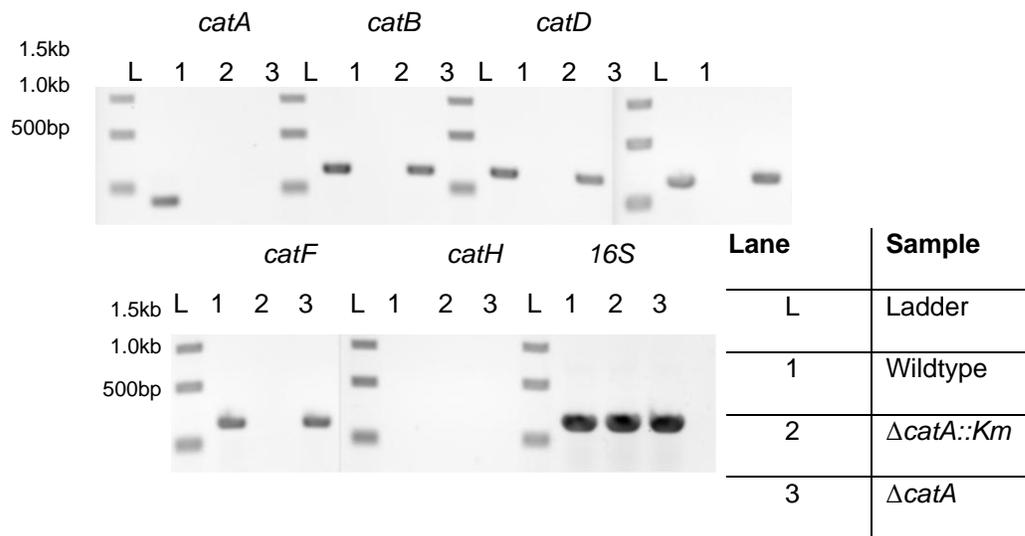
**Figure 2.4.3** RT-PCR analysis of wildtype and mutants Δ*catA* and Δ*catA*::Km.

RT-PCR analysis of the genes of the *cat* operon provides evidence that indeed the Δ*catA*::Km mutant exhibits a polar effect and as a consequence, the entire gene cluster *catABCDEFGH* is non-functional. Interestingly, the *catH* gene was not detected for any of the strains, suggesting it is not expressed under normal culturing conditions and regulated by an alternative mechanism. *catH* is predicted to encode a 4'-phosphopantetheinyl transferase, required for posttranslational modification of acyl and peptidyl carrier proteins.

With knowledge that the disruption of the *catA* gene with the kanamycin resistance cassette results in a polar effect of downstream genes, *in trans* complementation proceeded by introducing a plasmid containing the entire *cat* operon, including *cat* genes *A-H*, into the Δ*catA*::Km mutant. Additionally, to validate that the in-frame deletion mutant Δ*catA* did not impose a polar effect on downstream genes, a plasmid containing only *catA* was introduced into the Δ*catA* mutant. As expected, *in trans* complementation of the motility phenotype was achieved for the Δ*catA*::Km mutant containing a plasmid with the *catABCDEFGH* and for the Δ*catA* mutant containing a plasmid for the *catA* gene only (**Figure 2.4.4**). This further supports the hypothesis that a polar effect of the remaining *cat* operon exists, resulting in diminished

expression of these genes. In conjunction with the genetic complementation of the motility phenotype of the Δ*catA* by introduction of a plasmid containing *catA*, the restoration of motility for the Δ*catA*::Km mutant substantiates the notion that the *cat* operon serves an important role in the motility of *A. baumannii*.

Insight into the polar effect demonstrated in the Δ*catA*::Km mutant prompted the investigation of the polar effect on desiccation sensitivity. To this end, the desiccation sensitivity of the Δ*catA* mutant was assessed (**Figure 2.4.5**).

ΔcatA + empty plasmid

ΔcatA + catA

ΔcatA::Km + empty plasmid

ΔcatA::Km + entire *cat* pathway
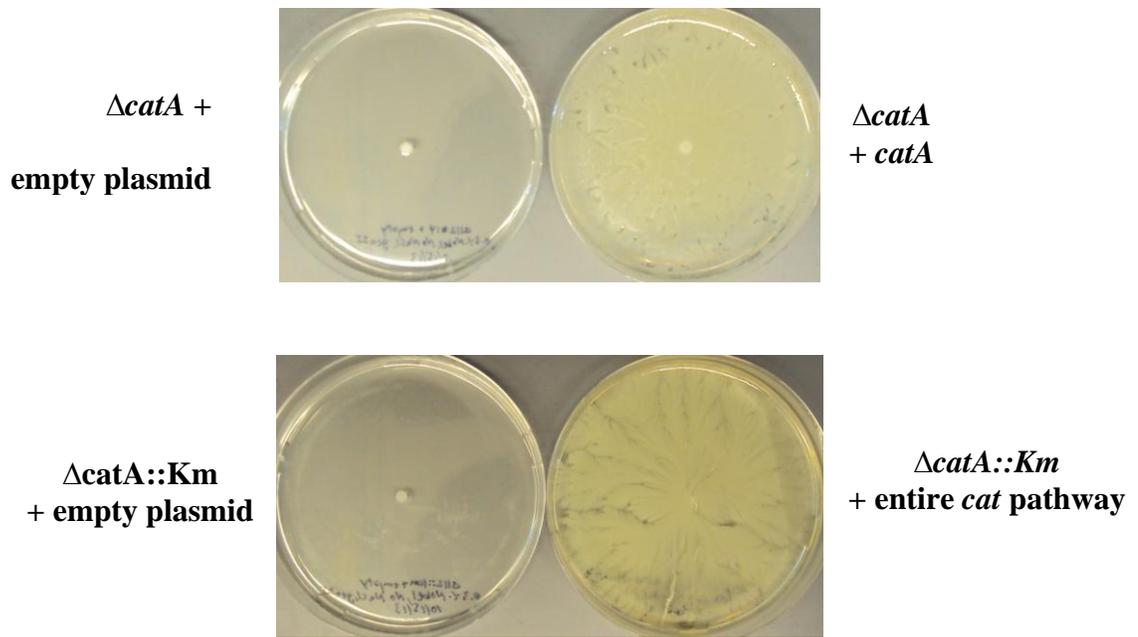
**Figure 2.4.4** Complementation of the motility phenotype  in  Δ*catA*::Km and Δ*catA* mutants.
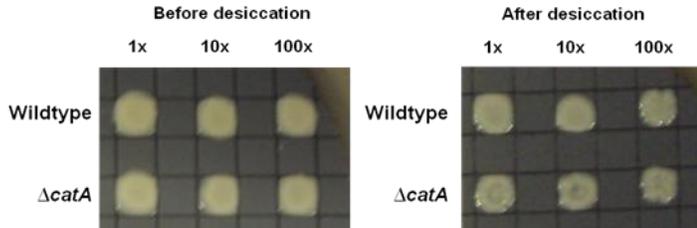


**Figure 2.4.5** Growth recovery of the wildtype and mutant with or without desiccation.

Both wildtype and the Δ*catA* mutant appeared to recover with equal vitality and all dilutions of inoculum had recovered after desiccation. In conjunction with the notable desiccation sensitivity seen for the Δ*catA*::Km mutant, this observation suggests that disruption of the entire *cat* operon is required for complete desiccation sensitivity. To demonstrate that the entire *cat* gene cluster is responsible for desiccation resistance, a genetic complementation experiment was executed. A plasmid containing the *cat* gene cluster was introduced to the Δ*catA*::Km mutant and partial restoration of the desiccation resistance phenotype was observed, supporting the attribution of the entire *cat* operon to observed desiccation sensitivity (**Figure 2.4.6**).
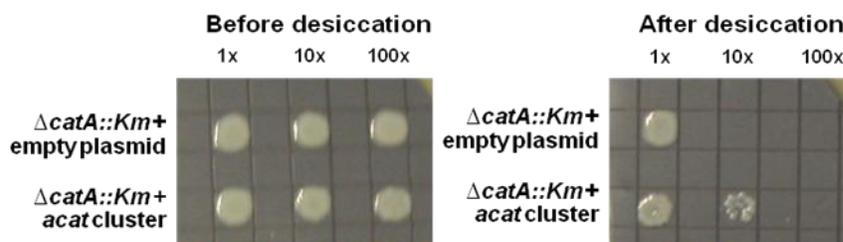


**Figure 2.4.6** Post-desiccation growth recovery of the Δ*catA*::Km mutant.

## 2.5    CAT PATHWAY REGULATION AND ITS SIGNIFICANCE

Quorum sensing (QS) is a phenomenon exhibited by bacteria to mediate population-dependent behavior. QS was first observed in *Vibrio fischeri,* whereby its luminescence was found to be dependent on population density. Specifically, individual *V. fischeri* bacterium found freely in the ocean did not display luminescence, while those inhabiting the squid *Euprymna scolopes* were found to exhibit luminescence. This led to the discovery of the *lux* operon, the genes responsible for luminescence.[45],[46]

Two hallmark features of the *lux* operon include the *luxI* and *luxR* genes. *luxI* encodes the autoinducer synthase responsible for biosynthesis of the acyl-homoserine lactone (AHL) signaling    molecule    that    can    freely    diffuse    across    the    membrane    to    interact

intracellularly.[47],[48]*luxR* encodes a transcriptional regulator that, upon binding to AHL, will transcribe QS-dependent genes. In high cell density, AHL concentrations achieve a threshold that, upon binding to intracellular LuxR protein, will enable the transcription of the *lux* operon. Therefore, the LuxR-AHL complex promotes the expression of *luxI* and *luxR*, resulting in a positive feedback loop of the QS pathways.[45],[46]

