# Biosynthesis of isonitrile lipopeptides by conserved nonribosomal peptide synthetase gene clusters in Actinobacteria

By

Nicholas Charles Harris

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Microbiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Wenjun Zhang, Chair
Professor Matthew Traxler
Professor Daniel K. Nomura

Spring 2019

Abstract

Biosynthesis of isonitrile lipopeptides by conserved nonribosomal peptide synthetase gene clusters in Actinobacteria
by

Nicholas Charles Harris

Doctor of Philosophy in Microbiology

University of California, Berkeley

Professor Wenjun Zhang, Chair

*Mycobacterium tuberculosis* is the leading causative agent of tuberculosis, from which millions die annually. A putative lipopeptide biosynthetic gene cluster has been shown to be essential for the survival of this pathogen in hosts, and homologous gene clusters have also been found in all pathogenic mycobacteria and other species of Actinobacteria. However, the specific functions of the encoded proteins have been elusive. Using both *in vivo* heterologous reconstitution and *in vitro* biochemical analyses, we have revealed that the five encoded biosynthetic enzymes are capable of synthesizing a family of isonitrile lipopeptides (INLPs) through a thio-template mechanism. The biosynthesis features the generation of isonitrile from a single precursor Gly promoted by a thioesterase and a nonheme iron(II)-dependent oxidase homolog and the acylation of both amino groups of Lys by the same isonitrile acyl chain facilitated by a single condensation domain of a nonribosomal peptide synthetase. Additionally, the biosynthesis of this bioactive isonitrile moiety is extraordinary and is distinct from the IsnA family of isonitrile synthases. We herein provide the first structural and biochemical evidence of an alternative mechanism for isonitrile formation.

# Table of contents

# List of figures

# List of tables

# Acknowledgments

I would like to first thank my advisor, Wenjun Zhang, for the incredible amount of support she has provided me over the past 5 years. Beyond teaching me how to think like a scientist, she taught me perseverance, patience, and how to move quickly.

Thanks to the rest of the Zhang Lab (current and past members) for being in it with me and going through this process together. Thank god for VP. And of course my undergraduate assistants Ryan Khalaf, Joelle Martin, Jordan Downey, and Brianna Larsen—we made a good team.

Beyond scientific gratitude, I would like to thank my wife, Jenny Harris, for unwavering support even during the times that I was repeating an *in vitro* reaction for the nth time. I don't think I could have done it without having Jenny's love and support.

I would like to thank my parents for being the best role models in the world. If everyone lived with as much adventure, energy, and generosity, there would be no problems in this world.

And lastly, I would like to thank Duende for showing me the way. I am grateful for the way life has flowed.

# Chapter 1. Introduction

## 1.1 Natural products in human health—drugs and disease

Bacteria often live in harsh, dynamic environments and have developed complex mechanisms for survival, including the biosynthesis of a plethora of bioactive molecules that are produced in response to environmental cues. These small-molecule secondary metabolites (SSMs) are produced as chemical weapons to combat competing organisms or as communication signals to control complex processes such as virulence, morphological differentiation, biofilm formation, and metal acquisition[1–3]. SSMs are often large complex structures or contain fascinating bioactive chemical moieties. Studying SSMs and their biosynthesis provides insight into the chemical logic used in nature to produce complex molecules that often have properties that are of value to humans. For instance, the widely used antibiotic Penicillin is produced by *Penicillium sp.*[4]*;* vancomycin, produced by the soil bacterium, *Amycolatopsis orientalis,* is used to treat *Clostrium difficile* infections;[5] and one of the most commonly prescribed pharmaceutical drugs, lovastatin, is produced by many species of fungi[6] (**Fig 1-1**). Moreover, studying the biosynthesis of SSMs provides important information about how the native producer interacts with its environment, which is a necessity for understanding the mechanisms underlying processes such as pathogenesis. By deciphering the chemical nature of a pathogen, we can better develop drugs to interfere with the chemical strategies utilized during the progression of an infection and thus more effectively combat disease.

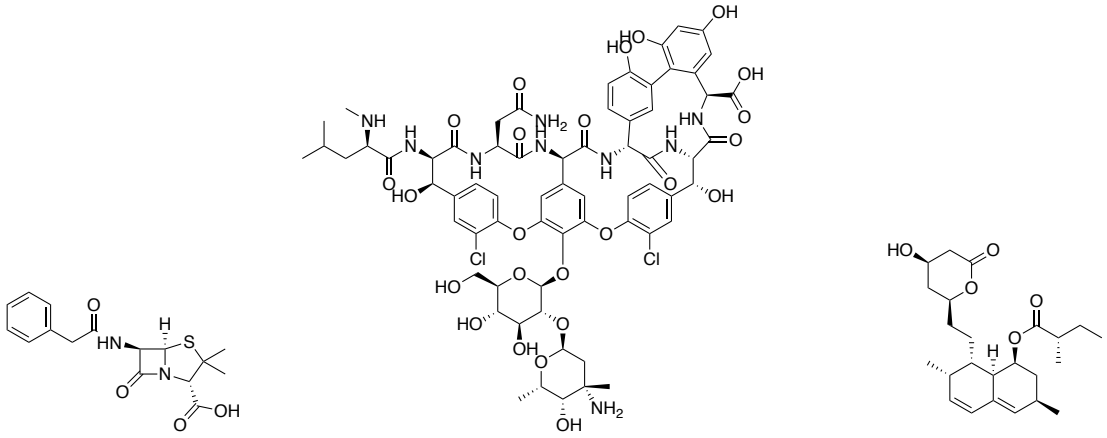



**Figure 1-1.** Examples of medicinally important natural products.

## 1.2 Motivation

The research described in this dissertation is intended to provide an in-depth view into the chemical logic used by *Mycobacteria sp.* to manipulate the host, evade the human immune response, and progress the disease state. Adding to the collective

knowledge base will hopefully move us closer to thwarting the current tuberculosis epidemic. During this research, we discovered and characterized a novel group of isonitrile-containing lipopeptides (INLPs) produced by Actinobacteria. INLP biosynthetic gene clusters are widespread and found in the genomes of many pathogenic Mycobacteria. Due to the bioactive nature of isonitrile and their common use as virulence factors, it is likely that these INLPs play a role in pathogenesis. Chapter 1 will provide an overview of the key virulence factors produced by *M. tuberculosis* and an overview of other isonitrile containing natural products, many of which serve virulence-related roles. Chapter 2 and 3 will describe the discovery of INLPs from Actinobacteria and the characterization of pathway enzymes. Since the work described in Chapter 2 and 3 was conducted using heterologous expression and *in vitro* biochemical analyses, Chapter 4 will detail work that has been conducted to identify INLPs produced during growth of the native Mycobacterial producers.

## 1.3 *Mycobacterium tuberculosis* ecology and pathogenesis

Perhaps the most relevant pathogen of the modern era is *M. tuberculosis*, the causative agent of tuberculosis. Despite years of public health initiatives, tuberculosis is now one of the top causes of death due to a single infectious agent.[7] *M. tuberculosis* latently infects 1/4 of the global population, causes more than 1.7 million deaths annually, and is the leading cause of death for individuals with HIV.[7] *M. tuberculosis* is an obligate pathogen that relies on human-to-human transmission to survive. Transmission occurs when an infected individual coughs and in doing so, transmits infectious air-borne droplets. Pulmonary infection is a hallmark aspect of *M. tuberculosis* infection and is necessary for transmission between humans.[8] Once inside the lungs, *M. tuberculosis* cells are phagocytosed by alveolar macrophages, but instead of the macrophage exercising its routine activity of phagosomal maturation, granuloma acidification and lysosomal targeting, *M. tuberculosis* controls entry into the macrophage by purposefully triggering receptors with effector molecules to encourage phagocytosis, and prematurely arrests phagosomal maturation, thus maintaining the macrophage as an ideal environment for residence.[7] The successful manipulation of the host immune response relies on an arsenal of virulence factors, many of which decorate the unique and practically impenetrable cell wall that distinguishes *M. tuberculosis* from other pathogens (**Fig 1-2**).[9]

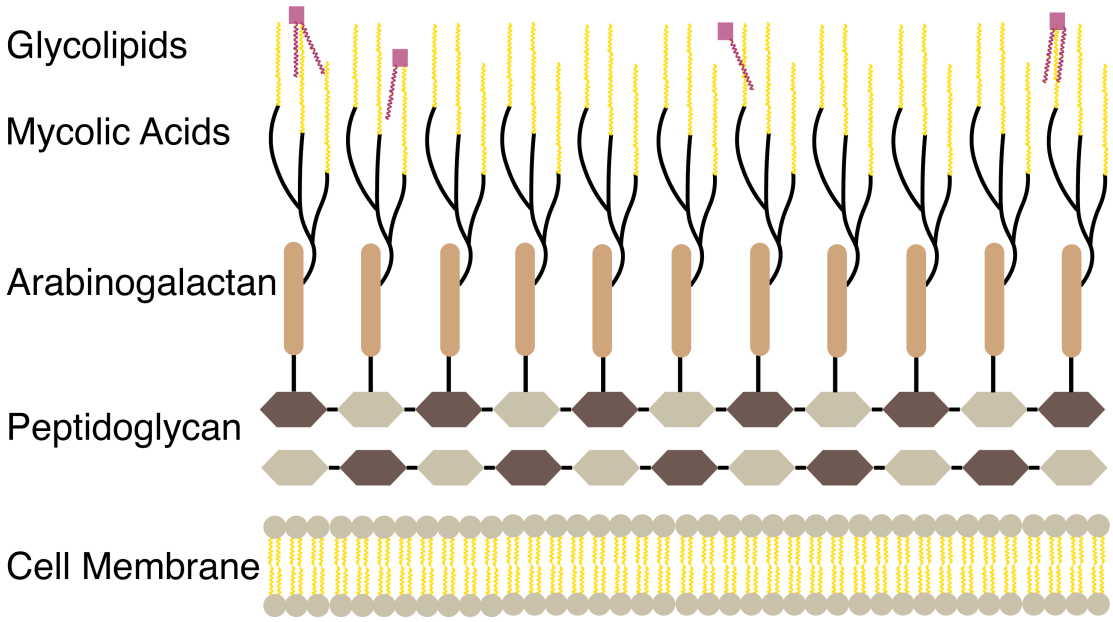



**Figure 1-2.** Mycobacterial cell wall.

The mycobacterial cell wall consists of a plasma membrane encased by a resilient peptidoglycan layer, which is further encapsulated by an arabinogalactan polysaccharide network that is threaded with phosphatidylinositol-containing lipomannans.[10] Extending outward from this heteropolysaccharide layer are long chain mycolic acids, which differ in chemical composition.[9] These 2-alkyl 3-hydroxy long-chain fatty mycolic acids are produced through the coordination of two fatty acid synthases, a type I, eukaryotic-like multienzyme complex (FAS-I) and a type II disaggregated (FAS-II).[10–12] The FAS-1 utilizes both acetyl- and malonyl-CoA substrate monomers to produce a bimodal distribution of C16-C18 and C24-C26 acids through multiple rounds of chain elongating Claisen condensations.[10] These fatty acids are released tethered to CoA, with many of the C16-C18 acyl-CoA products being utilized for the synthesis of membrane phospholipids or fed into FAS-II and the C24-C26 CoA products directed into a downstream portion of the mycolate pathway and eventually forming the α-branch of the final mycolic acid, the formation of which is mediated by the type I polyketide synthase (PKS), Pks13.[10] The mycobacterial FAS-II consists of multiple disassociated enzymes that together produce and release very long chain fatty acids, typically of around C56 in length. First the β-ketoacyl-acyl-carrier-protein synthase, FabH, catalyzes the addition of the FAS-I derived medium-chained acyl-CoAs to malonyl-ACP to form the corresponding β-ketoacyl-ACP. The FAS-II cycle of chain elongation then proceeds in a typical manner of ketone reduction by MabA, dehydration by HadAB/BC, enoyl reduction by InhA, and further chain elongation mediated by the ketosynthase KasA/B. The released long chain fatty acids are then activated by a fatty acyl amp ligase, FadD32, prior to the activity of Pks13. This PKS contains the minimal domain architecture of ketosynthase (KS), acyltransferase (AT), and thioesterase (TE), which enable the condensation of both FAS-1 and FAS-2 derived fatty acids and the release of α-alkyl β-

hydroxy mycolic acids of C86-95 in length. Additional modifications are achieved via methylation and cyclopropanation reactions and or the installment of ketone and methoxy moieties. Once mature mycolic acids are formed, they can be used as substrate monomers to form even larger and more complex molecules with distinct virulence associated bioactivities.[9,13] Mature mycolic acids are transported to the outer cell wall where they are either attached to the arabinogalactan layer via an ester linkage, or they are attached non-covalently but physically tethered to other mycolic acids and lipids in a web-like fashion.[9] Non-covalently tucked within these mycolic acids lie multiple types of lipid-derived virulence factors, as well as virulence associated cell-surface proteins, and components of signal transduction systems which enable quick changes in gene expression during the host immune response. In addition, these loosely attached mycolic acids, lipids, lipoglycans and proteins can be readily detached and released into extracellular space where they interact freely with the host's intracellular environment and modulate the host immune response. These outer membrane components play a critical role in pathogenesis.

### 1.3.1 Entry into the Macrophage

*M. tuberculosis* bacilli thrive within the cytosol of host immune cells, predominately macrophages. The bacilli tactfully encourage entry into the macrophage by orchestrating a controlled immune response, triggering receptors on the macrophage membrane with effector molecules and thereby stimulating phagocytosis.[14] Receptor-mediated phagocytosis is primarily promoted by *M. tuberculosis* derived lipoarabinomannan, phosphatidylinositol mannosides, and cell wall associated polysaccharides, which bind to the mannose receptor (MR) and the complement receptor CR3.[14,15] Activation of MR and CR3 lead to phagocytosis of the bacilli and thus entry into a promising niche, however, if not further controlled by additional virulence factors, the bacilli would be quickly killed. Therefore, the invading bacilli release various small molecules to stop phagosomal maturation and promote transport into the cytosol.[15]

During entry, the mycolic acid derived phtiocerol dimycoserosate (PDIM) is inserted into the macrophage plasma membrane. This causes a reorganization of the membrane and prevents acidification and development of the phagolysosome.[15] PDIM is therefore extremely important during the early stages of infection and is in fact essential for growth and virulence *in vivo*.[13] During progression of the disease state, PDIM-induced biophysical changes within the plasma membrane can additionally lead to autophagy. In coordination with the early secretory antigenic target, ESAT-6, a small protein (6-kDa) secreted by the type-VII secretion system, ESX-1, the membrane of the macrophage is ruptured through a currently unknown mechanism, ultimately causing cell death.[16] PDIM, although important for arresting phagosomal maturation, requires ESX/ESAT6 to cause autophagy.[17] Therefore gene expression programs during infection can change from favoring latent growth, via PDIM production alone, or toward a more deadly lytic phase in which both PDIM and ESX/ESAT6 cause coordinated events of autophagy.[16]

### 1.3.2 PDIM Biosynthesis

Malonyl and methyl malonyl CoA are used as starter units for PDIM biosynthesis and decreased availability of these substrates decreases virulence *in vivo*.[13] In addition, PDIM biosynthesis increases during *in vitro* growth when supplemented with odd-chain

fatty acids or when cultivated within host tissues. Host derived fatty acids are indeed catabolized by *M. tuberculosis* during infection and used for starter unit biosynthesis, which in turn allows for increased flux towards PDIM and other lipid-derived virulence associated small molecules.[13]

PDIM consists of two mycocerosic acids stitched to a single phthiocerol molecule via two ester linkages. Phthiocerol is biosynthesized by the multimodular non-iterative PKS, PpsA-E. Each module has a similar domain architecture consisting of a KS, AT, KR, PCP, for substrate loading, decarboxylative condensation, and reduction of newly introduced β-ketones, and PpsC and PpsD having additional dehydration (DH), and enoyl reducatse (ER) domains for hydroxyl group removal and subsequent acyl chain saturation.[18] FadD26, first activates and loads a saturated fatty acid of C12-C16 in length onto the thiolation domain of PpsA, and the chain is extended via the addition of malonyl CoA by KS domain mediated Claisen condensations. Two hydroxyl groups are introduced by the KR domains of PpsA and PpsB. A single methyl branch is installed by the AT domain of PpsD, which utilizes methyl malonyl CoA (all other AT domains on the other modules utilize malonyl CoA). Separately, mycocerosic acids are formed by mycocerosic acid synthase (MAS), an iterative PKS containing a domain architecture of KS, AT, DH, ER, KR, PCP. FadD28 activates and loads a saturated acid of C22-24 in length,[19] and the acyl chain is elongated via the incorporation of methyl malonyl CoA substrates, which leads to the incorporation of four methyl groups on the mature mycoserosic acid.[20] Flanking the *ppsA-E* gene cluster is a pthiocerol dimycocerosyl transferase, encoded by *papA5*.[21] PapA5 is structurally similar to CoA-dependent acyltransferases and mediates the diesterification of phthiocerol and the two mycocerosate molecules to form the final PDIM molecule. PapA5 catalyzes a similar diesterification of phthiodiolone with mycoserosates to form the virulence factor phthiodiolone dimycoserosate.[21]

Mature PDIM is exported to the cell surface by the transporter MmpL7, a resistance, nodulation, and cell division (RND) type permease that contains multiple transmembrane domains.[22] Domain 2 of MmpL7 physically interacts with the PDIM biosynthetic machinery acting as a scaffold to recruit Pps and MAS enzymes and facilitates the synthesis and transport of PDIM from the cytosol and across the cytoplasmic membrane.[23] In coordination with the transporter DrrC, PDIM is secreted to the outer leaflet of the cell wall, where it is then available to interact with the host.[24]

### 1.3.3 Metal acquisition

At the core of pathogenesis is the need to efficiently acquire iron from the host. Since many biological processes rely on iron (e.g. enzyme active sites, heme etc.), the host has developed systems for iron storage to both maintain a supply for its own processes as well as to safeguard against the invading pathogen. Preventing the pathogen from acquiring iron, as well as other essential nutrients, can prove a successful strategy for taming and starving out the invading cell. Macrophages exercise nutritional immunity by effectively transferring iron, both Fe(II) and Fe(III), to bone marrow cells, or binding free iron by transferrin, lactoferrin, and hemoglobin-haptoglobin.[25] These iron binding systems effectively maintain the intracellular iron concentrations well below the $10^{-7}$ M requirement for *M. tuberculosis* to grow efficiently.[26] Therefore it is essential for *M. tuberculosis* to produce siderophore molecules that bind Fe(III) with high affinity, to

appropriate iron from the macrophage. Without effective siderophore production, *M. tuberculosis* would quickly succumb to the immune response.