QS has also been demonstrated to be a crucial element in the virulence of human pathogens.[43] For example, QS mutants of the human pathogen *Pseudomonas aeruginosa*, the causative agent of chronic infection of patients with cystic fibrosis, were shown to exhibit reduced pathogenicity and host colonization in mouse model studies.[49],[50]

### *abaR*-Dependent Transcription of *catA*



| Group | Experiment |
| --- | --- |
| 1 | Wildtype + empty plasmid |
| 2 | Wildtype + AHL |
| 3 | Wildtype + *abaR* |
| 4 | Wildtype + *abaR* + AHL |

N = 4
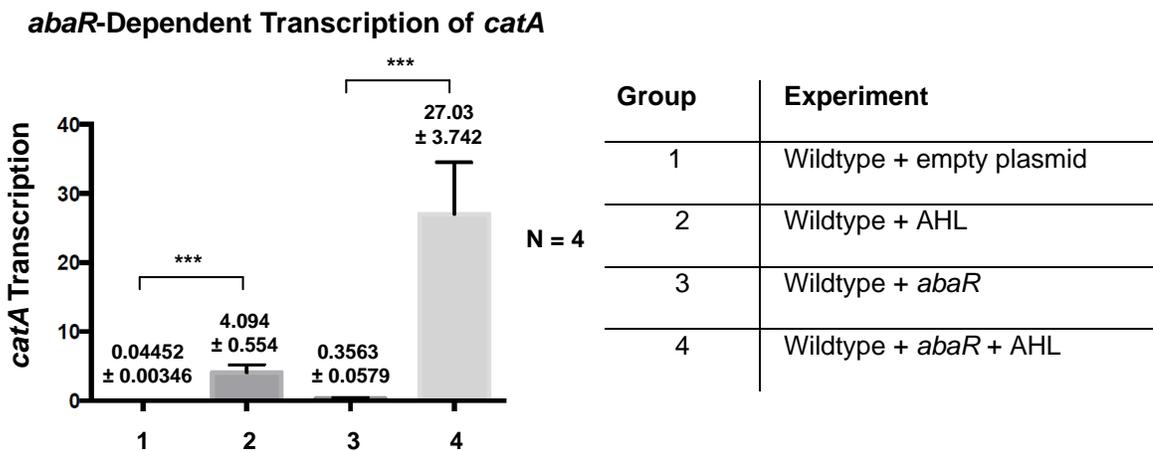
**Figure 2.5.1** qPCR measurements of *catA* transcription from wildtype *A. baumannii* expressed with the transcriptional regulator *abaR* with or without the AHL.

Recently, a QS system was identified and characterized in *A. baumannii* and its activation has been attributed to virulence-associated phenotypes including biofilm formation[26] and motility[30]. Two genes, *abaI* and *abaR*, were discovered to share high sequence similarity to genes *luxI* and *luxR*, respectively, and disruption of *abaI* resulted in the motility deficiency phenotype. Furthermore, transcriptional analysis of the *cat* operon was accomplished by the introduction of an exogenous AHL molecule, *N*-(3-hydroxy)-dodecanoylhomoserine lactone (3-

OH $C_{12}$-HSL), which was found to upregulate the expression of *catA* and presumably of the remaining genes of the *cat* operon.[30] With consideration of this previous study, quantitative PCR (qPCR) was implemented to substantiate these claims (**Figure 2.5.1**). In this experiment, overexpression of *abaR* and introduction of 3-OH $C_{12}$-HSL resulted in an approximately 100-fold increase in *catA* transcription compared to overexpression of *abaR* alone, which further supports the view that *abaR*, like *LuxR*, is transcriptionally active only when bound to its cognate homoserine lactone constituent.

Attempts to restore the mutant motility phenotype by overexpression of *abaR* in the presence of 3-OH $C_{12}$-HSL failed, suggesting that the *cat* gene cluster is essential for motility (**Figure 2.5.2**). Experiments to restore motility deficiency included both mutants Δ*catA* and Δ*catA*::Km overexpressed with *catR* in the presence of 3-OH $C_{12}$-HSL. These results serve as direct evidence that, relative to the *aba* genes, the *cat* pathway is a downstream regulatory point for the motility phenotype.

| **Wildtype**<br>**+ *abaR***<br>**+ 3-OH $C_{12}$-HSL** | **Δ*catA***<br>**+ *abaR***<br>**+ 3-OH $C_{12}$-HSL** | **Δ*catA*::Km**<br>**+ *abaR***<br>**+ 3-OH $C_{12}$-HSL** |
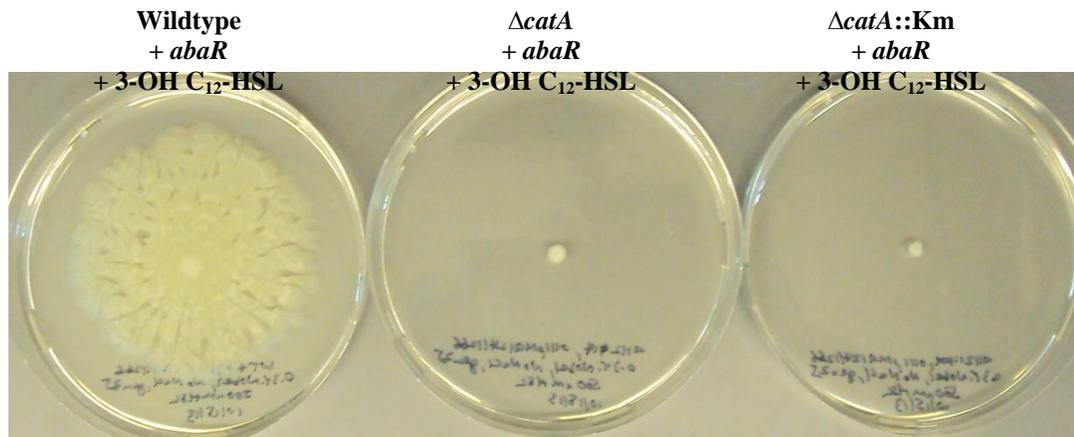


**Figure 2.5.2** Motility deficiency of the Δ*catA* and Δ*catA*::Km mutants not restored with *abaR* overexpression in presence of 3-OH $C_{12}$-HSL.

## 2.6    CONCLUSIONS


Investigations into the evolutionarily-conserved *cat* gene cluster of *A. baumannii* have provided insight into the importance of these genes and implications of the putative lipopeptide to the physiology of this organism. The *cat* gene cluster has been shown to be involved in motility, desiccation resistance, colony morphology, and wax worm virulence. It has also been shown that for desiccation sensitivity that only disruption of the entire pathway is required to observe pronounced mutant phenotypes. This complete pathway disruption was verified by both gene transcriptional analysis (qPCR) as well as genetic complementation experiments, restoring partially or completely motility and desiccation wildtype phenotype. Lastly, the *cat* pathway was found to be a downstream regulator of the motility phenotype.

In light of these mutant phenotypes, the *cat* gene cluster is an ideal drug target for next generation, virulence attenuation-based antibiotics. The phenotypes discussed are pertinent to bacterial pathogenicity and can provide insight into how *A. baumannii* has become a persistent nosocomial pathogen.

Although a number of virulence phenotypes have been established for the *cat* gene cluster, the isolation and characterization of the PKS-NRPKS putative lipopeptide product has remained elusive. Therefore, future directions in this project will entail the identification of the molecule of interest and chemical complementation experiments to restore the wildtype phenotypes in *cat* gene cluster mutants.

# 3.0 HAPALOSIN BIOSYNTHESIS IN HAPALOSIPHON WELWITSCHII AND ITS COMBINATORIAL GENERATION IN ESCHERICHIA COLI

Secondary metabolites from cyanobacteria of the order *Stigonematales* have proven to be prolific sources of medicinally valuable compounds, including hapalindoles[55], fischerellins[56], hapalosin[57], and ambigols[58], with roles in anticancer, anti-inflammatory, and anti-fungal activities. Despite these established bioactivities, nothing is known about their biosynthesis and only recently have the genomes of these cyanobacterial species been sequenced[59]. Moreover, these species are not genetically amenable and no genetic tools are available for these investigations.

Recent characterizations of sequenced genomes from species of the order Stigonematales have indicated that a significant portion of their genomes are dedicated to polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) genes. For example, approximately 5% of the *Fischerella* sp. genome was found to consist of PKS and NRPS genes.[59] Knowledge of the vast repertoire of secondary metabolite biosynthetic genes and the unavailability of genetic tools to exploit this biosynthetic machinery prompted the call to develop synthetic biology tools to investigate these pathways. To establish a proof of concept for the utility of novel gene expression plasmid tools, a large PKS-NRPS pathway hypothesized to produce hapalosin was discovered bioinformatically (i.e., "genome mining") from the cyanobacterial strain *Hapalosiphon welwitschii* UTEX1830 and was cloned from the native host and its function was explored heterologously in *E. coli*.

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

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 a collective effort to investigate the biosynthesis of hapalindole-type indole alkaloids. The available genome information of these cyanobacteria provided us an intermediate opportunity to mine other biosynthetic pathways related to known natural products of Stigonematales origin. To this end, we initiated an effort to search for the biosynthetic gene cluster responsible for the assembly of hapalosin. Hapalosin is a 12-membered cyclic depsi-lipopeptide that was originally discovered from *H. welwitschii* by Moore and coworkers[57]. It was shown to exhibit potent multidrug resistance (MDR) reversing activity of a human ovarian cancer cell line more effectively than verapamil, one of the first MDR-reversing agent studied clinically.[57] Due to the intricate biological activities, a number of total syntheses related to hapalosin have been reported.[60], [61], [62], [63], [64]

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, it was not clear from the onset how the biosynthesis will be initiated and terminated.