*M. tuberculosis* produces only the mycobactin type siderophore, which contain a signature salicylate derived oxazoline moiety connected to an acylated *N*-hydroxy-Lys residue by an amide bond.[27] The hydroxy-Lys is esterified at the α-carboxyl group with a β-hydroxy acid, which is bound to a seven membered lactam ring by an additional amide bond. Fe(III) is coordinated by the hydroxamic groups of the N-hydroxy-Lys, and the oxygen of the phenolate and nitrogen of the oxazoline moieties.[27] While Fe(III) is bound with high affinity, having a stability constant on the order of $10^{30}$, Fe(II) is poorly bound.[28] Production of mycobactins is essential for virulence as well as growth in iron replete media. *M. tuberculosis* produces two forms of mycobactin that differ in the *N*-acyl chain extending from the central hydroxylated Lys. The two structures differ in their aqueous solubility and it is thought that the more soluble product, containing a shorter fatty acyl chain and either a methyl ester or carboxylate at its terminus, is secreted into the aqueous environment of the macrophage cytosol to actively bind free iron as well as iron-containing proteins.[27] The more lipophilic mycobactin, contains a long fatty acyl chain with a methylated branch at its terminus and is thought to serve as an ionophore to transport the acquired iron across the mycobacterial cell envelope. In this way, the two mycobactin molecules might work in unison with one being responsible for iron scavenging and acquisition and the other for transport into the cell for growth and utilization.[27]

### 1.3.4 Mycobactin biosynthesis

Mycobactin siderophores are biosynthesized by the multimodular hybrid non-ribosomal peptide synthetases/polyketide synthase (NRPS/PKS) encoded by *mbtA-F*, and co-transcribed with *mbtH-J* as a single operon.[29] The domain architecture of the pathway consists of an adenylation domain (A) of MbtA; MtbB domains consisting of thiolation (T), cyclization (Cy), A, T, and TE; MbtE domains organized as condensation (C), A, T, C, and T; MbtC containing a single KS; MbtD having AT, KR, and T domains; and MbtF having domains organized as C, A, T, and an epimerization (E) domain.[30] Substrates for the NRPS/PKS megasynthase are first synthesized by MbtG, MbtI, and MbtJ, a salicylate synthase with both isomerase and lyase activities, that catalyzes the formation of salicylate from isochorismate, which forms the aryl acid that caps one end of the mycobactin side chain. MbtG catalyzes the *N*6-hydroxylation of L-Lys to form *N*6-hydroxy-L-Lys, which serves as the substrate for the A domain of MbtF. Additionally, a portion of the produced *N*6-hydroxy-L-Lys is modified by the activity of Rv1347c, an acyltransferase, which acylates the *N*6-hydroxy-L-Lys to form *N*6-acyl-*N*6-hydroxy-L-Lys, which then serves as the substrate for the A domain of MbtE. Once the substrates have been synthesized, the megasynthase MtbA-F pieces together monomers to construct the final mycobactin products. MtbA first activates and loads salicylate onto the first thiolation domain of MbtB. The A domain of MbtB is specific for L-Thr, which is loaded onto the second thiolation domain of MbtB. The Cy domain of MbtB is thought to catalyze both the condensation of the peptidyl carrier protein bound-salicylate and L-Thr, and also mediates a cyclization to form the salicyl-methythiazolinyl-MbtB intermediate providing the oxazoline ring on the final product. Although MbtB contains a TE domain, it is unclear if the MbtB product remains tethered to the phosphopantetheine arm of the

second T domain of MbtB or if the product is released before being incorporated into the MbtE bound substrate. Meanwhile, the A domain of MbtE activates and loads *N*6-acyl-*N*6-hydroxy-L-Lys onto the first T domain of the MbtE module. The C domain of MbtE catalyzes the condensation of *N*6-acyl-*N*6-hydroxy-L-Lys-MbtE with the upstream salicyl-methythiazolinyl intermediate. Additional C and T domains exist at the N-terminus of MbtE, however their contribution to mycobactin biosynthesis is not yet understood. The AT domain of MbtD is presumed to utilize acetoacetyl-CoA and the KR domain of MbtD likely reduces the ketone to a hydroxyl group and subsequently the β-hydroxy group attacks the carbonyl of the MbtE bound intermediate. The A domain of MbtF is specific for *N*6-hydroxy-L-Lys and the C domain catalyzes the condensation of the MbtF-bound substrate and the upstream intermediate formed by the activities of the MtbA-E NRPS/PKS modules. The product is released via a second cyclization of the terminal *N*6-hydroxy-L-lysyl moiety but the domain/enzyme that catalyzes the cyclization has not yet been established.

Initial biochemical work with the Mbt NRPS/PKS modules were unsuccessful due to the encoded protein products being insoluble. A major discovery was that *mbtH* encodes a small 8 kDa protein, with no assigned catalytic function, but aids in solubilizing components of the NRPS/PKS megasynthase. Co-purification of MbtH with MbtB, E, and F allowed for an other-wise insoluble and inactive NRPS/PKS to become soluble and active *in vitro*, thus allowing for confirmation of their activities.[31] Other mbtH-like proteins have since been discovered across the bacterial kingdom and each seem to possess a similar property of assisting with solubilizing pathway enzymes.[32,33]

### 1.3.5 Chronic infection and the progression of the disease state

Once the bacilli have established themselves in their niche within the macrophage, they perpetuate a state of chronic infection by triggering the formation of granulomas.[34] These granulomas consist of complex aggregates of immune cells around the macrophage and although they are formed as a part of the immune response, they tend to benefit the bacilli by providing a thick barrier that allows for persistent growth and the eventual spread of bacteria to additional tissues. The formation of granulomas is a hallmark event of *M. tuberculosis* infection that defines latency. It is only after the bacilli have spread that the final stage of disease progression is triggered in which granulomas burst, largely through the action of PDIM and ESAT/ESX, and massive inflammation causes severe tissue damage resulting in lung cavitation and pulmonary disease. Granuloma formation during *M. tuberculosis* infection is primarily triggered through binding and recognition of the bacterial glycolipid, trehalose 6-6′ dimycolate (TDM or cord factor), by the C-type lectin receptor, Mincle.[35,36] This binding event activates the Syk-Card9 signaling pathway and ultimately leads to NFkB release, the activation of the inflammatory cytokine response and in turn, the formation of granulomas.[34] TDM has been shown to activate Syk-Card9 through the binding of Mincle both *in vitro* and *in vivo*.[35,36]

### 1.3.6 TDM Biosynthesis

After mature mycolic acids are released from Pks13, they can be used as monomers for the formation of even larger and more intricate molecules such as TDM and PDIM. TDM is formed via the esterification of trehalose and two mature mycolic acid molecules with trehalose-6-monomycolate (TMM) being an intermediate with bioactive properties

that mimic those of TDM.[37] Ester formation is promoted by the antigen 85 (Ag85) complex which includes three acyltransferase domains that provide redundancy for the formation of TDM, this redundancy is important as TDM is necessary for virulence.[38–40] In addition to triggering the formation of granulomas, TDM molecules can link together in a cord-like fashion, hence the alias, cord factor, thus creating a thick lipid barrier on the outer leaflet of the cell wall. This thick hydrophobic TDM layer protects the bacilli from hydrophilic small molecules including both antibiotics and toxic, immunoderived chemicals produced by the host.[11]

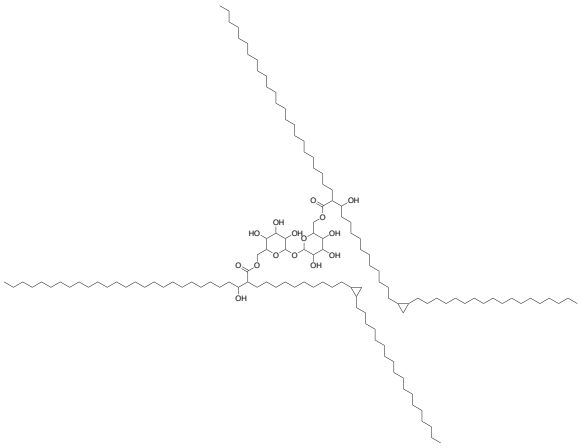


**Figure 1-3.** Structure of TDM.

## 1.4 Isonitrile: An important chemical warhead for virulence

The electron-rich functionality of the isonitrile lends itself as a biologically active warhead for naturally derived products. It acts as a carbon monoxide analog which participates in π back-bonding with metal atomic orbitals, with the C lone pair being more antibonding with respect to the CN bond and therefore allowing for coordination with metal atoms.[41] Owing to their ability to coordinate transition metals, isonitrile are often exploited for metal acquisition, detoxification, and virulence.[42–44] Indeed the resume of potent biologically active isonitrile containing natural products is vast, and many examples are detailed in this section.

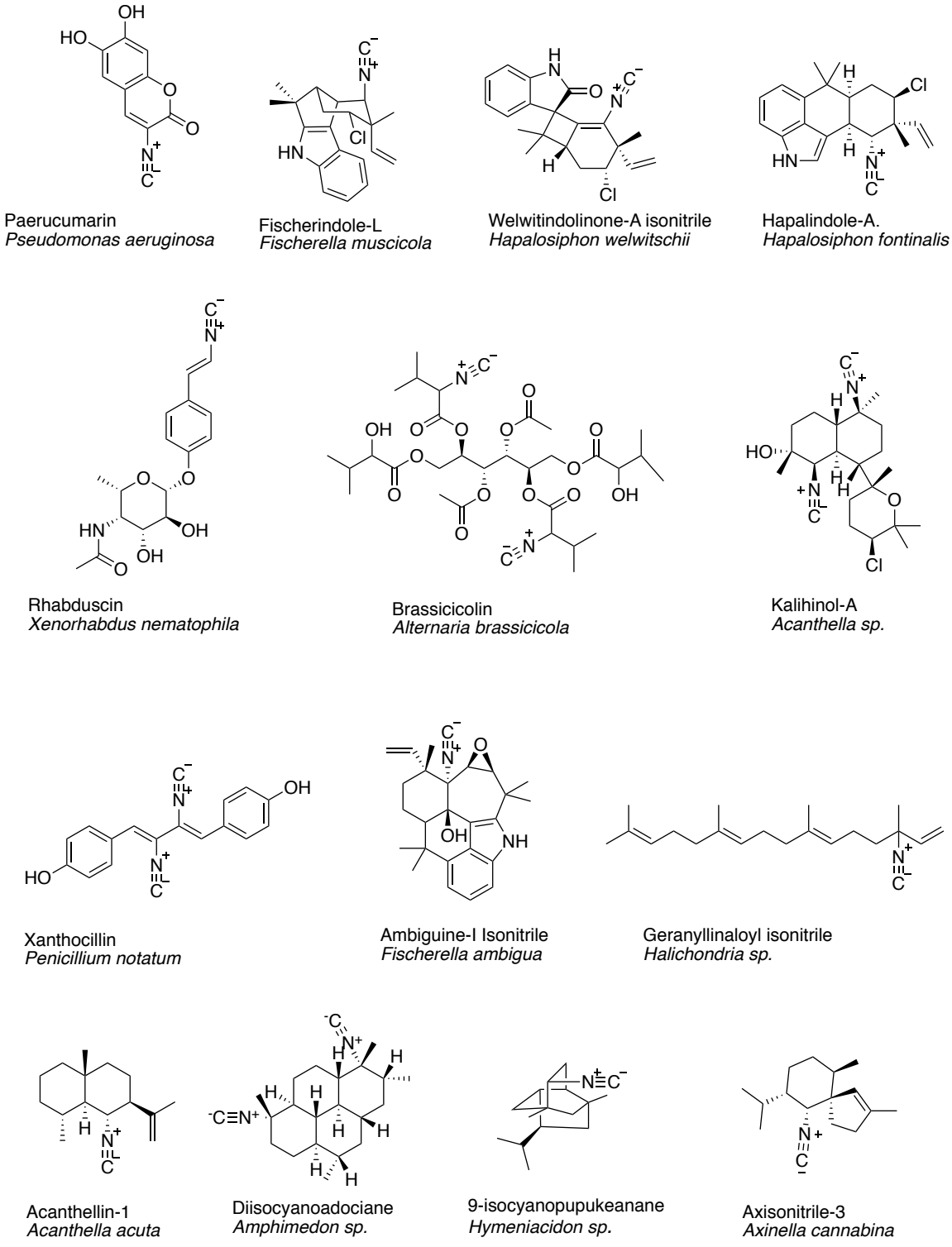



**Figure 1-4.** A small selection of isonitrile-containing natural products.[42,43,45–56]

### 1.4.1 History of isonitrile natural product research

The first isonitrile containing natural product was isolated and characterized in 1950 from *Penicillium notatum*, and named xanthocillin.[57] Despite the early discovery of naturally occurring isonitrile, it was not until 2005 that the first genes implicated in isonitrile formation were discovered and finally, 68 years after xanthocillin was isolated, its biosynthesis was elucidated.[57] The identification of the first isonitrile biosynthetic gene came to fruition by cloning DNA extracted from environmental samples (eDNA) into a cosmid library and heterologously expressing this library in *Escherichia coli*.[47] The library was screened for antibacterial activity using the classic growth inhibition assay using top agar containing *Bacillus subtilis* cells. Using a bioassay-guided small molecule purification/fractionation strategy, Brady and Clardy, isolated a potent antibiotic bearing the isonitrile moiety, a C-3 substituted indole with a trans olefin and terminal isonitrile. The producing cosmid was further genetically dissected using transposons, and two genes were identified as being responsible for the biosynthesis of the product with IsnA being directly involved in forming the isonitrile and IsnB being an non-heme iron(II) dependent α-ketoglutarate (α-KG) dependent dioxygenase that catalyzes the oxidative decarboxylation of the Trp carboxylate moiety to form a trans olefin. Feeding labeled substrates, the N of the isonitrile was confirmed to originate from the amine of Trp. Shortly after their initial study, Brady and Clardy, designed an elegant "inverse-labeling" experiment to elucidate the origin of the C of the isonitrile.[58] Brady and Clardy effectively dissected the metabolome by heterologously expressing *isnA/B* in *E. coli* knockout mutants missing genes at each major branch point of metabolism. Cultures were fed $^{13}C$ glucose to create a $^{13}C$ labeled metabolic background, and $^{12}C$ substrates were fed to chemically complement the mutants. They then looked for the final isonitrile product lacking a $^{13}C$ label and using this strategy, they were able to confirm ribulose-5-phospate as the substrate from which the C of the isonitrile originates. The exact mechanism employed by IsnA and its homologues to convert the α-amino group to isonitrile on an amino acid with ribulose-5-phosphate as a co-substrate is still unknown.