Using our assembled draft genome of *H. welwitschii* UTEX1830, a 26 kbp-long gene cluster that contains five PKS-NRPS-like genes was located. Based on its coding protein homology and related domain prediction, we hypothesize it is responsible for the biosynthesis of hapalosin and was therefore named the *hal* gene cluster (**Figure 3.1.1**).

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[65], 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[66]. *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.

With these bioinformatic predictions in hand, we set out to systematically address if our hypothesis is correct. As there is no validated genetic tool available for *H. welwitschii*, the characterization of the *hal* gene cluster is limited to either *in vitro* expression of each individual Hal protein *in vitro* or heterologous expression of the entire pathway in a different organism. Due to the sheer size of Hal proteins (the average size of the *halB-D* genes is 5858bp), we quickly ruled out the former approach. Instead, we resorted to the development of a more general synthetic biology toolbox that can allow us to quickly assemble large biosynthetic pathways and regulate their expression in a heterologous system.



| Gene | Size (bp) | Domain organization |
|------|-----------|---------------------|
| *halA* | 2127 | NRPS (AL-T) |
| *halB* | 5946 | PKS (KS-AT-MT-KR-T) |
| *halC* | 7047 | NRPS/PKS (C-A-A-KR-T) |
| *halD* | 4581 | NRPS (C-A-MT-T) |
| *halE* | 5457 | PKS (KS-AT-KR-T-TE) |
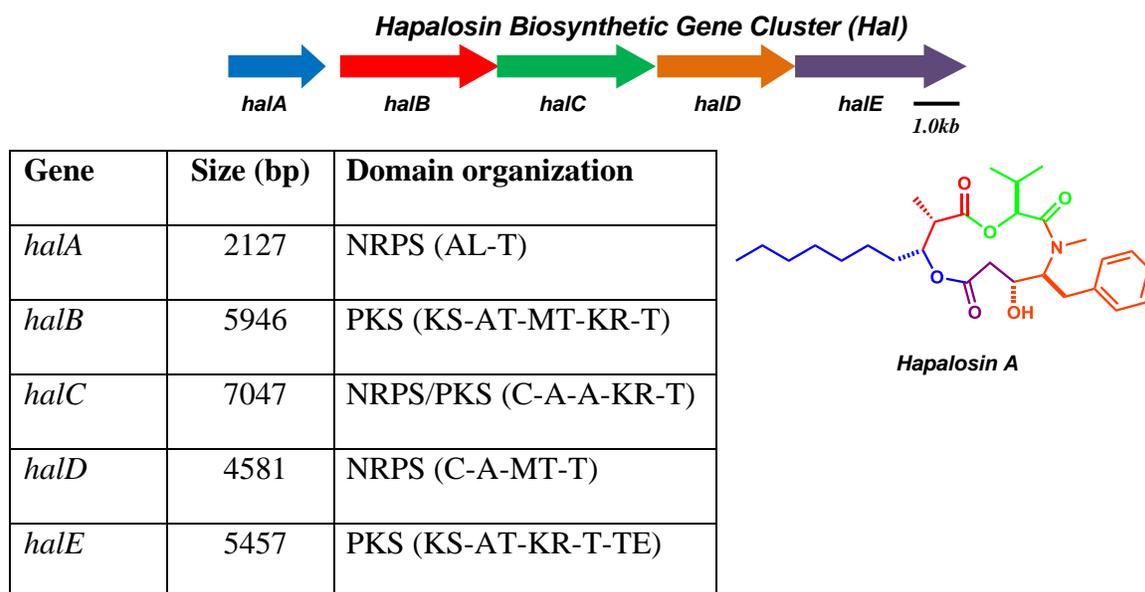
**Figure 3.1.1** Organization of the hapalosin gene cluster

## 3.2    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 by polymerase chain reaction (PCR) amplification and used to construct the expression plasmids via yeast-assisted homologous recombination (yeast recombineering), which exploits yeast's natural mechanism of DNA repair to seamlessly recombine homologous sequences of DNA together.[67],[68] 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 pXL3 and 4) or Tac (plasmid pXL2) promoters (**Figure 3.2.1**). 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 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 pXL2 and pXL4 were verified by restriction enzyme DNA digestion analyses (**Figure 3.2.2** and **Table 1**). The observed DNA fragments were consistent with the expected DNA products for each respective restriction enzyme digestion analysis.
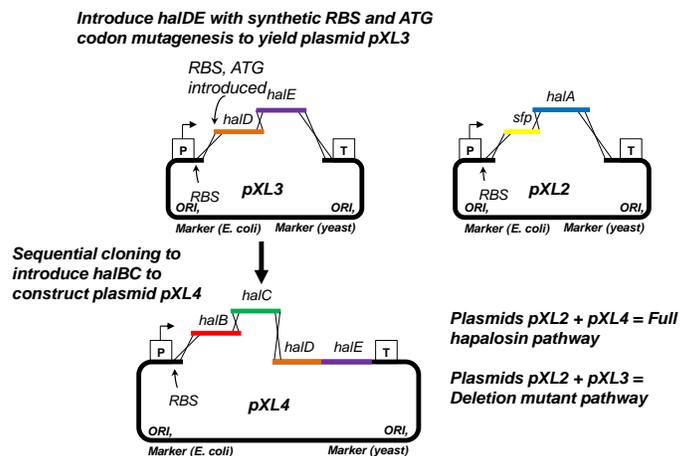
**Figure 3.2.1** Construction of hapalosin expression plasmids pXL2 – 4.
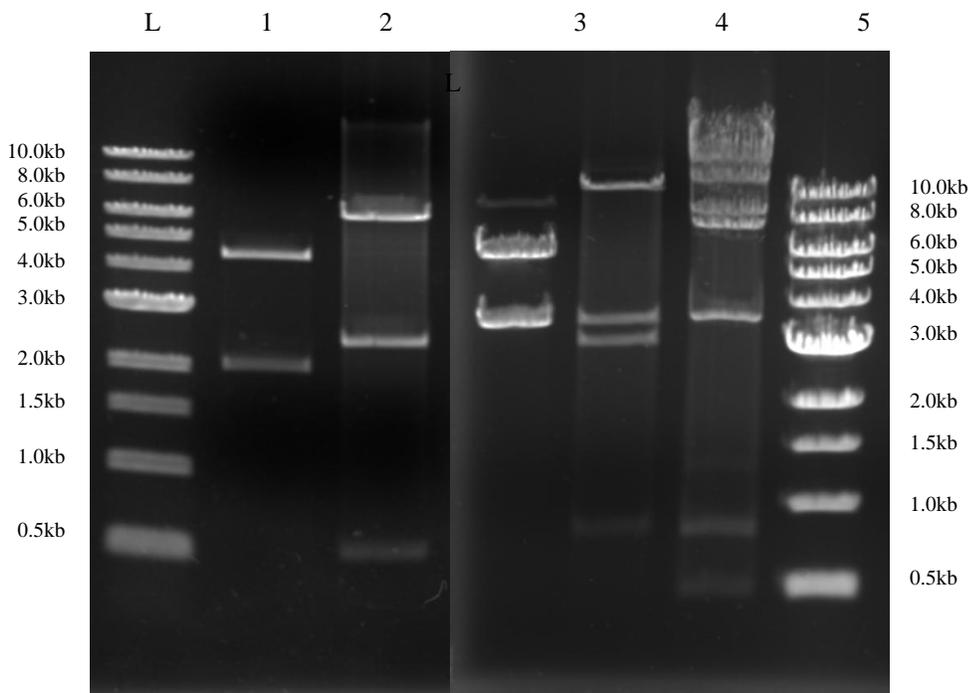


**Figure 3.2.2** Successful construction of expression plasmids containing the genes of the hapalosin
biosynthetic pathway.

**Table 1** Restriction enzyme digest analysis of hapalosin expression plasmids

| Lane | Plasmid | Digest analysis | Expected products (bp) |
| --- | --- | --- | --- |
| 1 | pMQ131 | HindIII | 4461, 2019, 13 |
| 2 | pXL2 | HindIII | 6124, 2406, 499, 13 |
| 3 | pMQ123i | BglII | 6076, 3621 |
| 4 | pXL3 | BglII | 11426, 3621, 3090, 837 |
| 5 | pXL4 | BglII | 11426, 8299, 7286, 3671, 837, 479 |

## 3.3    HETEROLOGOUS EXPRESSION OF HAPALOSIN GENE CLUSTER AND ITS DELETION MUTANT IN ESCHERICHIA COLI

To assess the production of the cyanobacterial natural product hapalosin A in a non-native host, the plasmids containing the genes for the complete biosynthetic pathway were introduced into an *E. coli* XL1-Blue expression strain. *E. coli* XL1-Blue strains containing the hapalosin biosynthetic pathway were grown and induced for gene expression. Following induction, culture supernatants were extracted with ethyl acetate and analyzed by liquid chromatography-mass (LC-MS) and proton nuclear magnetic resonance ($^1$H-NMR) spectrometries.

The biosynthesis of hapalosin A from *E. coli* was compared to a standard extract obtained from the *Hapalosiphon welwitschii* strain UTEX1830, the natural producer organism. The molecular mass and retention times by LC-MS analysis were used to verify that the putative hapalosin A compound obtained from *E. coli* was the same as that from the cyanobacterial species. 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 3.3.1**. Three other prominent peaks were also discovered in this expression culture, labeled **B**, **C,** and **D**. The characterization of these compounds will be discussed in further detail later in this document. Following LC-MS analysis of hapalosin A, the compound was isolated and analyzed by $^1$H-NMR (**Figure 3.3.2**). Hapalosin A was also analyzed by HR-MS$^2$ analysis, yielding a characteristic fragmentation pattern (**Figure 3.3.3**).

The introduction of genes *halABCDE* from *Hapalosiphon welwitschii* UTEX1830 into the heterologous host *E. coli*, the regulated expression of these genes, and the production and identification of hapalosin establishes for the first time the relationship between the genes responsible for biosynthesis and the final compound of interest, hapalosin A.

**E. coli *halABCDE* expression**

**UTEX1830 extract (standard)**

**E. coli *halDE* expression only (background)**
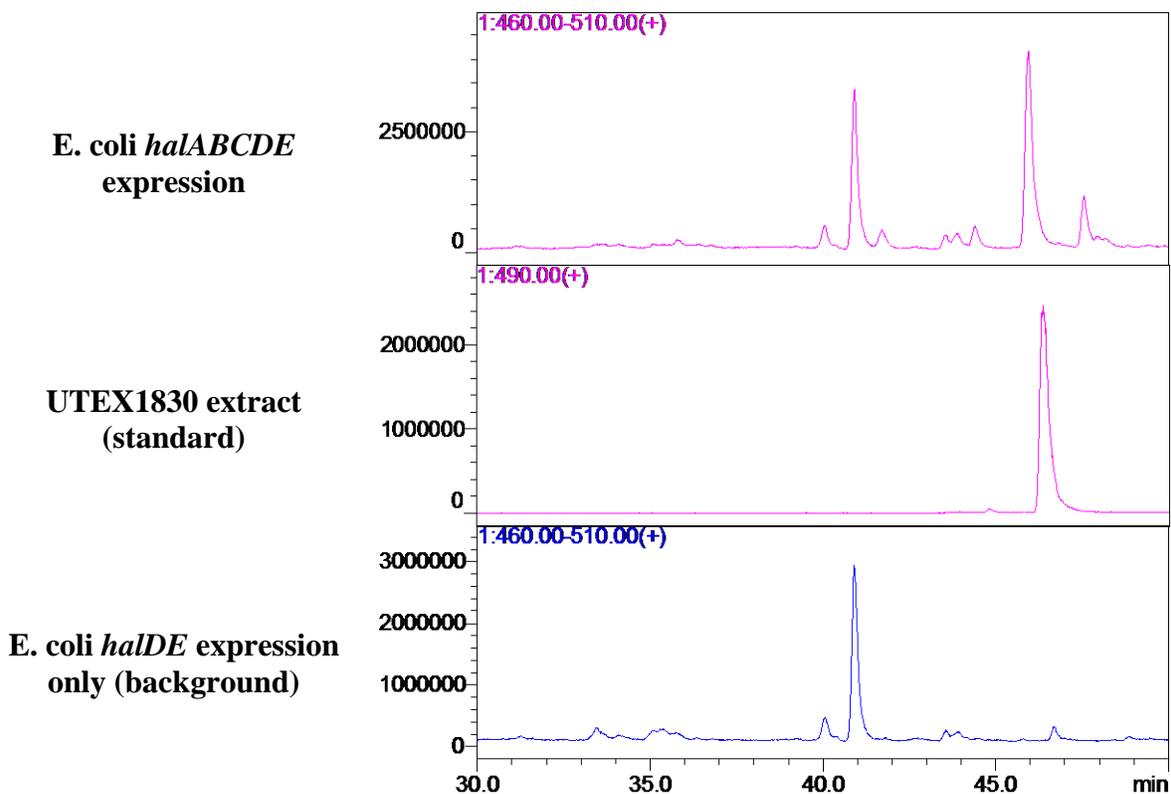
**Figure 3.3.1** Comparative LC-MS chromatograms of hapalosin biosynthetic pathway production

38
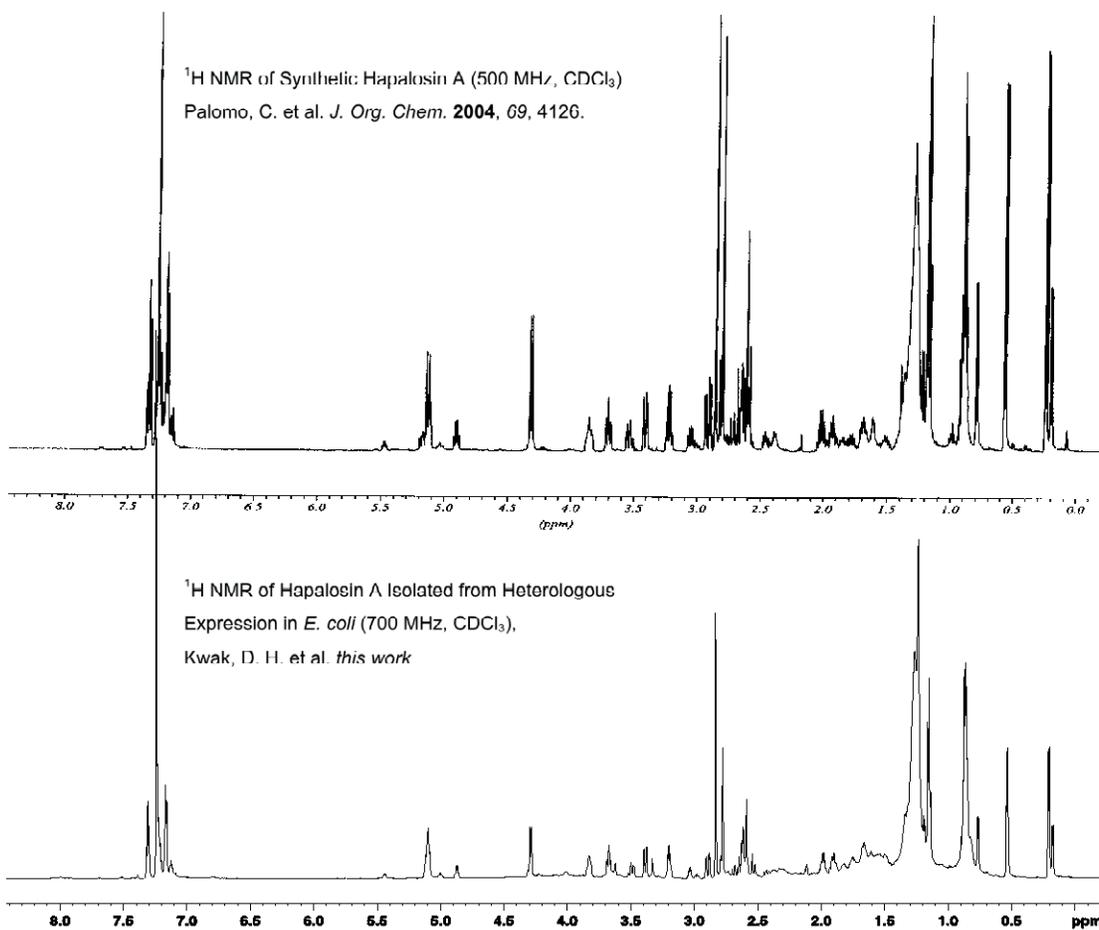
**Figure 3.3.2** Comparison of $^1$H-NMR spectra for hapalosin

39
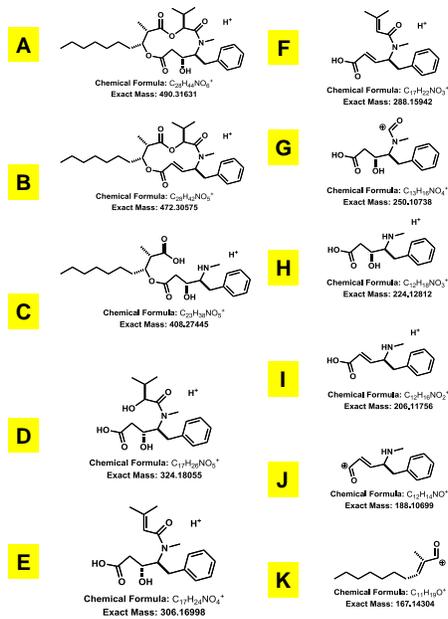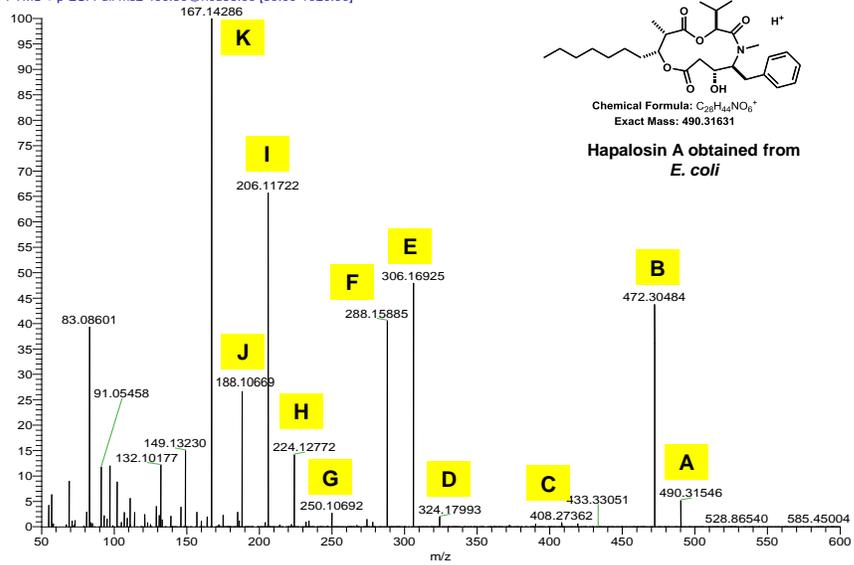
| Peak | Molecular formula | Mass (Da) | Observed Mass (Da) | Error (ppm) |
|---|---|---|---|---|
| A | $C_{28}H_{44}NO_6^+$ | 490.31631 | 490.31546 | 1.7 |
| B | $C_{28}H_{42}NO_5^+$ | 472.30575 | 472.30484 | 1.9 |
| C | $C_{23}H_{38}NO_5^+$ | 408.27445 | 408.27362 | 2.0 |
| D | $C_{17}H_{24}NO_4^+$ | 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 3.3.3** The HR-MS$^2$ fragmentation of hapalosin A obtained from heterologous expression