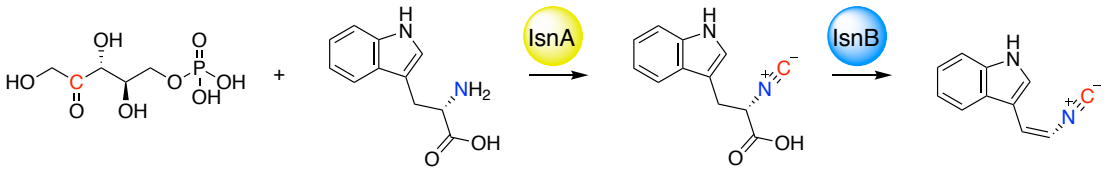



**Figure 1-5.** Formation of isonitrile by IsnA/B.

### 1.4.2 Diversity of isonitrile containing natural products

Since the initial discovery of isonitrile containing natural products and the elucidation of their biosynthesis, many other natural products containing the isonitrile moiety have been characterized. These natural products have been isolated from diverse organisms including fungi,[57,59] bacteria,[42,43] marine sponges,[60,61] nudibranch mollusks,[62] cyanobacteria,[63] and plants.[64]

### 1.4.3 Paerucumarin

The *pvcABCD* operon of the opportunistic human pathogen, *Pseudomonas aeruginosa*, encodes IsnA/B homologues and two additional enzymes required for the production of the iron-binding isonitrile-containing product, paerucumarin, aptly named for its cumarin scaffold.[42,65,66] Expression of *pvcA-D* is controlled by the PtxR transcription factor which is known to regulate important pathogenicity determinants.[67,68]

PvcA and PvcB are homologues of IsnA and IsnB, respectively, both exerting their activities on L-Tyr. PvcC and PvcD are HpaB- and HpaC-like two-component flavin adenine dinucleotide-dependent monoxygenases that oxidize phenols to dihydroxy phenols and catechols. PvcB is thought to mediate ring formation of the carboxylate moiety to the ring of L-Tyr and PvcC/D are proposed to oxidize the resulting intermediate to a catechol thus yielding the final paerucumarin product.[65] Homologous pathways exist in the plant pathogen *Erwinia carotovorum*, the insect pathogen *Photorhabdus luminescens*, the bacterial predator *Bdellovibrio bacteriovorus*, and the human pathogens *Burkholderia mallei, Legionella pneumophila*, and *Vibrio cholera*.[42]

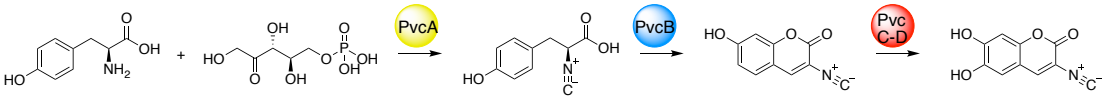



**Figure 1-6.** Formation of paerucumarin by PvcA-D.

### 1.4.4 Xanthocillins and brassicicolin

Xanthocillin is a diisonitrile chalkophore produced by *Aspergillus fumigatus*, a human pathogen, and *Penicillium sp.* and a similar organization of the Xan genes exists in many species of fungi.[57] The AceA and MacA, copper-fist transcription factors that are activated under copper-replete conditions, regulate the biosynthesis of xanthocillin. The Xan pathway utilizes L-Tyr as the α-amino N donating substrate. XanB, a non-heme iron(II) dependent α-KG dependent dioxygenase plays the analogous role of both IsnA and IsnB, forming both the isonitrile moiety and catalyzing the subsequent decarboxylation and desaturation of the L-Tyr scaffold. XanB functions twice in the pathway to covert two L-Tyr molecules to yield the intermediate (*Z*)-4-(2-isocyanovinyl)phenol molecules. XanG, a cytochrome P450 coexpressed with *xanB*, is thought to catalyze the oxidative dimerization of the two 4-(2-isocyanovinyl)phenol molecules to form 4,4'-((1*Z*,3*Z*)-2,3-diisocyanobuta-1,3-diene-1,4-diyl)diphenol. XanE is proposed to be a methyltransferase capable of modifying xanthocillin to form methylated derivatives. Ribulose-5-phosphate has not yet been confirmed as the source of C in the isonitrile moiety.

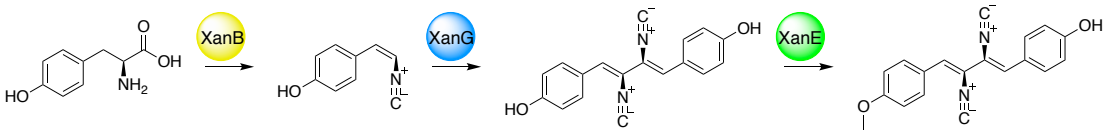



**Figure 1-7.** Formation of xanthocillin by XanB,G,E.

Brassicicolin is another fungal derived diisonitrile containing natural product produced by the plant pathogen *Alternaria brassicicola*, which causes black spot disease in *Brassica sp.*[59,69,70] The precise genes involved with brassicolin biosynthesis are not known but the formation of the isonitrile appears to be mediated by an IsnA-like mechanism with the amine of Val likely providing the N of the isonitrile and the C presumably originating from ribulose-5-phosphate.[70] The biosynthetic pathway likely utilizes mannitol as a central carbon scaffold to which four Val residues and two acetyl groups are attached via ester linkages.[70]

### 1.4.5 Rhabduscins

The insect pathogens *Xenorhabdus nematophila* and *Photorhabdus luminescens,* produce the isonitrile containing rhabduscins, to disrupt a key enzyme in the host insect's immune response system.[43] Rhabduscins are amidoglycosyl- and vinyl-isonitrile-functionalized L-Tyr derivatives. L-Tyr is functionalized by an IsnA/B mediated mechanism followed by glycosylation of the phenol by a glycosyltransferase. The glycosylated functionalized L-Tyr can be modified by the addition of an acetamide moiety extending from the sugar. Both rhabduscin molecules are located at the bacterial cell surface, as imaged using stimulated Raman scattering microcopy, and bind with high affinity to a Cu(II) containing insect phenoloxidase. Rhabduscins can inhibit this phenoloxidase at low nM concentration and thus prevent the host insect from producing melanin. Melanin is typically produced during an infection to wall-off the invading pathogen and to prevent the spread of infection.[71] However, the inhibition of the melanin phenoloxidase by rhabduscins allows for the quick proliferation and persistence of the pathogen.[72]

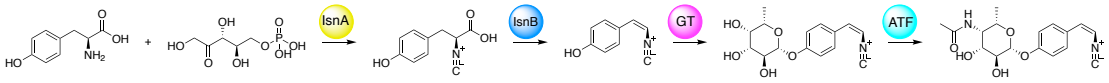



**Figure 1-8.** Formation of rhabduscins.

### 1.4.6 Ambiguines

Ambiguines are isonitrile containing hapalindole-type natural products constructed of terpenoid indole alkaloids and produced by the cyanobacteria *Fischerella sp.* and *Hapalosiphon sp.*[63] These isonitrile are formed from L-Tyr by the IsnA homologues AmbI1 and AmbI2 and the IsnB homologue, AmbI3. The functionalized indole is then fused to a single monoterpene originating from geranyl pyrophosphate by the proposed

activity of the prenyltransferase, AmbP1. From here, the product can remain unmodified as hapalindole U, further prenylated via the activity of the dimethylallyltransferase, AmbP3, or chlorinated to form hapalindole G or amibiguine A, by the activity of the non-heme, iron(II) and α-KG dependent dioxygenase, AmbO5. The resulting isonitrile products that form the hapalindole, ambiguine, and fischambiguine family of molecules have a broad range of bioactive activities including insecticidal,[73] antimicrobial,[55,74,75] and anti-cancer activities,[76] although the native biological functions have not yet been established.

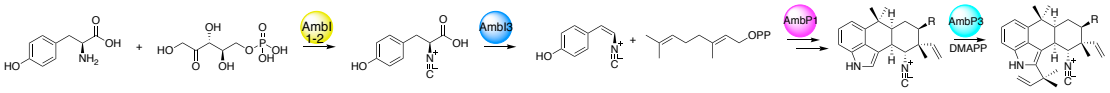



**Figure 1-9.** Formation of ambiguines.

### 1.4.7 Cyanide derived isonitrile from marine sponges

Many marine sponge derived isonitrile-containing natural products have been characterized with the vast majority being isocyanoterpenes.[60,61] Despite the large sample of natural products having been characterized, little is known about their biological activities or their biosynthesis. It is currently thought that isonitrile formation is not enzymatic but instead, free cyanide is incorporated into these marine natural products.[60,61] Multiple studies have fed labeled amino acid substrates as well as labeled acetate to sponge cultures, however the produced isonitrile products do not incorporate the labeled atoms.[77,78] However, feeding sodium-[$^{14}$C]-cyanide to the marine sponge of the genus *Amphimedon* results in the incorporation of $^{14}$C in the final diisocyaonadociane product[78] and feeding of doubly labeled [$^{13}$C, $^{15}$N]-cyanide to the sponge *Ciocalypta sp.* results in the retention of both labeled atoms in the final isonitrile-containing product.[79] These data provide support for an alternative mechanism of isonitrile natural product biosynthesis in which the CN is of abiotic origin and is incorporated into the natural product intact. It is additionally possible that sponges themselves are incapable of producing these isonitrile products altogether and instead, their bacterial symbionts may produce them.[45]

# Chapter 2. Biosynthesis of isonitrile lipopeptides by conserved nonribosomal peptide synthetase gene clusters in Actinobacteria

Parts of this chapter have been adapted from the following with permission:
Harris, N. C., Sato, M., Herman, N. A., Twigg, F., Cai, W., Liu, J., Zhu, X., Downey, J., Khalaf, R., Martin, J., Koshino, H., Zhang, W.
Biosynthesis of isonitrile lipopeptides by conserved nonribosomal peptide synthetase gene clusters in Actinobacteria. *Proc Natl Acad Sci* **114**(27) 7025-7030 (2017).

## 2.1 Introduction

One of the most important classes of secondary metabolites are nonribosomal peptides, which are typically biosynthesized by modular NRPSs in an assembly-line manner.[80] Two NRPS-encoding gene clusters (*mbt* and *Rv0096-0101*) have been identified from the genome of *M. tuberculosis*. Although the cluster of mbt has been characterized to biosynthesize mycobactin siderophores that form mycobactin-Fe(III) complexes for iron sequestration,[81] the role of *Rv0096-0101* remains obscure despite various biological studies that have indicated the production of a virulence factor by this gene cluster.[82–90] For example, using transposon-site hybridization, *Rv0098* to *Rv0101* were predicted to be required for *M. tuberculosis* survival in a mouse model of infection.[86] Consistently, a transposon insertion of *Rv0097* attenuated *M. tuberculosis* growth and survival in mice.[83]

An *in silico* homology search has revealed that gene clusters homologous to *Rv0096-0101* are conserved in pathogenic mycobacteria, such as *M. bovis*, *M. leprae*, *M. marinum*, *M. ulcerans*, and *M. abscessus* (**Fig 2-1**), but absent in nonpathogenic mycobacteria such as *M. smegmatis*, providing further indication of the virulence associated nature of the locus product in mycobacteria. Interestingly, in addition to the genus of *Mycobacterium*, related operons are found in the phylum of Actinobacteria across genera including *Streptomyces*, *Kutzneria*, *Nocardia*, and *Rhodococcus* (**Fig 2-1**), suggesting a widespread presence of this cluster. Further bioinformatic analysis has shown that five genes (*Rv0097-0101*) are conserved across all identified gene clusters and that these genes encode proteins homologous to an iron(II) and α-KG dependent oxidase, a fatty acyl-CoA thioesterase, an acyl-acyl carrier protein ligase (AAL), an acyl carrier protein (ACP), and a single- or dimodule NRPS, respectively (**Fig 2-1**). Although all of these five proteins are typically involved in secondary metabolite biosynthesis, the identity of the corresponding metabolite and the specific function of these proteins have not yet been fully elucidated.

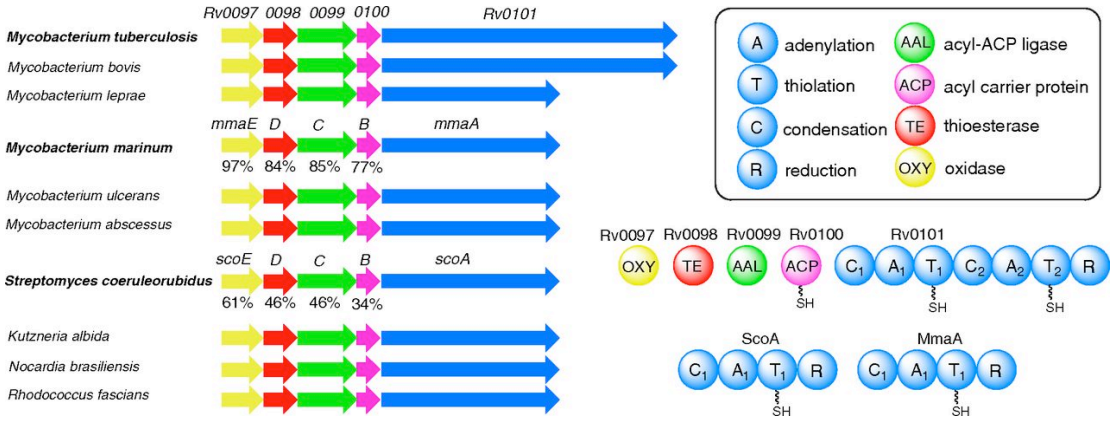



**Figure 2-1.** Schematic of the selected conserved biosynthetic gene clusters and their encoding protein products. Thousands of homologous gene clusters have been identified from published genomes. The similarity of protein homologs from *M. marinum* and *S. coeruleorubidus* to *M. tuberculosis* is shown below the gene clusters. $A_1$ in Rv0101, MmaA, and ScoA was predicted to activate Lys based on the 10-residue specificity sequence (DIEDVGSVVK, DIEDVGSVVK, and DTEDVGTVVK, respectively). $A_2$ in Rv0101 was predicted to activate Phe (DAWTVAAICK).

To better understand the role of this widespread gene cluster, we turned to reconstitution in *E. coli* as a means to quickly and systematically assess the function of the five conserved enzymes through metabolomic exploration. The enzymes from *M. tuberculosis* H37Rv, *M. marinum* strain M (an opportunistic human pathogen), and *Streptomyces coeruleorubidus* NRRL18370[91] (a known pacidamycin producer) were studied and compared, which revealed similarities and variations in biosynthetic functions (**Fig 2-1**). We discovered that these five conserved enzymes were necessary and sufficient to synthesize a unique group of isonitrile lipopeptides (INLPs). Based on both *in vivo* reconstitution and *in vitro* biochemical analysis, we scrutinized the timing and substrate specificity of these enzymes in INLP biosynthesis.

## 2.2 Results

### 2.2.1 Biosynthesis of INLPs by ScoA-E in *E. coli*

Previous biochemical studies of Rv0099-0101 suggested that a lipopeptide might be produced by these enzymes.[92,93] To unveil molecular features of the putative lipopeptide and gain preliminary knowledge of the functions of the five conserved biosynthetic enzymes, *scoA-E* from *S. coeruleorubidus* and *mmaA-E* from *M. marinum* (**Fig 2-1**) were cloned for *E. coli* heterologous expression and functional reconstitution, respectively. A negative control strain transformed with empty vectors was also constructed. Protein expression analysis showed that all 10 proteins were solubly expressed in *E. coli* (**Fig 2-2**). The *E. coli* cultures were extracted with organic solvents, concentrated, and analyzed by liquid chromatography-high resolution mass spectrometry (LC-HRMS) followed by

untargeted metabolomics analysis using XCMS[94,95] for the determination of metabolic profile differences and the identification of new metabolites.

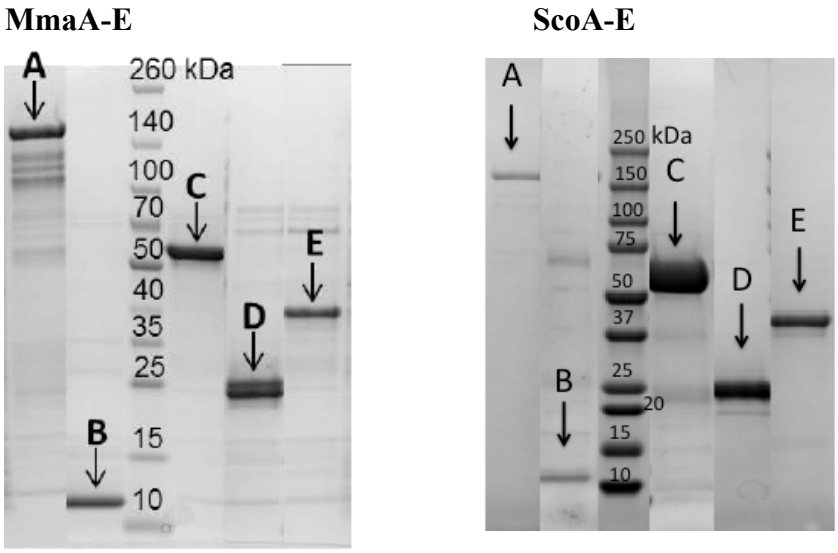



**Figure 2-2.** SDS-PAGE analysis of the *E. coli* purified proteins. All proteins have an *N*-terminal hexahistidine tag and were expressed as largely soluble. Criterion Tris-HCl gels (4-15% precast, Biorad) were used.