As discussed previously, comparative LC-MS analysis of extracts from *Hapalosiphon welwitschii* and *E. coli* expression strains containing the full *halABCDE* or the partial pathway *halDE* genes revealed the presence of two compounds present only in the *E. coli* strain expressing the entire biosynthetic pathway. These compounds are labeled **B**,**C** and **D** in **Figure 3.3.1** and their nominal masses were determined to be 504, 462, and 486 Da, respectively, whereas the nominal mass of hapalosin A is 490 Da. Inspection of these numbers suggested the possibility of hapalosin analogs varying only in $-CH_2-$ (14 Da) and $-C_2H_4-$ (28 Da) groups. Indeed, HR-MS$^2$ analysis verified the presence of hapalosin analogs corresponding to hapalosins B, C and D (**Figure 3.3.4**). Hapalosin B had only previously been reported to be produced naturally from the cyanobacteria strain *Nostoc sp.* 4A[70]; therefore, this finding represents the first discovery of hapalosin B from *Hapalosiphon welwitschii* UTEX1830.

The structural elucidation of hapalosins C and D and consideration of the enzymes responsible for each chemical transformation in the biosynthesis of hapalosin suggested differential synthesis at the fatty acid ligase domain of HalA and at the *C*-methyltransferase domain of HalB, respectively. Therefore, in addition to differential selectivity 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.

**Hal Pathway**





| 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 | $-CH_2CH_3$ | $-H$ | $-H$ |
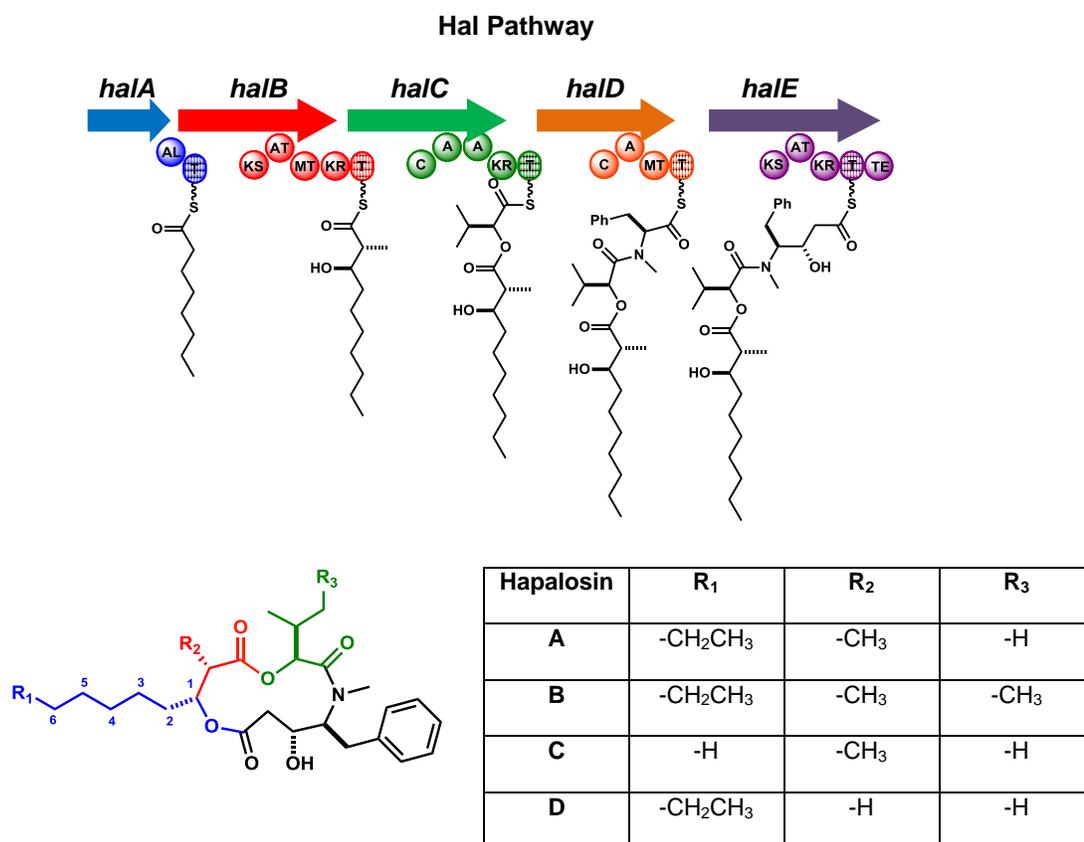
**Figure 3.3.4** Structures of hapalosins A-D provide insight into the differential biosynthesis of hapalosin products within the pathway.

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* was insufficient to biosynthesize hapalosin (**Figure 3.3.1**). Further investigations led to the discovery that expression of only *halBCDE* also enabled the synthesis of hapalosin, although production was comparably modest to that of the full pathway (**Figure 3.3.5**). This peculiar result suggested, then, that the function of HalA could be complemented by an enzyme not part of the native pathway.
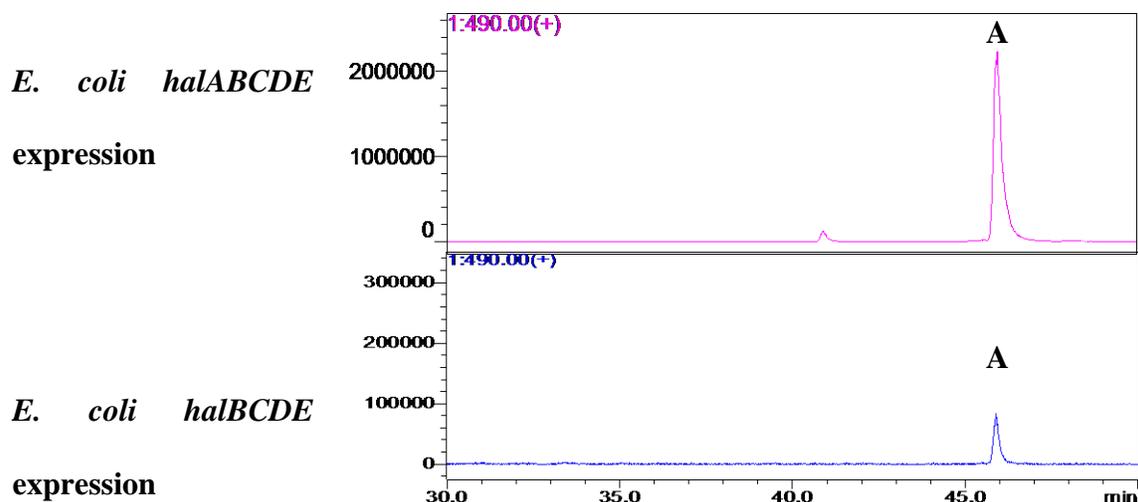
**Figure 3.3.5** LC-MS chromatogram comparison of *E. coli* expression strains

Lipoic acid in *E. coli* serves an essential role as a cofactor to key enzymes involved in primary metabolism, similar to the role of 4'-phosphopantetheinyl moiety to ACPs and PCPs of NRPS and PKS enzymes by shuttling biosynthetic intermediates along the primary metabolic pathway.[71]

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.[72]

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. Although there is no direct evidence of interactions between LipB and HalB, the production of hapalosin from HalBCDE in *E. coli* nonetheless demonstrates that domains of the Hal gene cluster can interact with those from *E. coli*, a rare event as interactions between domains are highly specific.[73] This

43

unexpected result, therefore establishes some capacity of enzyme-enzyme promiscuity between the enzymes of the Hal gene cluster.

In light of the biosynthetic capacity of the *hal* pathway and having established the genes necessary for the production of hapalosin and its analogs, the exact sequence of substrate incorporation within the complete hapalosin pathway was assigned. Evidence of substrate promiscuity within the hapalosin biosynthetic pathway expressed in *E. coli* consequently prompted the investigation into whether other hapalosin analogs could be discovered.

## 3.4 COMBINATORIAL 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. The presence of these hypothetical hapalosin analogs were then tested from the *E. coli* expression extracts via HR-MS$^2$.

The discovery of hapalosin D had demonstrated a capacity of the hapalosin intermediate to elude methylation by the *C*-methyltransferase of HalB, provoking the question of whether the *N*-methyltransferase of HalD could also allow hapalosin intermediates to proceed through the biosynthetic pathway to form the non-*N*-methylated products. The absence of the *N*-methyl group of hapalosin A adopts the minor *s-trans* amide hapalosin conformer[74], exhibiting reduced MDR reversing efficacy[75]. Careful examination of *E.coli* expression extracts did yield detectable amounts of the non-*N*-methylated hapalosin compound, confirming that both methyltransferases present within the pathway do not proceed with completion.

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. With this rationale, *E. coli* expression extracts 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.[76]

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 abundance of C16 saturated fatty acids within *E. coli* cells.[77] 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 binding pocket of the HalA fatty acid ligase.

The final domain of the Hal gene cluster consists of the thioesterase 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. Six different linear hapalosin products were discovered and verified by HR-MS$^2$.

The examples above demonstrate how structures can be hypothesized to exist based on enzymatic logic and substrate availability; the presence of hypothetical structures could then be verified by comparing HR-MS$^2$ data to the fingerprint fragmentation pattern established for hapalosin A. Additionally, hapalosin compounds displaying multiple substitutions − including

varying acyl chain length, both *N*- and *C*-methylation – were detected by HR-MS$^2$. A summary of the discovered substitutions is shown in **Figure 3.4.1**.



R$_1$ = H, C$_2$H$_5$, or C$_4$H$_9$
R$_2$ = CH$_3$, R$_3$ = H; or R$_2$ = R$_3$ = CH$_3$; R$_2$ = R$_3$ = H
R$_4$ = H or CH$_3$
R$_5$ = H or CH$_3$

Theoretical library of 72 combinations
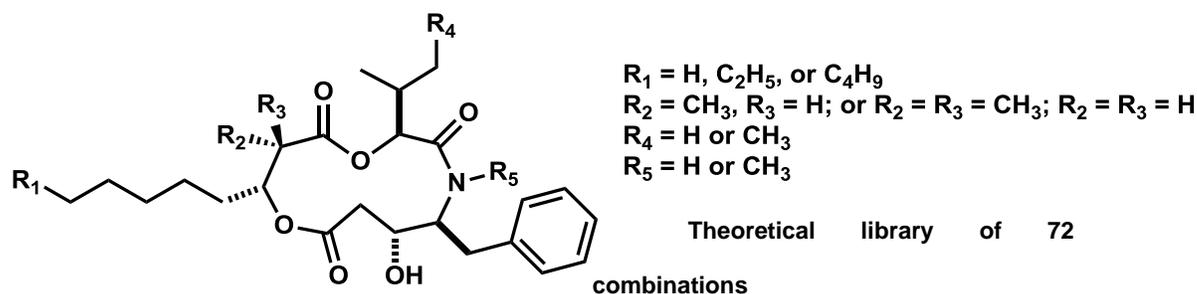
**Figure 3.4.1** Summary of hapalosin analogs detected

The theoretical combinatorial library size based on observed substitutions from the hapalosin biosynthetic pathway, including the linear analogs, was calculated to be 72: 3 acyl chain moieties (HalA), 3 malonyl substituents (HalB), 2 α-hydroxy acid substrates (HalC), 2 variations of methylation or non-methylation of the amide nitrogen (HalD), and 2 release product variants macrocyclic or linear (HalE). It should be noted, however, that only 15 have been verified by HR-MS$^2$ and the likely reason why others have not yet been detected is due to the low yields and/or suboptimal biosynthetic conditions. Also note that substitutions of the phenylalanine moiety to include instead, tyrosine or tryptophan, were not detected, suggesting an adenylation domain of HalD that is highly specific.

Inspired by the ability of HalA fatty acid ligase to accept longer and shorter acyl chain substrates for the biosynthesis of hapalosin, *E. coli* feeding experiments were proposed to introduce an azide group to the fatty acid chain of hapalosin. Despite prior knowledge of the high specificity of the adenylation domain of HalD, the incorporation of various substrates in place of the natural phenylalanine residue of hapalosin would enable the expansion of a combinatorial library of hapalosin analogs. In this effort, different halogen-substituted phenylalanine substrates were introduced to *E. coli* expression cultures and analyzed for non-natural substrate incorporation.

To this end, 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. That is, 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 this knowledge, 6-azidohexanoic, 7-azidoheptanoic, and 8-azidooctanoic acids were synthesized and used as feeding substrates to *E. coli* expression cultures (**Figure 3.4.2**).

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 concentration of substrate was doubled in the culture medium, compared to that of the other azido-alkanoic acid substrates.

This series of unnatural fatty acid feeding experiments therefore serves as 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 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 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. Although these azido-alkanoic acid feeding experiments did not explore shorter chain length substrates, it is presumed that 6 atom-long fatty acid substrates approaches the lower limit for chain length size, since the C6 hapalosin (hapalosin C) was measureable by LC-MS but comparably less than that of hapalosin A.
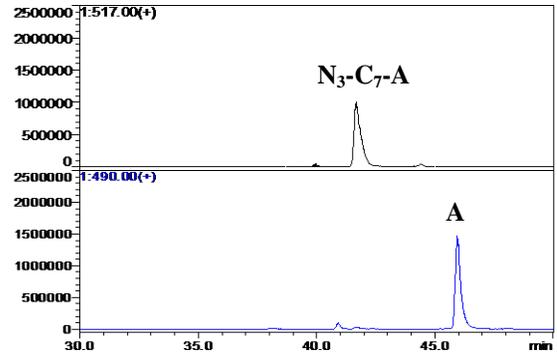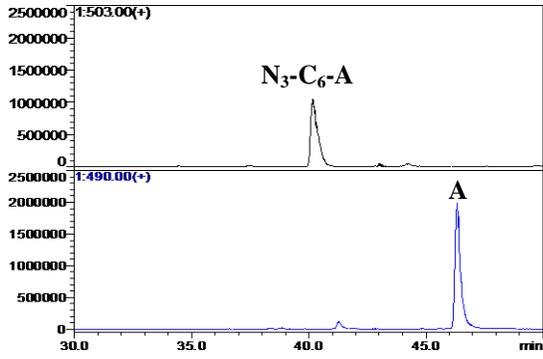
*6-azidohexanoic acid feeding*

*7-azidoheptanoic acid feeding*

*8-azidooctanoic acid feeding*
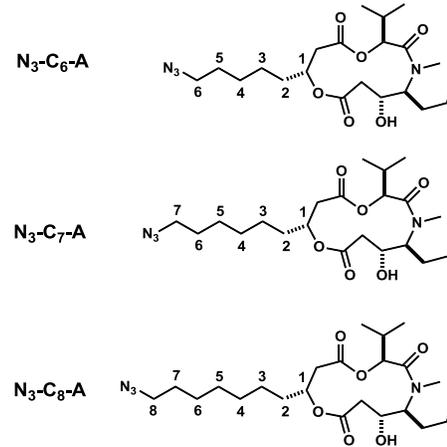
**Figure 3.4.2** Azido-alkanoic acid feeding experiments in *E. coli* expression cultures

Previous attempts to detect natural analogs of hapalosin bearing substitutions of the phenylalanine residue with tyrosine and tryptophan were unsuccessful and were attributed to high substrate selectivity of the adenylation domain of HalC. 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 3.4.3**).

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 pathway to generate the molecule of interest, it is important to note that the yields of such halogenated hapalosin structures were considerably lower than that of the azide-modified hapalosin structures and only trace amounts were detected by HR-MS$^2$. On the other hand, further optimization of halogenated phenylalanine substrate incorporation was largely left unexplored. Ultimately, 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 and from this platform alone, a total of 29 hapalosin analogs were characterized by HR-MS$^2$ from this single biosynthetic gene cluster, testifying to its ability to generate structural diversity from a rather modest number of simple substrates.

**Figure 3.4.3** HR-MS[2] analysis and fragmentation assignments of the fluorinated hapalosin analog.

## 3.5     CONCLUSIONS

The present study underscores the potential of the hapalosin gene cluster as a platform for generating a library of hapalosin analogs via precursor-directed biosynthesis[78] for this established multidrug-resistance reversing compound. Further, a series of feeding experiments involving azide-conjugated fatty acid substrates provides insight into the chain length limitation of the fatty acid ligase domain of HalA of the pathway. Indeed, promiscuity of the Hal pathway appears to be a theme: the HalB enzyme is capable of interacting with a non-native octanoyltransferase presumed to be LipB of the lipoic acid biosynthetic pathway to produce hapalosin; and, five points within the pathway introduce diversity to produce analogs of hapalosin.

A major direction forward in this project will involve the isolation of azide-conjugated hapalosin compounds. Although the MDR activity of hapalosin has been well established *in vitro*[57], direct evidence of its interaction with P-glycoprotein is lacking. Utilization of an azide-modified hapalosin as a chemical reporter could provide detailed insight into the interactions of this molecule with P-glycoprotein and, potentially, other proteins involved in the MDR phenotype of cancer cell lines. Such data could facilitate detailed analysis into structure-activity relationships as well as sites and mechanisms of action for hapalosin.