Although no new metabolite was immediately identified in the culture of *E. coli-mmaA-E* compared with the negative control strain, upon coexpression of scoA-E, two new metabolites with molecular formulas $C_{18}H_{28}N_4O_4$ (INLP **1**, calculated for $C_{18}H_{29}N_4O_4^+$: 365.2183; found: 365.2185) and $C_{16}H_{26}N_4O_3$ (INLP **2**, calculated for $C_{16}H_{27}N_4O_3^+$: 323.2078; found: 323.2079) that were absent in the negative control were found to be produced (**Fig 2-3 and Appendix A**). Both metabolites were UV-inactive above 210 nm and acid sensitive. The minor metabolite INLP **2** was further predicted to be deacetylated INLP **1** based on HRMS/MS analysis (**Appendix A**). To purify a sufficient quantity of INLP **1** and **2** for structural elucidation, a total of ~40 L of *E. coli* culture was prepared and extracted with chloroform, followed by purification via multiple rounds of HPLC using reverse-phase C18 columns. These purification steps yielded ~3 mg of pure compound INLP **1** and ~0.5 mg of INLP **2**. NMR spectra, including $^1H$, $^{13}C$, dqf-COSY, HSQC, HMBC, and ROESY spectra, were obtained for compound INLP **1** and used to determine its molecular connectivity (**Fig 2-3, Table 2-1, and Appendix B**). The presence of an isonitrile moiety in INLP **1** was further confirmed by comparison with the reported $^{13}C$-$^{14}N$ nuclear spin coupling constants and IR spectroscopy absorption (**Fig 2-3**).[96] The absolute configuration of the C3′ and C3″ (both R) was determined by acid hydrolysis to yield the monomer of 3-amino butyric acid, which was then reacted with Marfey's reagent and compared with the standards (**Fig 2-3**).[97] The NMR spectra of 2 confirmed that it is a C1 deacetylated analog of INLP **1** (**Fig 2-3, Table 2-2, and**

**Appendix C**). The molecular structures of INLP **1** and **2** were revealed to be similar to two known isonitrile antibiotics (SF2768 and SF2369) that were originally isolated from the culture filtrates of different Actinomycetes species (**Fig 2-3**),[98–100] strongly indicating that *E. coli* was a suitable heterologous host for the reconstitution of ScoA-E activities to produce relevant metabolites. Systematic removal of each of the five genes from *E. coli-scoA-E* abolished the production of INLP **1** and **2**, demonstrating the necessity of all five enzymes in synthesizing these two INLPs (**Fig 2-3**).

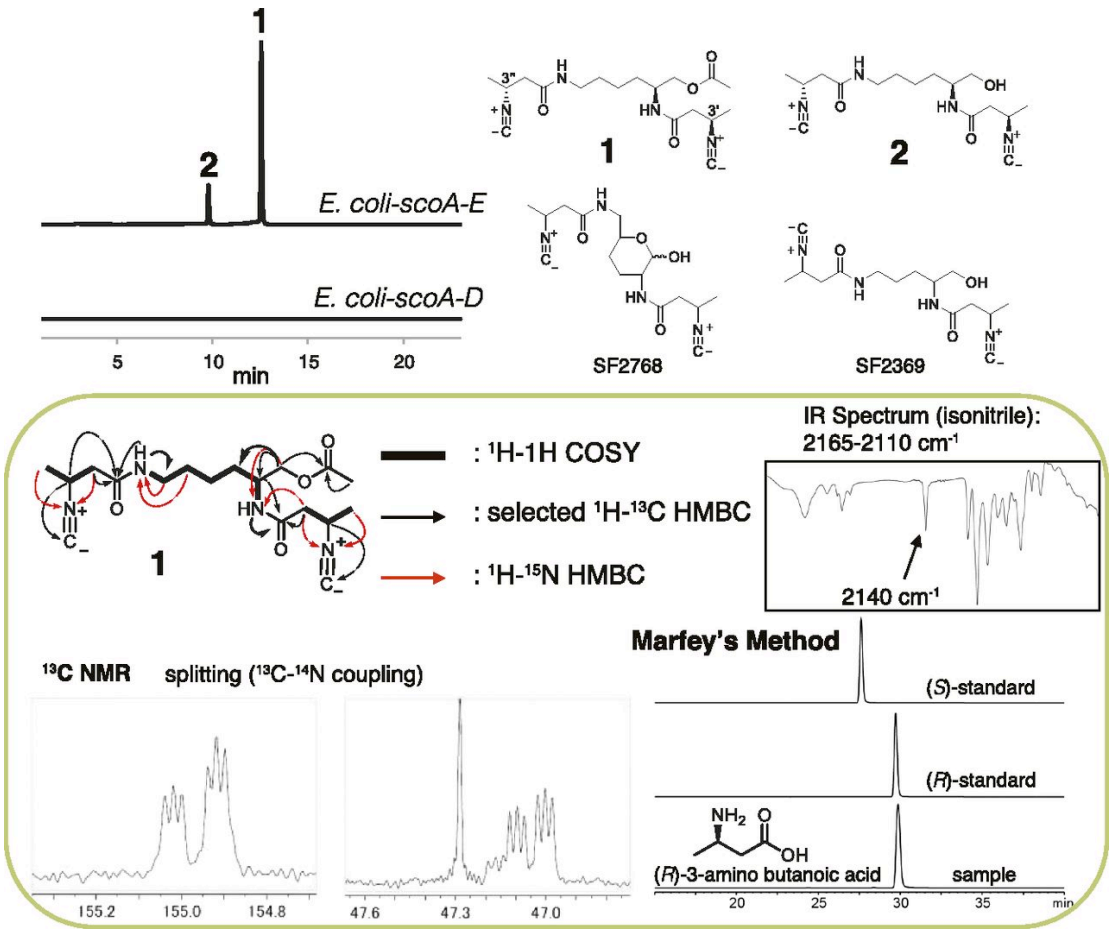
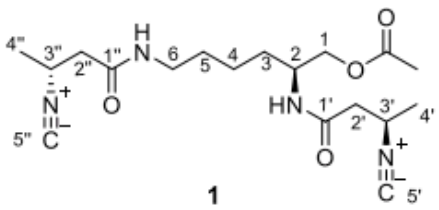



**Figure 2-3.** Biosynthesis of INLPs (INLP **1** and **2**) by ScoA-E in *E. coli.* Extracted ion chromatograms show production of INLP **1** and **2** by *E. coli-scoA-E*. The calculated masses for INLP **1** and **2** with 10-ppm mass error tolerance were used for each trace. The negative control strain with empty vectors and strains containing any of the four gene combinations did not produce INLP **1** and **2**, and only one representative trace is shown

here for simplicity. The structural determination of INLP **1** is boxed. Structures of two known Actinomycetes metabolites are also shown here.

**Table 2-1.** NMR data for compound **1** in DMSO-d6.

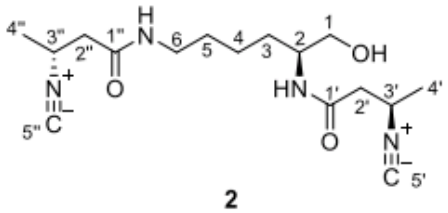


**1**

| Position | $\delta_H$ [ppm] | | mult. (*J* in Hz) | $^1H$-$^{13}C$ HMBC | $\delta_C$ [ppm] |
|---|---|---|---|---|---|
| 1 | 3.84 | 1H | dd (10.9, 6.5) | 2,3,1-OAc(C) | 65.4 |
| | 3.99 | 1H | dd (10.9, 4.8) | | |
| 2 | 3.94 | 1H | m | 1′,3,4 | 47.3 |
| 3 | 1.47 | 1H | m | 1,2,4,5 | 30.2 |
| | 1.34[a] | 1H | m | | |
| 4 | 1.34[a] | 1H | m | 3,6 | 22.6 |
| | 1.26 | 1H | m | | |
| 5 | 1.39 | 2H | m | 3,4,6 | 28.8 |
| 6 | 3.03 | 2H | m | 1″,4,5 | 38.3 |
| 1′ | | | | | 168.0 |
| 2′ | 2.41[b] | 1H | m | 1′,3′,4′ | 42.4 |
| | 2.47[c] | 1H | m | | |
| 3′ | 4.07[d] | 1H | m | 1′,2′,4′,5′ | 47.1 |
| 4′ | 1.31[e] | 3H | m | 1′,2′,3′ | 20.9 |
| 5′ | | | | | 155.0 |
| 1″ | | | | | 167.8 |
| 2″ | 2.37[b] | 1H | m | 1’’, 3’’, 4’’ | 42.3 |
| | 2.44[c] | 1H | m | | |
| 3″ | 4.07[d] | 1H | m | 1″,2″,4″,5″ | 47.0 |
| 4″ | 1.31[e] | 3H | m | 1″,2″,3″ | 21.0 |
| 5″ | | | | | 154.9 |
| 1-OAc | | | | | 170.2 |
| | 1.99 | 3H | s | 1, 1-OAc (C) | 20.6 |
| 2-NH | 7.93 | 1H | d (8.5) | 1,2,3,1′,2′,3″ | |
| 6-NH | 7.97 | 1H | m | 5,6,1″ | |

[a], [b], [c], [d], [e], overlapped signals.

**Table 2-2.** NMR data for compound **2** in DMSO-d6.



**2**

| Position | $\delta_H$ [ppm] | | mult. (*J* in Hz) | $^1H$-$^{13}C$ HMBC | $\delta_C$ [ppm] |
|---|---|---|---|---|---|
| 1 | 3.24 | 1H | m | | 63.3 |
| | 3.32 | 1H | m | | |
| 2 | 3.70 | 1H | m | | 50.5 |
| 3 | 1.24[a] | 1H | m | | 30.4 |
| | 1.53 | 1H | m | | |
| 4 | 1.24[a] | 1H | m | 3 | 22.8 |
| | 1.31[b] | 1H | m | | |
| 5 | 1.35 | 1H | m | | 29.1 |
| | 1.40 | 1H | m | | |
| 6 | 3.03 | 2H | m | 1″,4,5 | 38.5 |
| 1′ | | | | | 167.9 |
| 2′ | 2.44 | 1H | m | 3′ | 42.5 |
| | 2.48 | 1H | m | | |
| 3′ | 4.07[c] | 1H | m | 2′, 4′ | 47.2 |
| 4′ | 1.30[b] | 3H | m | 2′, 3′ | 21.0 |
| 5′ | | | | | 154.9 |
| 1″ | | | | | 168.1 |
| 2″ | 2.37 | 1H | m | 3″ | 42.3 |
| | 2.38 | 1H | m | | |
| 3″ | 4.07[c] | 1H | m | 3″ | 47.1 |
| 4″ | 1.30[b] | 3H | m | 3″ | 21.0 |
| 5″ | | | | | 153.6 |
| 1-OH | 4.65 | 1H | m | | |
| 2-NH | 7.76 | 1H | d (8.5) | 1′, 2 | |
| 6-NH | 7.98 | 1H | m | | |

[a], [b], overlapped signals.

### 2.2.2 Proposed Biosynthetic Pathways for INLPs by the Five Conserved Biosynthetic Enzymes

The heterologous production of INLP **1** and **2** in *E. coli* enabled the assignment of function to the five conserved biosynthetic enzymes, especially the role of the two modification enzymes (ScoD and ScoE), in a unique isonitrile moiety synthesis. We have thus proposed putative enzymatic pathways for INLP biosynthesis (**Fig 2-4**). The INLP **2** is presumably generated through a hybrid AAL-NRPS assembly-line–based mechanism. The assembly line starts with the activation of crotonic acid by ScoC through adenylation, and the adenylated acid is loaded onto an ACP (ScoB) for further processing. The α,β-unsaturated fatty acyl-ACP is then modified by a thioesterase homolog (ScoD) and a nonheme iron(II)- dependent oxidase (ScoE) to generate a β-isonitrile fatty acid intermediate. In addition to being a thioesterase as suggested by the previous biochemical and structural analyses of Rv0098,[101] we propose ScoD to additionally function as a reverse lyase-like enzyme that catalyzes a Michael addition of Gly to the β-position of an α,β-unsaturated fatty acyl-ACP to yield an N-carboxymethyl-3-aminoacyl-ACP and then further hydrolyzing this product to form N-carboxymethyl-3-aminobutanoic acid. ScoE presumably catalyzes the subsequent oxidation and decarboxylation to yield a β-isonitrile moiety. The resulting 3-isocyanobutanoic acid is then adenylated by ScoC and loaded back onto ScoB. The isonitrile-modified fatty acyl chain is then condensed to both amino groups of Lys promoted by the NRPS (ScoA) and reductively released by the reduction (R) domain of the NRPS to form a terminal alcohol product INLP **2**. This reductive release mechanism is consistent with the identified activity of the R domain in Rv0101, which contains a conserved Ser/Tyr/Lys catalytic triad and was demonstrated to catalyze a four electron thioester reduction using the purified truncated R domain and a synthetic substrate.[93] We propose that the extra acetyl moiety found in INLP **1** is due to the activity of a promiscuous acetyltransferase endogenous to *E. coli* for possible detoxification, and similar acetylation events have been observed in other shunt product biosynthesis.[102] An analogous biosynthetic pathway involving MmaA-E has also been proposed, differing in the chain length of the fatty acyl moiety (**Fig 2-4**). Although ScoC seems to prefer a short-chain fatty acid substrate (such as crotonic acid as shown in INLP **1** and **2**), MmaC likely favors a medium-chain fatty acid substrate (such as 2-decenoic acid) because previous biochemical and structural analyses of Rv0099, a homolog of MmaC from the same genus of *Mycobacterium*, showed that this AAL activated fatty acids of medium-chain length.[92,93]

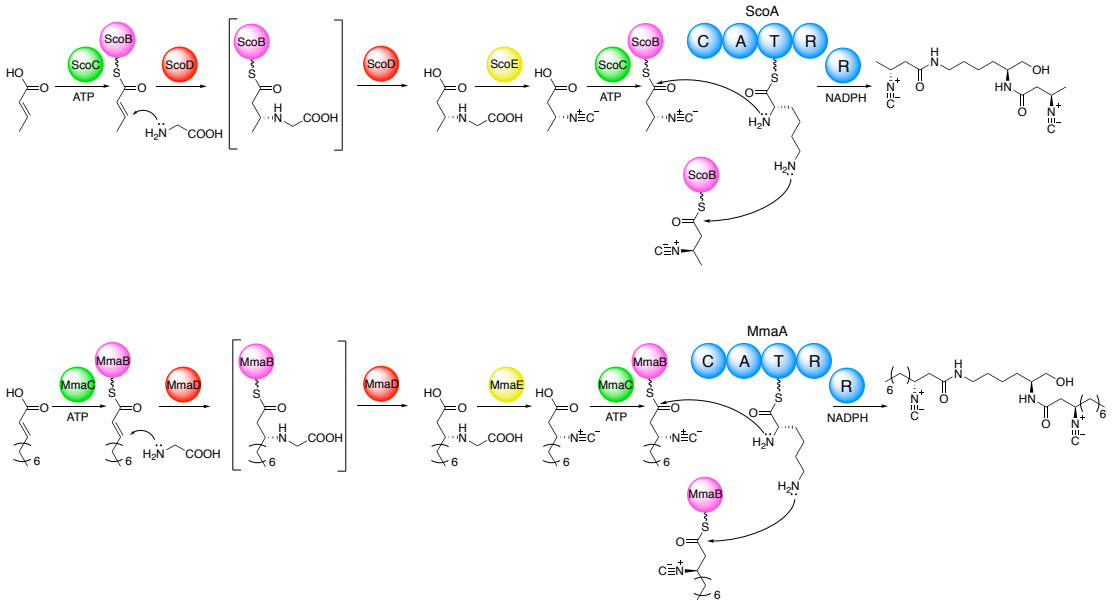
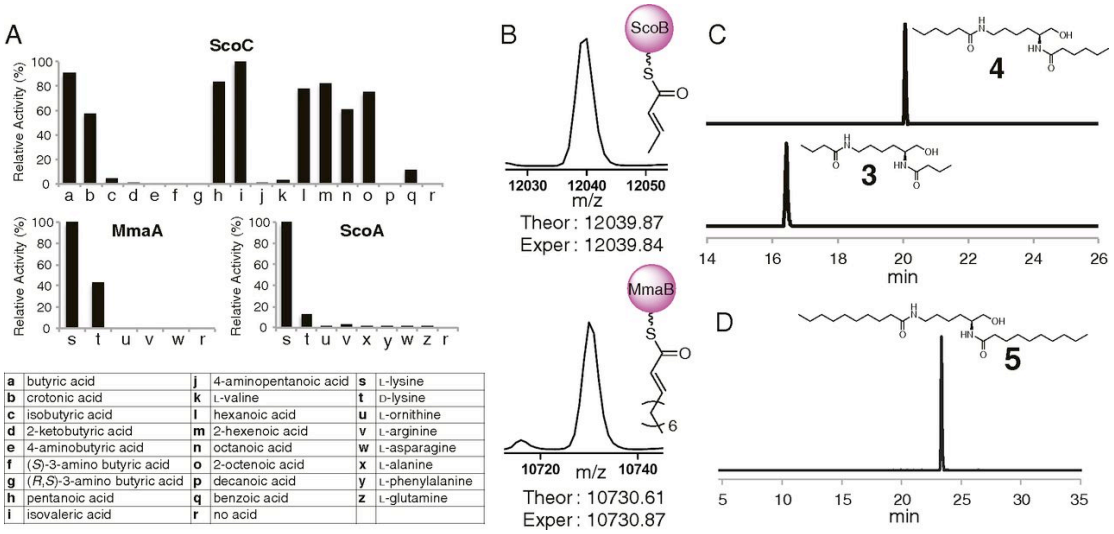