**EXPERIMENTAL METHODS USED TO STUDY THE BIOGENESIS, FUNCTION AND STRUCTURE OF A CRYPTIC ACINETOBACTER TOXIN**

## A.1    METHODS AND MATERIALS

### A.1.1   Plasmids and primers used in this study

**Table 2** Plasmids used in this study

| Plasmid | Major features | Reference |
|---------|----------------|-----------|
| pMQ124 | $_{ori}$ColE1/pRO1600, *aacC1*, P$_{BAD}$-*lacZα, CEN6* | [28] |
| pXLAB1 | pMQ124[28] derivative containing *ori* from *A. calcoaceticus* | This work, [37] |
| 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 3** 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 | TATCAGACCGCTTCTGCGTTCTGATTTAATCTGTA TCATTACTTTTGACACCAGACCAACTGG | 3' primer to clone full length *lacZ* |
| abaRFW | TACTGTTTCTCCATACCCGTAGGAGGAAAAAATG GAAAGTTGGCAAGAAGATTTATTATC | 5' primer to clone *abaR* |
| abaRRV | CAAATTCTGTTTTATCAGACCGCTTCTGCGTTCTG ATCTACAAAAGCCCTAGCATTAC | 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 |

### 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[51]. A. bamannii 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.DNA manipulations

### A.1.3 DNA manipulations

DNA polymerase kit for PCR was purchased from KAPA Biosystems (Boston, M.A., U.S.). All restriction enzymes used for cloning and restriction digest analysis were purchased from New England Biolabs (Ipswich, M.A., U.S.). Manufacturer recommended parameters were used for these experiments.

### A.1.4 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.[28] 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 a highly transformable yeast strain VL6-48N[52]. 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.1.5  Mutant generation

The protocol used for kanamycin disruption mutant generation was adapted from a previously reported procedure.[27] 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.[53] The suicide plasmid vector to delete the catA gene was constructed using a 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.1.6 Construction of A. baumannii Expression Plasmids

The A. baumannii-E. coli-S. cerevisiae shuttle vector and derivatives thereof were constructed via yeast recombineering.[28] Primers WH1266ORIFW and WH1266ORIRV were used to clone the origin of replication sequence from an A. baumannii replicable plasmid pWH1266[37]. 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[54]. 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. catAFWand 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 PBAD 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.1.7 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 Na2HPO4, 40mM NaH2PO4-H2O, 10mM KCl, 1mM MgSO4-7H2O, 0.0028% (v/v) β-mercaptoethanol, pH 7.0 per 500mL H2O), 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 for at room temperature (30 min). After incubation, Na2CO3 (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. LacZ activity (A420nm) was normalized to the OD600 measured from the A. baumannii overnight liquid culture.

### A.1.8 Motility assays

The procedure for executing motility assays was adapted from a previously reported protocol.[30] 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 C12-HSL (250nM) were supplemented. Plates were inoculated by dipping a 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.1.9 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.1.10 Desiccation resistance

This assay was adapted from a previously reported procedure.[32] Overnight liquid cultures of A. baumannii were diluted 100-fold and allowed to grow to OD600 = 0.5. This liquid culture was diluted further 10- to 100-fold 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 3, 6, or 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.1.11 Wax Worm G. mellonella Cytotoxicity

A. baumannii virulence against the wax worm G. mellonella was executed using a procedure previously reported.[31] 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.1.12 Heat shock viability

A previously reported procedure was adapted for these experiments.[32] Overnight liquid culture of A. baumannii strains were diluted 100-fold in LB media and allowed to grow to OD600 = 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.1.13 Reverse Transcriptase (RT) and Quantitative (q) PCR Analysis

cDNA libraries were generated using standard kits and techniques available from Promega and Qiagen. Primers for RT-PCR analysis were designed to amplify an approximately 500bp internal sequence of each gene. qPCR analysis proceeded by amplifying a 500bp internal sequence of catA using fluorogenic ddCT substrates.

**EXPERIMENTAL METHODS TO STUDY HAPALOSIN BIOSYNTHESIS IN HAPALOSIPHON WELWITSCHII AND ITS COMBINATORIAL GENERATION IN ESCHERICHIA COLI**

## B.1    METHODS AND MATERIALS

### B.1.1    Plasmids and primers used in this study

To reconstruct the entire hapalosin biosynthetic pathway, genes halA and sfp were cloned by PCR and yeast recombineered into parent vector pMQ131[67] 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. 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 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. 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 B1.

**Table 4** 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 |
| Hap1FW | TTCTCCATACCCGTAGGAGGATATAAATGAACAACAATTTATTTTCAG | 5' primer to clone *halA* |
| Hap1RV | CTGTTTTATCAGACCGCTTCTGCGTTCTGATGGGCCCTTAAACACTAACTGAAGAGG | 3' primer to clone *halA* |
| Hap2FW | CGAATTCGAGCTCGGTACCCGGGGAAGGAGATATACATATGAGCAATATCTCTAAAAAAACC | 5' primer to clone fragment 1 |
| Hap2RV | TTCTTCTTCGGCGGCTACACTAGAG | 3' primer to clone fragment 1 |
| Hap3FW | GCGTTGAGTATTTGTTGCAGG | 5' primer to clone fragment 2 |
| Hap3RV | CCTTGTTCTAGTAGCTTCGAG | 3' primer to clone fragment 2 |
| Hap4FW | GTCAATTGATGCAGCAATTGC | 5' primer to clone fragment 3 |
| Hap4RV | TATCAGACCGCTTCTGCGTTCTGATTTAATCTGTATCATCAATCATCTGCCTGTGACCG | 3' primer to clone fragment 3 |
| Hap5FW | GAATTGGATCCTCTAGATTCTCCATACAGGAGGAATAATATGAATAGAGAACCAATCGCC | 5' primer to clone fragment 4 |
| Hap5RV | CTTGTGCTAAATTTGACCCAC | 3' primer to clone fragment 4 |
| Hap6FW | CACTCGCGGCTAATGCATCGTC | 5' primer to clone fragment 5 |
| Hap7RV | GTTTGGGGTGGGAGTGCTACC | 3' primer to clone fragment 5 |

**Table 5** Plasmids used in this study

| Plasmid | Description | Reference |
|---------|-------------|-----------|
| pMQ131 | $_{ori}$pBBR1, *aphA-3*, $P_{lac}$-*lacZa*, *oriT*, *URA3*, *CEN6/ARSH4* | [67] |
| pMQ123i | $_{ori}$ColE1/RK2, *aacC1*, $P_{tac}$-*gfpmut3*, *oriT*, *URA3*, *2μ* | [67] |
| pXL1 | pMQ131 with *sfp* | This study |
| pXL2 | pXL1 with *hapA* | This study |
| pXL3 | pMQ123i with *hapD* and *hapE* | This study |
| pXL4 | pXL3 with *hapB* and *hapC* | This study |

### B.1.2 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 (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 Na2SO4 and the organic solvent was evaporated in vacuo under gentle heating (37°C).

### B.1.3 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 (21.1×103 g, 5min). The organic supernatant was then

used for subsequent liquid chromatography-mass spectrometric (LC-MS) analysis. 2µL of MeOH-dissolved supernatant extract was separated by reverse phase high performance liquid chromatography, R-HPLC, (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.

### B.1.4  High resolution tandem-MS analysis of hapalosin analog production in *E. coli*

The same MeOH-soluble extract was subjected to LC-HR-MS2 analysis. LC separation proceeded by R-HPLC (Phenomenex Jupiter® C18, 370Å, 3µm, 2×150mm) using a Dionex® Ultimate3000 HPLC and analyzed on a ThermoScientific Q-Exactive HR-MS instrument using ESI positive ionization mode (35.0V collision energy). 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[79],[80] 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.

### B.1.5  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 (21.1×103 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 ($21.1 \times 10^3$ g, 5min), and the organic supernatant was used for preparative HPLC (Dionex® Ultimate3000 HPLC system with variable wavelength detector) using a reverse phase column (Phenomenex® LUNA C18(2), 100Å, 10µm, 250×21.20mm diameter). 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.

## B.1.6  Cultivation of *Hapalosiphon welwitschii* UTEX 1830 and metabolite analysis

UTEX B1830 was cultured as described previously.[57] 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°) to

yield the dried crude extract. The crude extract was redissolved in MeOH and analyzed via LC-MS and LC-HR-MS2 using the same protocol as described above for hapalosin obtained from E. coli.

### B.1.7  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.

### B.1.8  Azido-alkanoic acid synthesis

6-bromohexanoic and 8-bromooctanoic acids were both obtained from Sigma. 7-bromoheptanoic acid was obtained from Matrix Scientific. The procedure to synthesize azido-alkanoic acid substrates for feeding experiments was adapted from a protocol previously reported.[81] 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.

### B.1.9  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.

# BIBLIOGRAPHY

1.     Vining, L.C., Functions of Secondary Metabolites. Annual Review of Microbiology, 1990. 44(1): p. 395-427.

2.     Newman, D.J. and G.M. Cragg, Natural Products as Sources of New Drugs over the Last 25 Years⊥. Journal of Natural Products, 2007. 70(3): p. 461-477.

3.     Cyclosporine: A Review. Journal of Transplantation, 2012. 2012.

4.     Berdy, J., Bioactive Microbial Metabolites. J Antibiot, 2005. 58(1): p. 1-26.

5.     Fuqua, C. and E.P. Greenberg, Listening in on bacteria: acyl-homoserine lactone signalling. Nat Rev Mol Cell Biol, 2002. 3(9): p. 685-695.

6.     Lamont, I.L., et al., Siderophore-mediated signaling regulates virulence factor production in Pseudomonas aeruginosa. Proceedings of the National Academy of Sciences, 2002. 99(10): p. 7072-7077.

7.     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.

8.     Donadio, S., et al., Modular organization of genes required for complex polyketide biosynthesis. Science, 1991. 252(5006): p. 675-679.

9.     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.

10.    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.

11.    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.