**Figure 2-4.** Proposed function of ScoA-E and MmaA-E in INLP biosynthesis.

### 2.2.3 AAL, ACP, and NRPS Promote Diacylated Lipopeptide Biosynthesis

We next performed additional *in vivo* metabolic analyses using various combinations of biosynthetic genes and *in vitro* biochemical analyses using purified enzymes to dissect the proposed biosynthetic pathways for INLPs (**Fig 2-4**). The substrate specificity of NRPSs, including ScoA and MmaA, was tested using the classical ATP-[$^{32}$P]PP$_i$ exchange assay. As expected, both enzymes demonstrated a strong preference for the activation of L-Lys (**Fig 2-5A**), which is consistent with the molecular structures of INLP **1** and **2** and supports the assignment of the absolute configuration of C2 in INLP **1** to be 2S (**Fig 2-3**). Although L-Ornithine was not activated by ScoA/MmaA, it could be a preferred substrate for other conserved NRPSs based on the structure of SF2369 (**Fig. 2**). We next probed the fatty acid substrate specificity of AALs encoded by *scoC* and *mmaC*, respectively. The ability of ScoC to reversibly adenylate various acids was tested using the ATP-[$^{32}$P]PP$_i$ exchange assay. ScoC exhibited a strong preference for the activation of fatty acids with a short-chain length (C4–C8), and as expected, α,β-unsaturated fatty acids were well recognized (**Fig 2-5A**). The subsequent loading of selected fatty acyl moieties onto ScoB was further confirmed by HRMS analysis (**Fig 2-5B**). MmaC demonstrated an intrinsic ATPase activity in the ATP-PP$_i$ exchange assay that prohibited the determination of substrate specificity using this method. Nonetheless, the direct substrate activation and loading assays confirmed that, similar to Rv0099, MmaC preferentially activates fatty acids of medium-chain length with tolerance toward α,β-unsaturation (**Fig 2-5B**). One of the unusual events in the proposed biosynthetic pathway is the condensation of the same fatty acyl moiety to both amino groups of Lys, presumably promoted by a single C domain of the NRPS (**Fig 2-4**). We hypothesized that this C domain has relaxed substrate specificity, and the product formation assays using ScoA-C or MmaA-C successfully yielded lipopeptides (LP **3**–**5**) with the amide bond

formation at both amino positions (**Fig 2-5C and D and Appendix D**). Negative controls missing any of the enzyme or substrate (fatty acid, Lys, ATP, and NADPH) abolished the production of LP **3–5**. Products with a terminal alcohol were generated in these assays, confirming the four-electron reduction activity of the conserved R domain in ScoA and MmaA.




| | | | | | |
|---|---|---|---|---|---|
| **a** | butyric acid | **j** | 4-aminopentanoic acid | **s** | L-lysine |
| **b** | crotonic acid | **k** | L-valine | **t** | D-lysine |
| **c** | isobutyric acid | **l** | hexanoic acid | **u** | L-ornithine |
| **d** | 2-ketobutyric acid | **m** | 2-hexenoic acid | **v** | L-arginine |
| **e** | 4-aminobutyric acid | **n** | octanoic acid | **w** | L-asparagine |
| **f** | (*S*)-3-amino butyric acid | **o** | 2-octenoic acid | **x** | L-alanine |
| **g** | (*R,S*)-3-amino butyric acid | **p** | decanoic acid | **y** | L-phenylalanine |
| **h** | pentanoic acid | **q** | benzoic acid | **z** | L-glutamine |
| **i** | isovaleric acid | **r** | no acid | | |

**Figure 2-5.** Biochemical analysis of AALs and NRPSs. (*A*) Substrate specificity of ScoA, MmaA, and ScoC determined by ATP-[$^{32}$P]PP$_i$ exchange assays. (*B*) Detection of crotonyl-*S*-ScoB and decenoyl-*S*-MmaB by MS with maximum entropy deconvolution. (*C*) Extracted ion chromatograms showing the production of LP **3** and **4** in ScoA-C assays using butyric acid and hexanoic acid as a substrate, respectively. (*D*) Extracted ion chromatogram showing the production of LP **5** in the MmaA-C assay using decanoic acid as a substrate. The calculated masses for LP **3–5** with 10-ppm mass error tolerance were used for each trace.

### 2.2.4 Homologs of Rv0097 and Rv0098 Promote Isonitrile Biosynthesis through an Unprecedented Mechanism

Although the isonitrile functionality has been found in quite a few natural products, only one biosynthetic pathway has been identified in which one carbon is transferred to an amino group catalyzed by an isonitrile synthase such as IsnA.[47,65,103,104] Our proposed pathway for isonitrile synthesis is mechanistically distinct from the known pathway by using a different set of enzymes. Rv0098 has previously been shown to be a long-chain fatty acyl-CoA thioesterase, although structurally it lacks a general base or a nucleophile that is conserved in the thioesterase catalytic site. In addition, very low hydrolysis activities were obtained in the biochemical analysis of Rv0098, which solicited further detailed characterization of this hypothetical protein.[101] The recent biochemical characterization of CmiS1, a homolog of ScoD (identity/similarity = 47%/56%), showed that CmiS1 catalyzed the Michael addition of Gly to the β-position of a non–2- enoic acid thioester in the biosynthesis of the macrolactam antibiotic cremimycin.[105] We thus have

proposed that a similar reaction of Michael addition of Gly to the β-position of an α,β-unsaturated fatty acyl-ACP could be promoted by ScoD/MmaD/Rv0098 to yield an N-carboxymethyl-3-aminoacyl-ACP (**Fig 2-4**). This is consistent with the failed activation of 3-amino butyric acid by ScoC (**Fig 2-5A**), which argued against the known pathway using an isonitrile synthase. To confirm the proposed function of MmaD, a biochemical reaction using MmaB-D, ATP, 2-decenoic acid, and Gly was performed, and the product was analyzed by LC-HRMS after release from the protein by base hydrolysis. This yielded the expected Gly adduct (calculated for $C_{12}H_{24}NO_4^+$: 246.1700; found: 246.1701), and its mass was shifted by +1 using [2-$^{13}$C]Gly or by +2 using [2-$^{13}$C, $^{15}$N]Gly as an alternative substrate (**Fig 2-6A and Appendix E**). Negative controls omitting any of the three proteins or substrate abolished the production of the Gly adduct, indicating that the formation of the Gly adduct requires the activation of the fatty acid substrate and occurs on a thio-templated assembly line (**Fig 2-6A**). Similar reactions were also performed using ScoB-D, and the activity of ScoD in forming a Gly adduct was also confirmed (**Appendix F**). We further observed that the Gly adduct on ACP (either MmaB or ScoB) was readily released by hydrolysis *in vitro*, likely promoted by MmaD/ScoD, consistent with the observed thioesterase activity of Rv0098 and CmiS1.[81,106] We thus could not originally reconstitute the activity of MmaE/ScoE, the iron(II), and the α-KG– dependent oxidase homolog in promoting the subsequent oxidation and decarboxylation *in vitro*, most likely due to substrate limitation. Chapter 3 will discuss further biochemical characterization of ScoE.

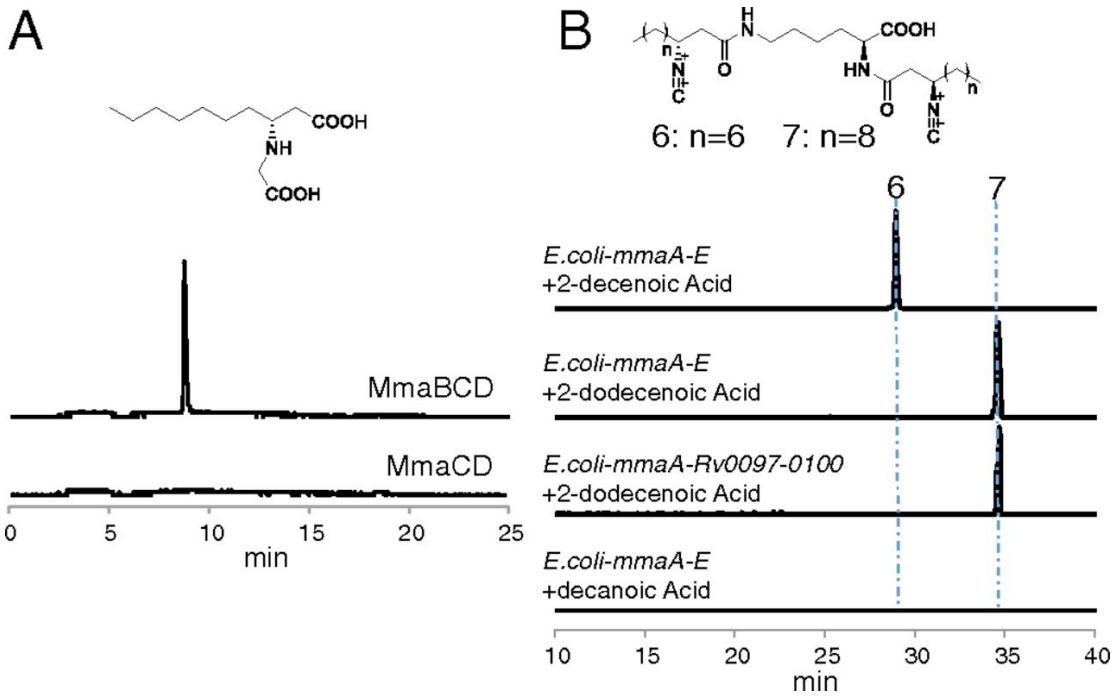



**Figure 2-6.** *In vitro* and *in vivo* analysis of isonitrile formation. (*A*) Extracted ion chromatogram showing the production of Gly adduct in the MmaB-D assay. The calculated mass with 10-ppm mass error tolerance was used for each trace. Control assays missing any of the protein or substrate (ATP, 2-decenoic acid, and Gly) abolished the

production of Gly adduct, and only one representative trace with no MmaB is shown for simplicity. (*B*) Extracted ion chromatograms showing the *E. coli*-based production of INLP **6** and **7** after feeding of 2-decenoic acid and 2-dodecenoic acid, respectively. The calculated masses for INLP **6** and **7** with 10-ppm mass error tolerance were used for each trace. Strains containing any of the four gene combinations, or feeding of decanoic acid and dodecanoic acid, did not produce INLP **6** and **7**, and only one representative trace with the feeding of decanoic acid to *E. coli-mmaA-E* is shown here for simplicity.

Because reconstituting the activity of MmaE/ScoE *in vitro* was not originally successful, we turned to *in vivo* reconstitution coupled with feeding of isotope-labeled Gly to further confirm the proposed activity of enzymes in isonitrile biosynthesis. Although coexpression of *scoA-C* in *E. coli* yielded expected diacylated lipopeptides such as LP **3** and acetylated LP **3**, the addition of *scoD* or *scoE* to *E. coli-scoA-C* did not change the metabolic profile with no new metabolite identified through comparative metabolomics, indicating that the modification of the fatty acyl chain occurs before the biosynthetic intermediate release by the NRPS and that the C or R domain of ScoA is intolerant of a bulky side group on the fatty acyl chain (Gly adduct). Only upon the addition of both *scoD* and *scoE* to *E. coli-scoA-C* were new major products of INLP **1** and **2** identified (**Fig 2-3**). We thus reasoned that ScoD functions on the assembly line, most likely on the free-standing ACP, ScoB. Additionally, feeding of 10 mM of Gly to the culture of *E. coli-scoA-E* boosted the titer of INLP **1** by over 20-fold, and feeding of [2-$^{13}$C, $^{15}$N]Gly showed that C(2)-N of Gly was incorporated into the isonitrile group of INLP **1** and **2**, supporting the proposed role of ScoD and ScoE in isonitrile biosynthesis (**Appendix A**).

To further confirm the role of MmaA-E in INLP biosynthesis, we reattempted the *in vivo* reconstitution of activities of MmaA-E in *E. coli*. Based on the substrate specificity of MmaC, we reasoned that the *in vivo* substrate limitation could be the reason for the failed INLP production by *E. coli-mmaA-E*. The 2-decenoic acid or 2-dodecenoic acid was then fed to the culture of *E. coli-mmaA-E*, and untargeted metabolomics analysis was performed to search for new metabolites. This led to the identification of a trace amount of two new metabolites with molecular formulas $C_{28}H_{48}N_4O_4$ (INLP **6**, calculated for $C_{28}H_{49}N_4O_4^{+}$: 505.3748; found: 505.3748) and $C_{32}H_{56}N_4O_4$ (INLP **7**, calculated for $C_{32}H_{57}N_4O_4^{+}$: 561.4374; found: 561.4376), respectively (**Fig 2-6B and Appendix G**). HRMS analysis suggested that INLP **6** and **7** are INLPs similar to INLP **1** and **2**, but contain two longer fatty acyl chains and a C1 acid moiety (**Appendix G**). The presence of the isonitrile was further confirmed by IR spectroscopy to show signature absorption of 2,132 $cm^{-1}$ . As expected, the production of INLP **6** and **7** was dependent on the coexpression of all five genes of *mmaA-E*, and feeding of [2-$^{13}$C, $^{15}$N]Gly to the culture of *E. coli-mmaA-E* demonstrated that C(2)-N of Gly was incorporated into INLP **6** (**Fig 2-6B and Appendix G**). It is notable that feeding of decanoic acid or dodecanoic acid did not lead to the production of INLP **6** and **7**, supporting our hypothesis that the biosynthesis of INLPs requires the activation of an α,β-unsaturated fatty acid (**Fig 2-4**). In addition, replacement of *mmaB-E* by *Rv0097-0100* in *E. coli* produced INLP **7** upon the feeding of 2-dodecenoic acid, demonstrating the interchangeable property of the encoding four enzymes (**Fig 2-6B**). The INLP alcohols could not be detected, suggesting that 2-decenoic acid and 2-dodecenoic acid may not be the native substrate for the *M. marinum* proteins that led to the impaired activity of the AALNRPS assembly line and

low titers of the corresponding INLPs in *E. coli*. Indeed bioinformatic analysis showed that putative fatty acid modification enzymes are encoded in close proximity to *mmaA-E* in the *M. marinum* genome.

## 2.3 Discussion

We have revealed the function of a biosynthetic gene cluster that is widespread in Actinobacteria including pathogenic mycobacteria such as *M. tuberculosis* and *M. marinum*. Using both in vivo and in vitro analyses, we discovered that the conserved five biosynthetic enzymes were capable of synthesizing a family of INLPs by using a thio-template mechanism. This unusual biosynthetic pathway starts with the activation and loading of an α,β-unsaturated fatty acid onto an ACP by a promiscuous AAL. A Michael addition of Gly to the β-position of the α,β-unsaturated fatty acyl-ACP and subsequent release of N-carboxymethyl-3-aminobutanoic acid is then promoted by a thioesterase homolog, followed by the oxidation and decarboxylation, presumably catalyzed by a nonheme iron(II)-dependent oxidase, to generate a β-isonitrile fatty acid, which is then reloaded onto an ACP, by the AAL. This isonitrile intermediate is then condensed to both amino groups of Lys promoted by an NRPS and reductively released to form a terminal alcohol product. The identified isonitrile biosynthetic pathway is distinct from the canonical mechanism in which one carbon is transferred to an amino group to form isonitrile catalyzed by an isonitrile synthase.[47,65,103,104] In addition, the condensation of an acyl moiety to both amino groups of Lys, presumably promoted by a single C domain of the NRPS, is rare in nonribosomal peptide biosynthesis. Further mutagenesis study in *M. marinum* suggests that this biosynthetic gene cluster plays a role in metal transport. This study thus sheds light on a metal transport system that is critical for the virulence of pathogenic mycobacteria. Future study of native INLP metallophores in mycobacteria and their mode of action during infection may inspire new therapies to combat pathogenic mycobacteria, in particular *M. tuberculosis*, from which millions of people die every year.

## 2.4 Materials and Methods

### 2.4.1 Construction of Plasmids for Protein Expression in *E. coli*.

Individual genes were PCR amplified from genomic DNA and cloned into pET24b, pETDuet, or pCDFDuet-1 by restriction enzyme digestion (Thermo Scientific) and ligation with Quick T4 DNA ligase (New England Biolabs). All primers used are listed in Table S3. Plasmids were isolated using a QIAprep Spin Miniprep Kit (Qiagen) and confirmed by DNA Sequencing (UC Berkeley DNA Sequencing Facility).

**Table 2-3.** Primers used in this study.

| Primer | Sequence (5' -> 3') |
|---|---|
| Duet-mmar101-F | aaagaattcgGTGACCGCGCCCGAAATCGG |
| Duet-mmar101-R | tataagcttCTAGGCCAGTGCGGGTCTTG |
| pET24b-mmar100-F | aaacataTGGCTGACCCGGTGCGCCA |
| pET24b-mmar100-R | aaactcgagCGTAGGACTTTCCAGTTCGG |

| | |
|---|---|
| **Duet-mmar99-F** | **aaaggatccgATGTCCGACTTACCCGCTAC** |
| **Duet-mmar99-R** | **aaactgcagTCAGCCACCGGTTTGGACCT** |
| **Duet-mmar98-F** | **aaaggatccgcatATGAGCACCACCGATTTGAC** |
| **Duet-mmar98-R** | **tataagcttcaattgCTAGGGGATGTTCAGGGCCG** |
| **Duet-mmar97-F** | **aaaggatccgATGACGCTCAACGTGAAAGG** |
| **Duet-mmar97-R** | **tataagcttTCATGCCGGGTAGCCCGGCG** |
| **MmaNRPS-ETDuet-F** | **aaactgcagGTGACCGCGCCCGAAATCGG** |
| **MmaNRPS-ETDuet-R** | **aaaaagcttCTAGGCCAGTGCGGGTCTTG** |
| **MmaAAL-RSFDuet-F** | **aaagaattcgATGTCCGACTTACCCGCTAC** |
| **MmaAAL-RSFDuet-R** | **aaaaagcttTCAGCCACCGGTTTGGACCT** |
| **MmaACP-RSFDuet-F** | **aaaaaaacatatgCTGACCCGGTGCGCCA** |
| **MmaACP-RSFDuet-R** | **aaactcgagCGTAGGACTTTCCAGTTCGG** |
| **MmaOXY-CDFDuet-F** | **aaacatatgATGACGCTCAACGTGAAAGG** |
| **MmaOXY-CDFDuet-R** | **aaactcgagTCATGCCGGGTAGCCCGGCG** |
| **MmaTE3-CDFDuet-F** | **aaactgcagATGAGCACCACCGATTTGAC** |
| **MmaTE3-CDFDuet-R** | **aaaaagcttCTAGGGGATGTTCAGGGCCG** |
| **MmaTe3-ACYCDuet-F** | **aaactgcagATGAGCACCACCGATTTGAC** |
| **MmaTe3-ACYCDuet-R** | **aaaaagcttCTAGGGGATGTTCAGGGCCG** |
| **AAL-RSFDuet-F** | **aaagaattcgATGGACCGGCTCCACCACCC** |
| **AAL-RSFDuet-R** | **aaaaagcttTCAGTTGACCTTGCGTGCGG** |
| **ACP2-RSFDuet-F** | **aaaaaacatatgCTGCTCCCCTCACGCT** |
| **ACP2-RSFDuet-R** | **aaactcgagTCATGCGGTGACATGGCCCG** |
| **Oxy-CDFDuet-F** | **aaacatatgATGCAGATCGACGAACAGCC** |
| **Oxy-CDFDuet-R** | **aaactcgagTCATGCCGCCTGGATCCCGT** |
| **TE3-CDFDuet-F** | **aaactgcagATGACGGACGAAGCCCTGCT** |
| **TE3-CDFDuet-R** | **aaaaagcttTCAGGGGACGTTGACGAAGG** |
| **NRPS-ETDuet-F** | **aaactgcagATGTCACCGCATGACGACGC** |
| **NRPS-ETDuet-R** | **aaaaagcttCTACTTGGCGGGCATTGCCG** |
| **TE3-ACYCDuet-F** | **aaactgcagATGACGGACGAAGCCCTGCT** |
| **TE3-ACYCDuet-R** | **aaaaagcttTCAGGGGACGTTGACGAAGG** |
| **Duet-sco101-F** | **aaagaattcgATGTCACCGCATGACGACGC** |
| **Duet-sco101-R** | **tataagcttCTACTTGGCGGGCATTGCCG** |
| **pET24b-sco100-F** | **aaacataATGCCTGCTCCCCTCACGCT** |
| **pET24b-sco100-R** | **aaactcgagTCATGCGGTGACATGGCCCG** |
| **Duet-sco99-F** | **aaaggatccgATGGACCGGCTCCACCACCC** |
| **Duet-sco99-R** | **aaactgcagTCAGTTGACCTTGCGTGCGG** |
| **Duet-sco98-F** | **aaaggatccgcatATGACGGACGAAGCCCTGCT** |
| **Duet-sco98-R** | **tataagcttcaattgTCAGGGGACGTTGACGAAGG** |
| **Duet-sco97-F** | **aaaggatccgATGCAGATCGACGAACAGCC** |
| **Duet-sco97-R** | **tataagcttTCATGCCGCCTGGATCCCGT** |
| **pMSG360 Hyg-F** | **ATCTGGATCCACGAAGCTTC** |

| | |
|---|---|
| **pMSG360 Hyg-R** | **GGGGATCCTCTAGAGTCCTG** |
| **pMSG360 vector-F** | **GGGTACCTGAGAGCCTTCAA** |
| **pMSG360 vector-R** | **ATCCTGCAGGAATTCCTCGA** |
| **MmaE KO-F** | **ggatttcggacaggactctagaggatccccGAGGCCGTCCCGTGCTTTGC** |
| **MmaE KO-R** | **gctgactgggttgaaggctctcaggtacccACGCTCAACGTGAAAGGCGA** |
| **MmaA KO-F** | **accgtacgtctcgaggaattcctgcaggatCTAGGCCAGTGCGGGTCTTG** |
| **MmaA KO-F** | **gggccaccatgaagcttcgtggatccagatCGCGGTCCCTGATCAACTGC** |

### 2.4.2 Identification of *E. coli-scoA-E* and *E. coli-mmaA-E* metabolites through untargeted, comparative metabolomics.

*E. coli* BAP1 strain harnessing plasmids containing either *scoA-E* or *mmaA-E* were grown in triplicate along with a control harnessing empty vectors. 30 mL cultures were grown at 37ºC in LB containing appropriate antibiotics until $OD_{600}$ 0.5 and induced with 0.5 mM IPTG. For *E. coli-mmaA-E*, decenoic acid or dodecenoic acid was fed to a final concentration of 1 mM at the time of induction. After induction, the temperature was decreased to 20ºC, and compound production was allowed to proceed for approximately 48 h. The entire culture (both pellet and supernatant) was extracted with an equal amount of chloroform. The organic extract was dried and redissolved in methanol and analyzed via liquid chromatography-high resolution mass spectroscopy (LC-HRMS). LC-HRMS analysis was normally performed using an Agilent Technologies 6520 Accurate-Mass Q-TOF LC-MS instrument and an Agilent Eclipse Plus C18 column (4.6 x 100 mm). A linear gradient of 2-98% $CH_3CN$ (vol/vol) over 45 min in $H_2O$ with 0.1% formic acid (vol/vol) at a flow rate of 0.5 mL/min was used. XCMS software was used for untargeted metabolomic comparisons of extracts from strains containing *scoA-E* or *mmaA-E* to the empty vector control. In a typical small-scale metabolomics analysis, for compounds **1** and **2**, A linear gradient of 10-50% $CH_3CN$ (vol/vol) over 12 min in $H_2O$ with 0.1% (vol/vol) formic acid at a flow rate of 0.5 mL/min was used. For compounds **6** and **7**, a linear gradient of 50-98% $CH_3CN$ (vol/vol) over 12 min in $H_2O$ with 0.1% (vol/vol) formic acid at a flow rate of 0.5 mL/min, or a linear gradient of 10 to 98% over 30 min followed by 10 min at 98%. HRMS/MS analysis was conducted using targeted MS/MS with collision energy of 5-30V.

### 2.4.3 Large-scale Production, Purification, and Characterization of 1 and 2.

A total of 40 L (40 ×1-L) of the *E. coli-scoA-E* was cultured in LB with appropriate antibiotics. The cultures were inoculated with 10 mL of a seed culture and grown at 37ºC to OD600 ≈ 0.5 – 0.7 before induction with 0.5 mM IPTG. After induction, the temperature was dropped to 20ºC, and compound production was allowed to proceed for approximately two days. The cells were pelleted by centrifugation (6371 × g, 15 min) and compounds **1** and **2** were extracted from the culture medium using two volumes of chloroform. The solvent was removed by rotary evaporation, and the combined residue was re-dissolved in methanol. Purification by HPLC was conducted using an Agilent 1260 HPLC with a C18 Vydac 218TP1022 column 10μm (22 x 250 mm) using a linear gradient of 5-30% $CH_3CN$ (vol/vol) over 50 min in $H_2O$ without formic acid at a flow rate of 5 mL/min and an injection of 1 mL. Fractions were screened using LC-HRMS

using an Agilent Eclipse Plus C18 column (4.6 x 100 mm) and a linear gradient of 10-50% $CH_3CN$ (vol/vol) over 12 min in $H_2O$ with 0.1% (vol/vol) formic acid at a flow rate of 0.5 mL/min. Fractions containing **1** were combined and those containing **2** were also combined, separately. Further HPLC purification was conducted using a C18 Vydac 218TP1022 column 10μm (22 x 250 mm) and an isocratic program of 15% $CH_3CN$ (vol/vol) in $H_2O$ at a flow rate of 5 mL/min. Fractions containing **1** and **2** were again determined by LC-HRMS and combined, separately. A final round of HPLC purification was conducted using an Inertsil ODS-4 column (6 mm x 250 mm) using an isocratic program of 23% $CH_3CN$ (vol/vol) in $H_2O$ at a flow rate of 1 mL/min. The resulting 3 mg of purified **1** and 0.5 mg of purified **2** were dried and analyzed by LC-HRMS and NMR. NMR spectra (1D: $^{1}H$, $^{13}C$ and 2D: HSQC, dqf-COSY, HMBC, and ROESY) were recorded on a Bruker Biospin 900 MHz spectrometer with a cryoprobe in DMSO-d6 (Cambridge Isotope Laboratories).

**2.4.4 Overexpression and Purification of Proteins.**

Expression and purification for all proteins with a His6-tag followed the same general procedure and is detailed as follows: cells were grown at 37°C in 700 mL of LB with the appropriate antibiotic to an OD600 of 0.5. The cells were then cooled on ice for 10 min and induced with 0.12 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 16 h at 16°C. Subsequently, the cells were harvested by centrifugation (6371 × g, 15 min, 4°C), resuspended in 30 mL lysis buffer (25 mM HEPES, pH 8, 0.5 M NaCl, 5 mM imidazole), and lysed by homogenization on ice. Cellular debris was removed by centrifugation (27216 × g, 1 h, 4°C). Ni-NTA agarose resin was added to the supernatant (1.5 mL/L of culture), and the solution was nutated at 4°C for 1 h. The protein-resin mixture was loaded onto a gravity flow column, and proteins were eluted with increasing concentrations of imidazole in Buffer A (20 mM HEPES, pH 8.0). Purified proteins were concentrated and buffer exchanged into Buffer A + 10% glycerol using Amicon Ultra spin filters. Proteins were flash frozen in liquid nitrogen and stored at -80°C. The approximate proteins yields were 8 mg/L for ScoA (158 kDa), 12 mg/L for ScoB (12 kDa), 14 mg/L for ScoC (57 kDa), 40 mg/L for ScoD (19 kDa), 15 mg/L for ScoE (34 kDa), 14 mg/L for MmaA (152kDa), 7 mg/L for MmaB (10 kDa), 30 mg/L for MmaC (58 kDa), 32 mg/L for MmaD (23 kDa), and 18 mg/L for MmaE (34 kDa).

**2.4.5 ATP-$PP_i$ Exchange Assays.**

Assays were performed in 100 μL of reaction buffer (50 mM Tris-HCl/2 mM $MgCl_2$, pH 8) containing 5 mM ATP, 1 mM $Na_4[^{32}P]$-pyrophosphate ($PP_i$) (~3 x $10^6$ cpm/mL), 1 mM TCEP, 5 mM substrate, and 5 μM enzyme. Reactions were incubated at 25°C for 2 h, then quenched by the addition of a charcoal suspension (1.6% w/v activated charcoal, 0.1 M $Na_4PP_i$, 3.5% $HClO_4$). Free $[^{32}P]PP_i$ was removed by centrifugation of the sample followed by washing twice with wash solution (0.1 M $Na_4PP_i$ and 3.5% $HClO_4$). Charcoal-bound radioactivity was measured on a Beckman LS 6500 scintillation counter.

**2.4.6 LC-HRMS Analysis of ScoB and MmaB-bound Biosynthetic Intermediates.**

Assays were performed in 50 μL of 50 mM HEPES (pH 8.0) containing 5 mM ATP, 2 mM $MgCl_2$, 5 mM acid substrate (C4-C12), 0−50 μM of each protein. The reaction mixture was mixed gently and incubated for two hours at room temp. After incubation,

450 uL of water was added and then the mixture was filtered (0.2 μm). LC-HRMS analysis was normally performed using an Agilent Technologies 6520 Accurate-Mass Q-TOF LC-MS instrument and a Phenomenex Aeris widepore XBC18 column (250 x 21 mm) with a linear gradient of 15 to 98% $CH_3CN$ (v/v) over 21 min in $H_2O$ with 0.1% (v/v) formic acid, at a flow rate of 0.15 mL $min^{-1}$. The data were analyzed using Agilent MassHunter Qualitative Analysis software using the maximum entropy deconvolution feature.

**2.4.7 Product Formation Assays Using ScoABC and MmaABC.**

Assays were performed in 50 μL of 50 mM HEPES (pH 8.0) containing 5 mM ATP, 2 mM $MgCl_2$, 5 mM acid substrate (C4-C12), 0−50 μM of each protein, 1 mM TCEP, 2 mM NADPH, and 5 mM Lys. The reaction mixture was mixed gently and incubated for two hours at room temp. After incubation, the reaction was quenched with 50 uL MeOH and protein residues were removed by centrifugation. LC-HRMS analysis was normally performed using an Agilent Eclipse Plus C18 column (4.6 Å~ 100 mm) with a linear gradient of 5 to 98% $CH_3CN$ (v/v) over 30 min in $H_2O$ with 0.1% (v/v) formic acid (for ScoABC), or 15 to 98% over 20 min followed by 10 min at 98% $CH_3CN$ (v/v) in $H_2O$ with 0.1% (v/v) formic acid (for MmaABC), at a flow rate of 0.5 mL $min^{-1}$.

**2.4.8 Biochemical Assays of ScoBCD and MmaBCD.**

Assays were performed in 50 μL of 50 mM HEPES (pH 8.0) containing 5 mM ATP, 2 mM $MgCl_2$, 5 mM acid substrate (C4-C12), 0−50 μM of each protein, and 200 mM Gly. For labeling experiments, [2-$^{13}$C]Gly, [$^{15}$N]Gly, or [2-$^{13}$C, $^{15}$N]Gly was used. The reaction mixture was mixed gently and incubated at room temp. At various time points, the reaction was quenched with 50 uL MeOH and incubated on ice for 10 minutes. The protein residues were pelleted by centrifugation for 10 min at 4ºC and the supernatant was used for LC-HRMS analysis. The protein pellet was then resuspended in 100 uL of MeOH and centrifuged again and the supernatant was discarded (this process was repeated three times). The pellet was then re-dissolved in 50 uL of 100 mM KOH and heated at 70ºC for 10 min. The solution was then neutralized with 50 uL of 100 mM HCl. The solution was spin filtered (Amicon Ultra Centrifugal Filters 3 kDa, 0.5 mL) to remove any particles and the flow-through was used for LC-HRMS analysis, which was normally performed using an Agilent Eclipse Plus C18 column (4.6 Å~ 100 mm). For ScoBCD experiments, a linear gradient of 10 to 50% $CH_3CN$ (v/v) over 12 min in $H_2O$ with 0.1% (v/v) formic acid, at a flow rate of 0.5 mL $min^{-1}$ was used. For MmaBCD experiments, a linear gradient of 20 to 98% $CH_3CN$ (v/v) over 18 min in $H_2O$ with 0.1% (v/v) formic acid, at a flow rate of 0.5 mL $min^{-1}$ was used.

MS and MS/MS data have been deposited to the EMBL-EBI MetaboLights[107] database (DOI: 10.1093/nar/gks1004. PubMed PMID: 23109552) with the identifier MTBLS454. The complete dataset can be accessed here http://www.ebi.ac.uk/metabolights/MTBLS454

# Chapter 3. Isonitrile Formation by a Non-Heme Iron(II)-Dependent Oxidase/Decarboxylase

Parts of this chapter have been adapted from the following with permission:
Harris, N. C., Born, D. A., Cai, W., Huang, Y., Martin, J., Khalaf, R., Drennan, C. L., Zhang, W. Isonitrile Formation by a Non-Heme Iron(II)-Dependent Oxidase/Decarboxylase. *Angew Chem Int Ed Engl* **57**(31) 9707-9710 (2018).