12.     Metzker, M.L., Sequencing technologies [mdash] the next generation. Nat Rev Genet, 2010. 11(1): p. 31-46.

13.     Fraser-Liggett, C.M., Insights on biology and evolution from microbial genome sequencing. Genome Research, 2005. 15(12): p. 1603-1610.

14.     Davies, J. and K.S. Ryan, Introducing the Parvome: Bioactive Compounds in the Microbial World. ACS Chemical Biology, 2011. 7(2): p. 252-259.

15.     Zerikly, M. and G.L. Challis, Strategies for the Discovery of New Natural Products by Genome Mining. ChemBioChem, 2009. 10(4): p. 625-633.

16.     Bentley, S.D., et al., Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature, 2002. 417(6885): p. 141-147.

17.     Gerwick, William H. and Bradley S. Moore, Lessons from the Past and Charting the Future of Marine Natural Products Drug Discovery and Chemical Biology. Chemistry & Biology, 2012. 19(1): p. 85-98.

18.     Feling, R.H., et al., Salinosporamide A: A Highly Cytotoxic Proteasome Inhibitor from a Novel Microbial Source, a Marine Bacterium of the New Genus Salinospora. Angewandte Chemie International Edition, 2003. 42(3): p. 355-357.

19.     Kale, A.J., et al., Bacterial Self-Resistance to the Natural Proteasome Inhibitor Salinosporamide A. ACS Chemical Biology, 2011. 6(11): p. 1257-1264.

20.     Anderson, R.M., The pandemic of antibiotic resistance. Nat Med, 1999. 5(2): p. 147-149.

21.     Perez, F., et al., Global Challenge of Multidrug-Resistant Acinetobacter baumannii. Antimicrobial Agents and Chemotherapy, 2007. 51(10): p. 3471-3484.

22.     Peleg, A.Y., H. Seifert, and D.L. Paterson, Acinetobacter baumannii: Emergence of a Successful Pathogen. Clinical Microbiology Reviews, 2008. 21(3): p. 538-582.

23.     Davies, J. and D. Davies, Origins and Evolution of Antibiotic Resistance. Microbiology and Molecular Biology Reviews, 2010. 74(3): p. 417-433.

24.     Rasko, D.A. and V. Sperandio, Anti-virulence strategies to combat bacteria-mediated disease. Nat Rev Drug Discov, 2010. 9(2): p. 117-128.

25.     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.

26.     Niu, C., et al., Isolation and Characterization of an Autoinducer Synthase from Acinetobacter baumannii. Journal of Bacteriology, 2008. 190(9): p. 3386-3392.

27.	Aranda, J., et al., A rapid and simple method for constructing stable mutants of Acinetobacter baumannii. BMC Microbiology, 2010. 10(1): p. 279.

28.	Shanks, R.M.Q., et al., New yeast recombineering tools for bacteria. Plasmid, 2009. 62(2): p. 88-97.

29.	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.

30.	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.

31.	Antunes, L.C.S., et al., Deciphering the Multifactorial Nature of <italic>Acinetobacter baumannii</italic> Pathogenicity. PLoS ONE, 2011. 6(8): p. e22674.

32.	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.

33.	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.

34.	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.

35.	Yildiz, F.H. and G.K. Schoolnik, Vibrio cholerae O1 El Tor: Identification of a gene cluster required for the rugose colony type, exopolysaccharide production, chlorine resistance, and biofilm formation. Proceedings of the National Academy of Sciences, 1999. 96(7): p. 4028-4033.

36.	Yura, T., H. Nagai, and H. Mori, Regulation of the Heat-Shock Response in Bacteria. Annual Review of Microbiology, 1993. 47(1): p. 321-350.

37.	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.

38.	Choi, A.H., et al., The pgaABCD locus of Acinetobacter baumannii encodes the production of poly-beta-1-6-N-acetylglucosamine, which is critical for biofilm formation. Journal of Bacteriology, 2009. 191(19): p. 5953-5963.

39.	Landick, R., J. Carey, and C. Yanofsky, Translation activates the paused transcription complex and restores transcription of the trp operon leader region. Proceedings of the National Academy of Sciences, 1985. 82(14): p. 4663-4667.

40. Oppenheim, D.S. and C. Yanofsky, TRANSLATIONAL COUPLING DURING EXPRESSION OF THE TRYPTOPHAN OPERON OF ESCHERICHIA COLI. Genetics, 1980. 95(4): p. 785-795.

41. Adhin, M.R. and J. van Duin, Scanning model for translational reinitiation in eubacteria. Journal of Molecular Biology, 1990. 213(4): p. 811-818.

42. Schümperli, D., et al., Translational coupling at an intercistronic boundary of the Escherichia coli galactose operon. Cell, 1982. 30(3): p. 865-871.

43. Rex, G., et al., The mechanism of translational coupling in Escherichia coli. Higher order structure in the atpHA mRNA acts as a conformational switch regulating the access of de novo initiating ribosomes. Journal of Biological Chemistry, 1994. 269(27): p. 18118-18127.

44. Baughman, G. and M. Nomura, Localization of the target site for translational regulation of the L11 operon and direct evidence for translational coupling in Escherichia coli. Cell, 1983. 34(3): p. 979-988.

45. Engebrecht, J., K. Nealson, and M. Silverman, Bacterial bioluminescence: Isolation and genetic analysis of functions from Vibrio fischeri. Cell, 1983. 32(3): p. 773-781.

46. Engebrecht, J. and M. Silverman, Identification of genes and gene products necessary for bacterial bioluminescence. Proceedings of the National Academy of Sciences, 1984. 81(13): p. 4154-4158.

47. Kaplan, H.B. and E.P. Greenberg, Diffusion of autoinducer is involved in regulation of the Vibrio fischeri luminescence system. Journal of Bacteriology, 1985. 163(3): p. 1210-1214.

48. de Kievit, T.R. and B.H. Iglewski, Bacterial Quorum Sensing in Pathogenic Relationships. Infection and Immunity, 2000. 68(9): p. 4839-4849.

49. Rumbaugh, K.P., et al., Contribution of Quorum Sensing to the Virulence ofPseudomonas aeruginosa in Burn Wound Infections. Infection and Immunity, 1999. 67(11): p. 5854-5862.

50. Tang, H.B., et al., Contribution of specific Pseudomonas aeruginosa virulence factors to pathogenesis of pneumonia in a neonatal mouse model of infection. Infection and Immunity, 1996. 64(1): p. 37-43.

51. Kouprina, N. and V. Larionov, Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast Saccharomyces cerevisiae. Nature Protocols, 2008. 3(3): p. 371+.

52.     Kouprina, N. and V. Larionov, Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast Saccharomyces cerevisiae. Nat. Protocols, 2008. 3(3): p. 371-377.

53.     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.

54.     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.

55.     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.

56.     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.

57.     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.

58.     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.

59.     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.

60.     Wagner, B., et al., Total synthesis and conformational studies of hapalosin, N-desmethylhapalosin and 8-Deoxyhapalosin. Bioorganic & Medicinal Chemistry, 1999. 7(5): p. 737-747.

61.     Hermann, C., et al., Total Synthesis of Hapalosin and Two Ring Expanded Analogs. Tetrahedron, 2000. 56(43): p. 8461-8471.

62.     Ghosh, A.K., et al., A Convergent, Enantioselective Total Synthesis of Hapalosin: A Drug with Multidrug-Resistance Reversing Activity. Angewandte Chemie International Edition in English, 1996. 35(1): p. 74-76.

63.     Dai, C.-F., et al., Diversity-Oriented Asymmetric Synthesis of Hapalosin: Construction of Three Small C9/C4/C3-Modified Hapalosin Analogue Libraries. Journal of Combinatorial Chemistry, 2007. 9(3): p. 386-394.

64. Dinh, T.Q. and R.W. Armstrong, A Convergent Total Synthesis of the Multidrug Resistance-Reversing Agent Hapalosin. The Journal of Organic Chemistry, 1995. 60(25): p. 8118-8119.

65. 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.

66. 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.

67. Shanks, R.M.Q., et al., New yeast recombineering tools for bacteria. Plasmid, 2009. 62(2): p. 88-97.

68. Ma, H., et al., Plasmid construction by homologous recombination in yeast. Gene, 1987. 58(2–3): p. 201-216.

69. Palomo, C., et al., A Practical Total Synthesis of Hapalosin, a 12-Membered Cyclic Depsipeptide with Multidrug Resistance-Reversing Activity, by Employing Improved Segment Coupling and Macrolactonization†. The Journal of Organic Chemistry, 2004. 69(12): p. 4126-4134.

70. 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.

71. 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.

72. 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.

73. 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.

74. Dinh, T.Q., X. Du, and R.W. Armstrong, Synthesis and Conformational Analysis of the Multidrug Resistance-Reversing Agent Hapalosin and Its Non-N-methyl Analog. The Journal of Organic Chemistry, 1996. 61(19): p. 6606-6616.

75.     Dinh, T.Q., C.D. Smith, and R.W. Armstrong, Analogs Incorporating trans-4-Hydroxy-l-proline That Reverse Multidrug Resistance Better Than Hapalosin. The Journal of Organic Chemistry, 1997. 62(4): p. 790-791.

76.     Khosla, C., et al., TOLERANCE AND SPECIFICITY OF POLYKETIDE SYNTHASES. Annual Review of Biochemistry, 1999. 68(1): p. 219-253.

77.     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.

78.     Thiericke, R. and J. Rohr, Biological variation of microbial metabolites by precursor-directed biosynthesis. Natural Product Reports, 1993. 10(3): p. 265-289.

79.     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.

80.     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.

81.     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.