## 3.1 Introduction

The electron-rich functionality of the isonitrile lends itself as a biologically active warhead for naturally derived products. Due to its ability to coordinate transition metals, it is often exploited for metal acquisition, detoxification, and virulence.[42–44] Indeed the resume of potent biologically active isonitrile containing natural products is vast, and examples include xanthocillin, an antiviral agent;[108] rhabduscin, a virulence associated phenoloxidase inhibitor;[43] and many marine sponge derived metabolites. Despite the widespread utility of isonitrile in nature, its biosynthesis has long been considered endemic to the IsnA family of isonitrile synthases, which typically convert an α-amino group to isonitrile on an amino acid and require ribulose-5-phosphate as a co-substrate (**Fig 3-10**).[42,47,65,104,109]

Our recent genome mining of a conserved gene cluster widely present in Actinobacteria indicated an alternative route for isonitrile formation.[110] In particular, we identified and proposed the function of five genes required for the biosynthesis of a unique class of isonitrile lipopeptides (INLPs) that are critical for the virulence of pathogenic mycobacteria. Taking the pathway from *Streptomyces coeruleorubidus* as an example, the biosynthesis was proposed to start with the activation and loading of crotonic acid onto ScoB, an acyl carrier protein (ACP) by ScoC, an acyl-ACP ligase. A Michael addition of Gly to the β-position of crotonyl-ScoB is then promoted by ScoD, a thioesterase to form a Gly adduct **8**, followed by oxidation and decarboxylation, presumably catalyzed by ScoE, a non-heme iron(II)-dependent oxygenase, to generate a β-isonitrile fatty acyl-ACP intermediate **9**. This β-isonitrile acyl moiety is then condensed to both amino groups of Lys promoted by ScoA, a single-module non-ribosomal peptide synthetase (NRPS), and reductively released to form a terminal alcohol product **10** (**Fig 3-1**). ScoE thus represents a new family of enzymes distinct from isonitrile synthases that promote the transfer of one carbon from ribulose-5-phosphate to an amino group to form isonitrile. Although the function of ScoE was reconstituted in *E. coli* for **10** biosynthesis, *in vitro* reconstitution of its activity based on the proposed biosynthetic pathway repeatedly failed.

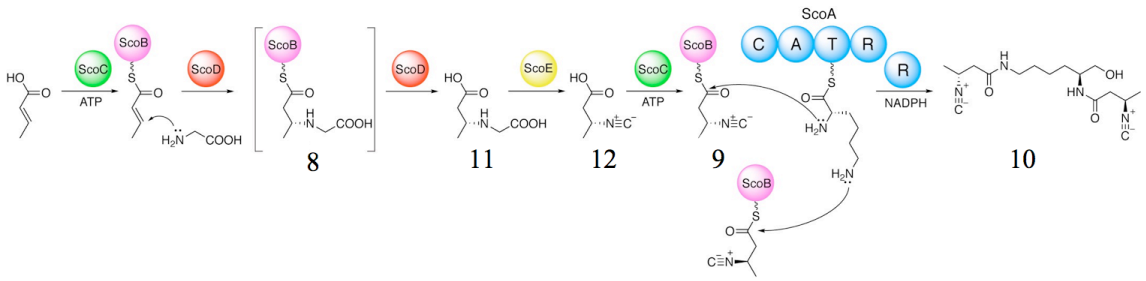



**Figure 3-1.** Schematic of isonitrile lipopeptide biosynthetic pathway. This revised pathway shows that ScoE utilizes a free acid substrate, **11**, and after isonitrile formation, the product **12** is reactivated and loaded onto ScoB by ScoC.

## 3.2 Results

We previously proposed that ScoE functions on an ACP-bound intermediate because biochemical analysis of pathway enzymes showed that the formation of **1** requires ScoB, and the action of the NRPS, ScoA, also requires a ScoB-bound substrate. The proposed pathway thus accounts for the necessity and sufficiency of these five core biosynthetic enzymes for INLP synthesis (**Fig 3-1**). However the recent biochemical and structural characterization of three ScoD homologues indicated that these thioesterases have dual functions with enzymatic hydrolysis occurring immediately after the Michael addition, both steps mediated by a single Gly residue in the active site.[106,111,112] These results raised the question of what the true substrate of ScoE is. We thus initiated an effort to reconstitute the *in vitro* activity of ScoE using various substrates that were chemically or chemoenzymatically synthesized. In particular, we chemically synthesized (*R*)-3-((carboxymethyl)amino)butanoic acid (CABA) **(11)** and CABA-CoA that was used alone or with the phosphopantetheinyl transferase from *Bacillus subtilis* (Sfp) to form CABA-ScoB **(8)** (**Appendix H-J**). Initial attempts to reconstitute the activity of ScoE were unsuccessful regardless of the substrate used.

Meanwhile, we obtained an X-ray crystal structure of ScoE to 1.8 Å-resolution (**Fig 3-3, 3-4 and Table 3-1**). The structure of ScoE is similar to the TauD family of non-heme iron(II) enzymes with root mean square deviation (RMSD) values of 2.1-2.7 Å (for $C_\alpha$ atoms) to structurally characterized TauD enzymes, although the sequence identity is only 20-28% to these same enzymes (**Fig 3-5**).[113] The 2-His-1-Asp facial triad (H132, D134, and H295) and an Arg (R310) within the substrate binding site are conserved with TauD family members. Putative ScoE substrate and αKG binding sites appear occupied by exogenous ligands (**Fig 3-3, A and C**). Particularly, a Zn(II) is bound by the 2-His-1-Asp facial triad (H132, D134, and H295). The fourth coordination position on the Zn(II) is occupied by an acetate molecule that is acquired from the crystallization condition and is bound analogously to αKG in the structure of TauD (**Fig 3-3 2, C and D**).[114,115] In the putative substrate binding pocket, a $Cl^-$ is bound in a position similar to that of the sulfonate moiety of taurine in a substrate-bound structure of TauD (**Fig 3-3, C and D**).[114,115] Adjacent to the $Cl^-$ is electron density that is consistent with a molecule of choline, oriented such that the positive charge of the trimethylamine moiety is pointing toward $Cl^-$ (**Figure 3-3C**).

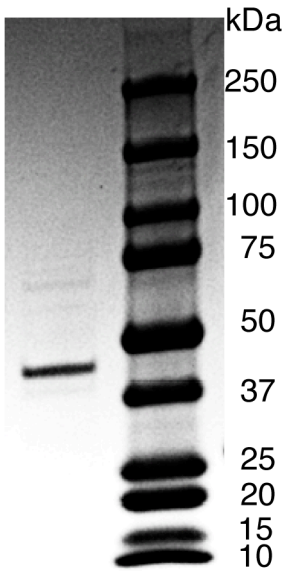



**Figure 3-2.** Protein gel of ScoE purified from M9 medium. Criterion Tris-HCl gel (4-15% precast, Biorad) was used.

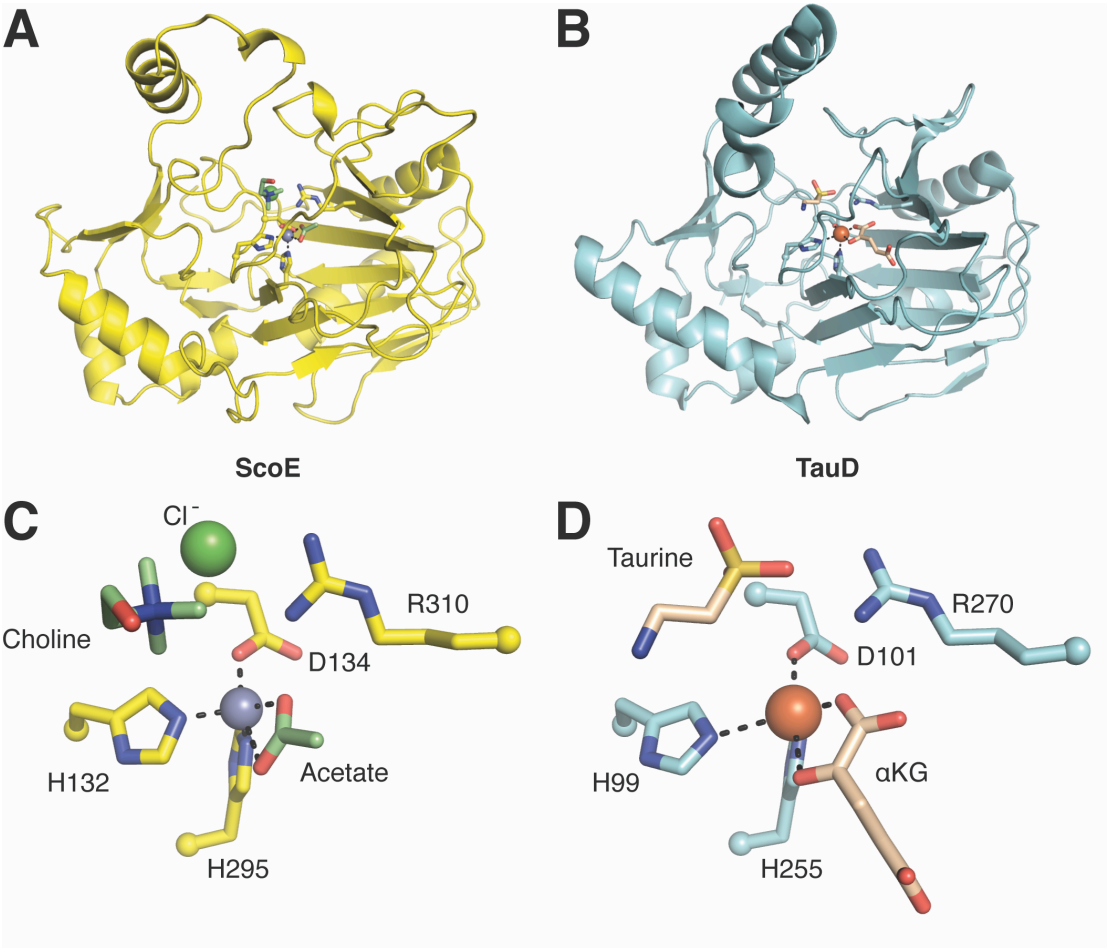



**Figure 3-3.** Structure of ScoE (PDB 6DCH) compared to TauD (PDB 1OS7).[15] A) Overall structure of a ScoE protomer shown in yellow ribbon representation. The metal-coordinating 2-His-1-Asp facial triad and a conserved active site Arg are shown in yellow ball-and-stick representation. Acetate and choline ligands within the active site are shown in green ball-and-stick. Only one of the two observed conformations of choline is shown for clarity. Zn(II) and $Cl^-$ are shown as gray and green spheres, respectively. B) Overall structure of a TauD protomer shown in teal ribbon representation and oriented similarly to ScoE in A). The metal-coordinating 2-His-1-Asp facial triad and a conserved active site Arg are shown in teal ball-and-stick representation. Taurine and αKG within the active site are shown in tan ball-and-stick. Fe(II) is shown as an orange sphere. C) Zoomed in view of the active site of ScoE. D) Active site of TauD displayed similarly to C).

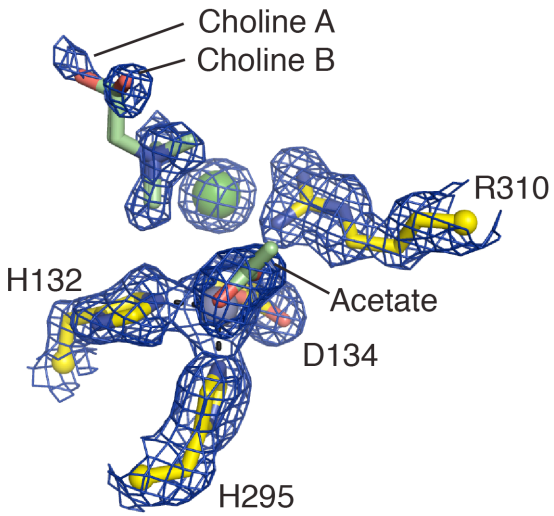



**Figure 3-4.** Electron density of the ScoE active site. Composite omit electron density contoured at 1 σ is shown in blue mesh. Conserved active site residues and active site ligands are shown in ball-and-stick representation. Chloride and Zn(II) and shown as green and gray spheres, respectively. Two conformations of choline are modeled, shown in green ball-and-sticks carbon atoms.

**Table 3-1.** Crystallographic data collection and refinement statistics for ScoE PDB 6DCH.

| **Data collection** | |
|---|---|
| Space group | $P4_32_12$ |
| Cell dimensions | |
| *a, b, c* (Å) | 62.1, 62.1. 167.9 |
| Wavelength (Å) | 0.9791 |
| Resolution (Å)‡ | 58.2-1.8 (1.86-1.8) |
| $R_{meas}$ (%)‡ | 22.0 (164) |
| $R_{pim}$ (%)‡ | 4.5 (32.0) |
| CC1/2‡ | (0.72) |
| $I/\sigma_I$‡ | 14.2 (2.6) |
| Completeness (%)‡ | 99.0 (100) |
| Redundancy‡ | 24.4 (26.2) |
| No. of unique reflections‡ | 31084 (3080) |
| **Refinement** | |
| Resolution (Å) | 1.8 |
| No. of reflections used | 31077 |
| Average B-factors (Å$^2$) | |
| Protein | 23.0 |
| Acetate | 27.3 |
| Choline | 20.0 |
| Chloride | 19.9 |
| Zinc | 18.5 |
| Water | 34.5 |
| $R_{work}/R_{free}$ (%)§ | 17.9/20.7 |
| No. atoms | |

| | |
|---|---|
| Protein | 2336 |
| Acetate | 4 |
| Betaine | 14 |
| Chloride | 1 |
| Zinc | 1 |
| Water | 331 |
| R.m.s. deviations | |
| Bond lengths (Å) | 0.007 |
| Bond angles (°) | 1.22 |
| Rotamer outliers (%) | 0 |
| Ramachandran angles | |
| Outlier (%) | 0 |
| Allowed (%) | 1.0 |
| Favored (%) | 99.0 |

‡ Values in parentheses indicate highest resolution bin.

§ Reflections used to calculate $R_{free}$ comprise 6% of observed reflections.

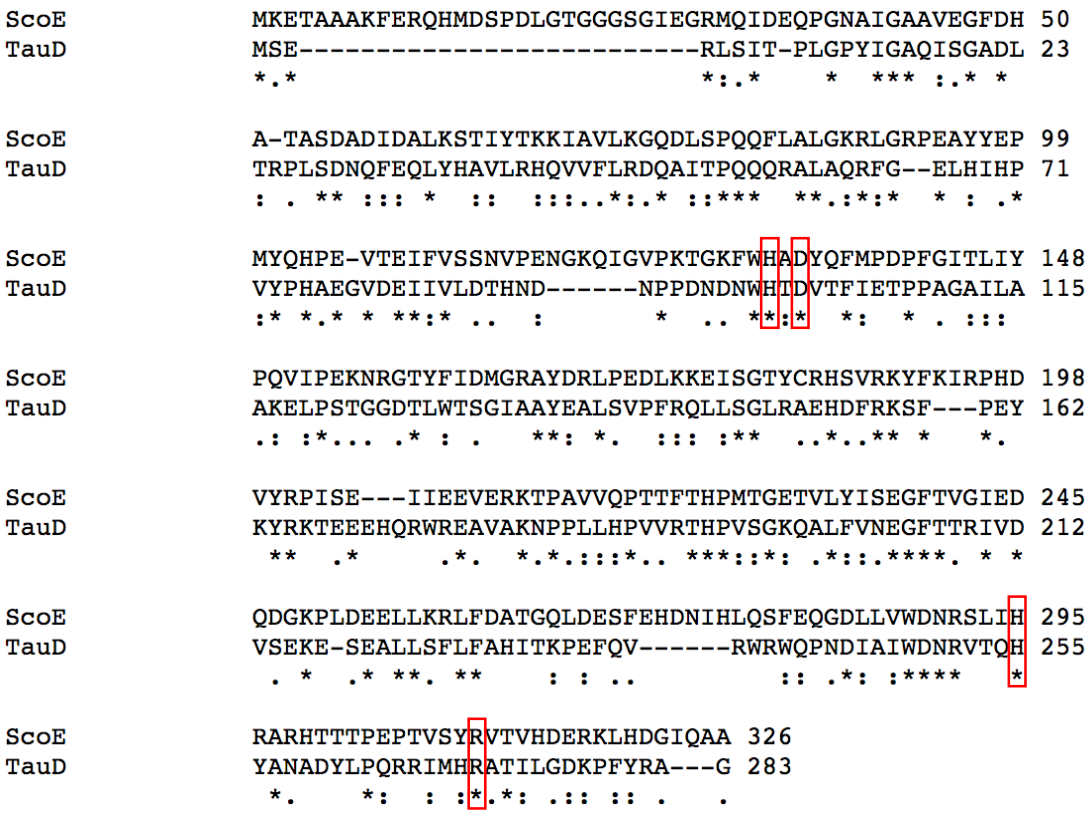



**Figure 3-5.** Amino acid sequence alignment of ScoE PDB 6DCH and TauD PDB 10S7. Key catalytic residues are highlighted with red boxes. Analysis was performed using T-Coffee.[116,117]

Since choline and Zn(II) were identified within the crystal structure of ScoE but absent from the crystallization condition, we hypothesized that choline and Zn(II) were co-purified with ScoE during protein purification and they could potentially interfere with substrate binding and Fe(II) reconstitution of the holo-enzyme. To mitigate the problem of unwanted choline that may be abundant in LB media used for protein purification,[118] we then used M9 defined medium for future purifications of ScoE (**Figure 3-2**). ScoE was reconstituted immediately before each assay with fresh Fe(II) to form the holo-protein.

We then incubated holo-ScoE with αKG, ascorbate, and either CABA-ScoB (**8**), CABA (**11**), or CABA-CoA. Upon incubation of ScoE with **11** and subsequent LC-HRMS analysis, we observed a mass spectrum associated with the formation of (*R*)-3-isocyanobutanoic acid **(5)** (**Fig 3-6A and Appendix K**). This product was not observed when either **11**, αKG, or ScoE was omitted, or when boiled ScoE was added to the reaction. To further confirm this result, we synthesized (*R*)-3-((carboxymethyl)amino)butanoic-5-$^{13}$C-acid as a substrate (**Appendix L**). When ScoE was incubated with this labeled substrate, the expected mass spectral shift of the product was observed (**Appendix K**). This mass spectrum was observed only when all necessary components of the ScoE reaction were included and only when the labeled substrate was utilized. We also confirmed the identity of the product **12** from the enzymatic reaction by comparing to a chemically synthesized standard (**Fig 3-6A**). In addition, the presence of the unique isonitrile functionality in **12** was confirmed by click reactions with tetrazines (**Fig 3-7)**. The absolute configuration of the C3 of the enzymatic product **12** was determined to be *R*, demonstrating that the chirality at this position was retained during the ScoE-catalyzed reaction (**Fig 3-8**). We further determined the kinetic parameters of ScoE toward **11** ($K_{\mathbf{m}}$ = 286 ± 93 μM**,** $k_{cat}$ = 21.9 ± 2.1 min$^{-1}$) and αKG ($K_{\mathbf{m}}$ = 20.9 ± 3.2 μM**,** $k_{cat}$ = 0.50 ± 0.02 min$^{-1}$) using LC-MS to monitor the formation of **12** and the succinate formation assay to monitor NADH oxidation, respectively (**Fig 3-9)**. No activity of ScoE was observed when **8** or CABA-CoA was used as a substrate.

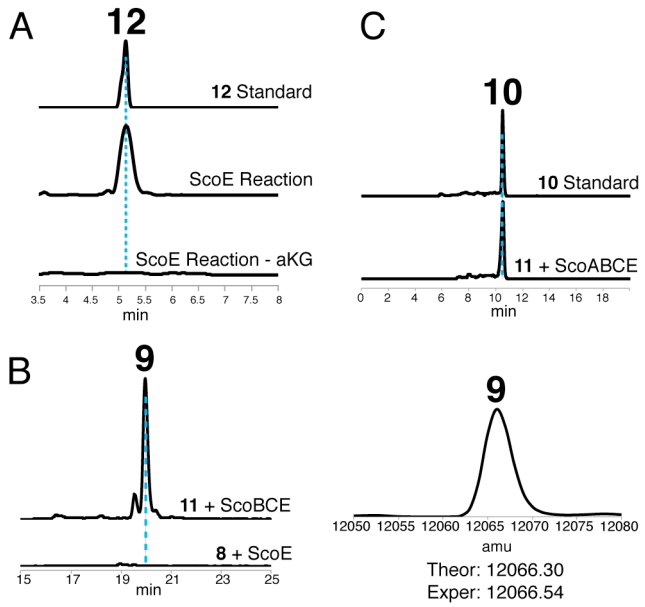



**Figure 3-6.** *In vitro* characterization of ScoE. A) Extracted ion chromatograms showing the conversion of **11** to **12** catalyzed by ScoE. For simplicity, only the no αKG assay is shown as a representative negative control. The calculated mass of **12** with a 10-ppm mass error tolerance was used. B) Extracted ion chromatograms showing the production of 2. Bottom trace displays the assay with 1 and ScoE. Top trace shows the coupled reaction containing ScoE, ScoB, ScoC and **11**. The calculated mass of **9** with a 10-ppm mass error tolerance was used. The deconvoluted mass spectrum of **9** is displayed on the right. C) Extracted ion chromatograms showing the formation of **10** in the total enzymatic synthesis using ScoA, ScoB, ScoC, ScoE and **11**. The calculated mass of **10** with a 10-ppm mass error tolerance was used.

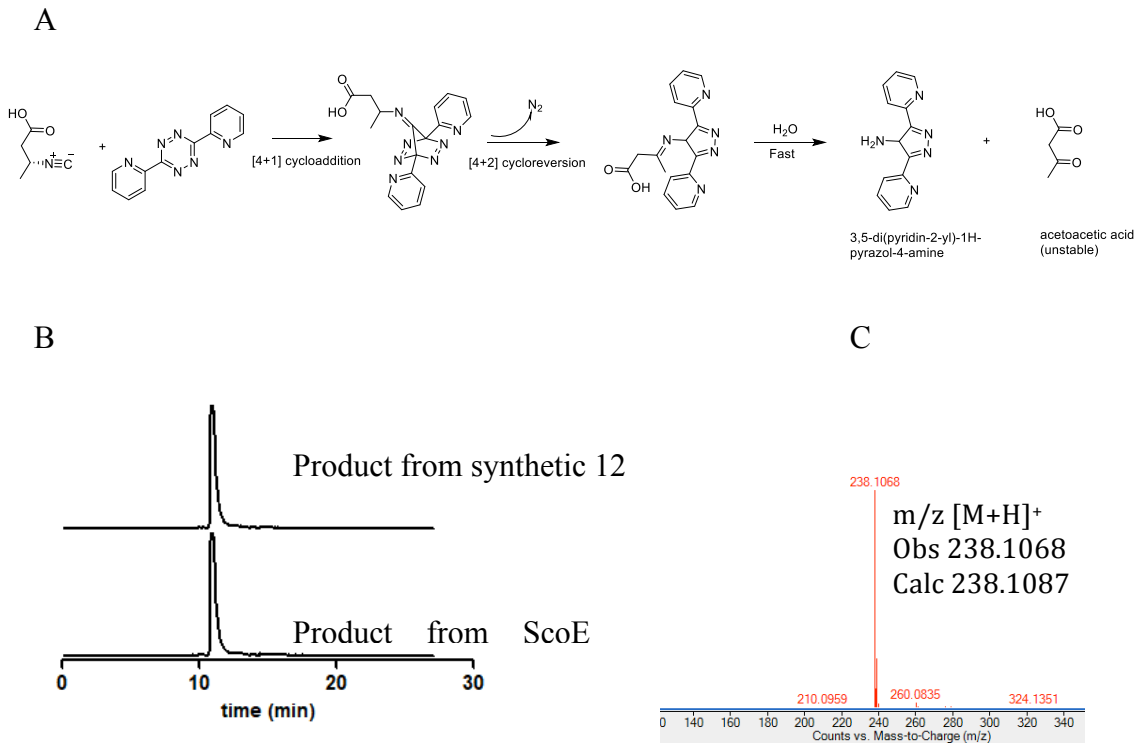



**Figure 3-7.** Tetrazine click reaction with synthetic and ScoE derived **12**. A) Schematic of [4+1] cycloaddition between **12** and tetrazine. B) LC-MS analysis of isonitrile-tetrazine click chemistry. upper trace: 3,5-di(pyridin-2-yl)-1H-pyrazol-4-amine was formed from click reaction using synthesized **12**; lower trace: 3,5-di(pyridin-2-yl)-1H-pyrazol-4-amine was formed from click reaction using ScoE + **11** produced **12**. C) LC-HRMS analysis of 3,5-di(pyridin-2-yl)-1H-pyrazol-4-amine derived from synthesized **12**.

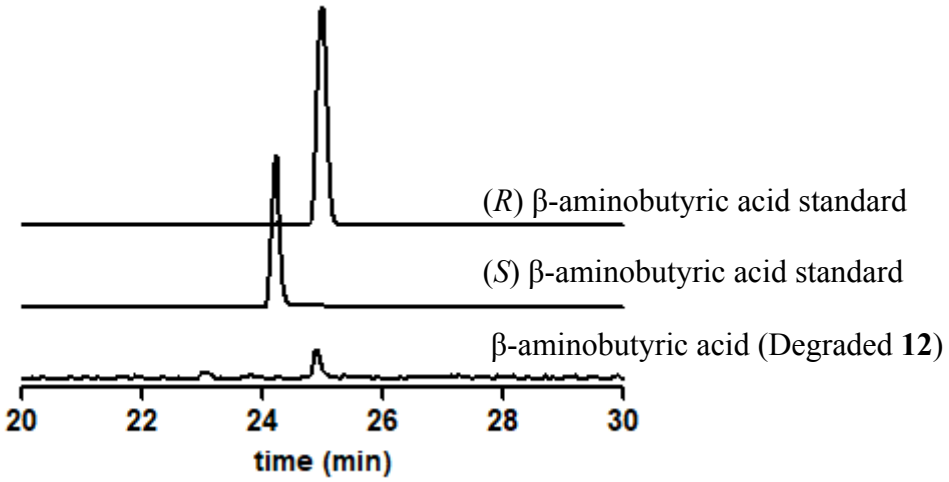



**Figure 3-8.** Chirality of **12** from ScoE *in vitro* reaction determined by Marfey's analysis. Chirality of (*R*)-3-((carboxymethyl)amino)butanoic acid (**11**) is preserved during ScoE

catalyzed formation of (*R*)-3-isocyanobutanoic acid (**12**). β-aminobutyric acid was obtained from the degradation of **12** generated from ScoE *in vitro* reaction and compared to (*R*) and (*S*) standards of β-aminobutyric acid after reacting with Marfey's reagent and subsequent LC-MS analysis.

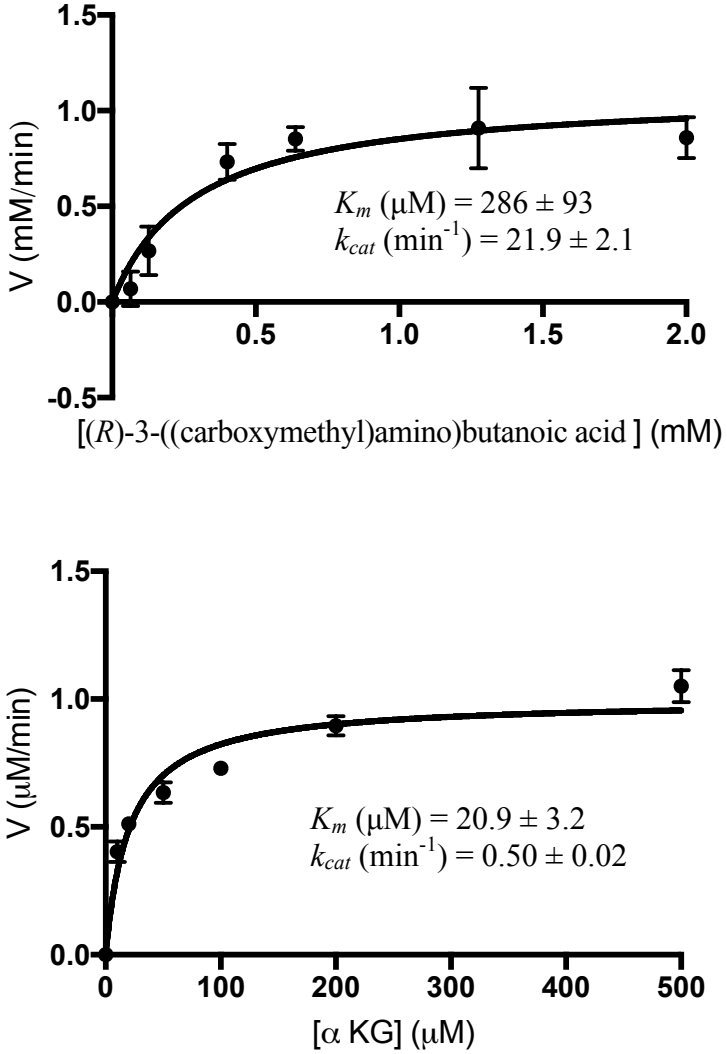



**Figure 3-9.** ScoE kinetic parameters determined for **11** and αKG. Parameters were determined for **11** (top panel), with 50 μM ScoE and the concentration of αKG fixed at 10 mM. Formation of the product **12** was monitored by LC-MS. Parameters were determined for αKG (bottom panel), with 2 μM ScoE and the concentration of **11** fixed at 1 mM. Succinate formation was monitored using a coupled NADH oxidation reaction and monitoring at 340 nm. Rates of reactions were calculated by subtracting the values from identical reactions lacking ScoE. Kinetic parameters were determined and plotted using Prism. Error bars represent standard deviations from three independently performed experiments.

Our *in vitro* biochemical analysis provided direct evidence for a second mechanism of isonitrile formation by a non-heme iron(II) and αKG dependent oxidase/decarboxylase. We propose that ScoE functions similarly to TauD and TfdA,[114,119] utilizing an enzyme-bound iron-oxo species for oxidation of **11** which likely goes through two sequential steps with an imine intermediate (**Fig 3-10**). The position of the choline hydroxyl group and $Cl^-$ in our structure of ScoE may indicate the binding mode of the two carboxylate groups of **11**. However, a high-resolution structure of ScoE with substrate bound is necessary to determine the precise substrate binding arrangement and provide insight in substrate activation, which is currently under way.

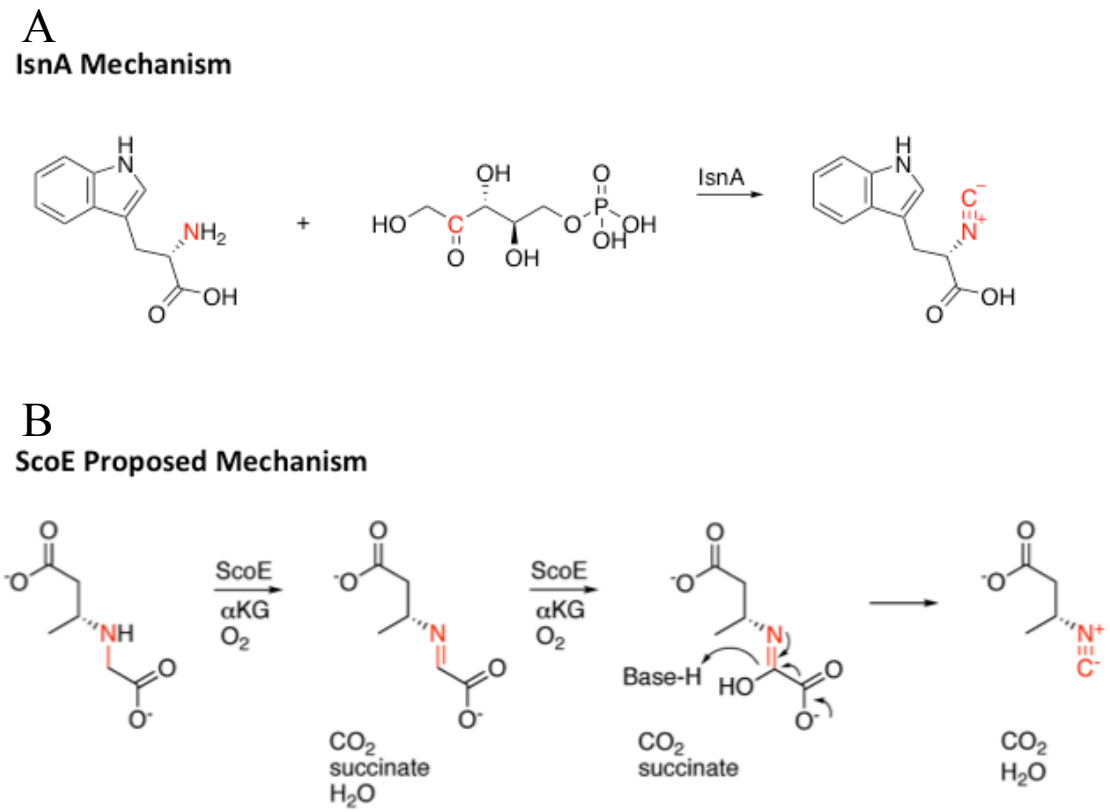



**Figure 3-10.** A) IsnA catalyzed formation of isonitrile.[58],[109] Isonitrile synthases typically convert an α-amino group to isonitrile on an amino acid and require ribulose-5-phosphate as a co-substrate. B) Proposed mechanism for isonitrile formation catalyzed by ScoE. ScoE likely functions similarly to TauD and TfdA, utilizing an enzyme-bound iron-oxo species for oxidation of **11** which likely goes through two sequential steps with an imine intermediate.

The reconstitution of ScoE activity using the free acid substrate, **11**, raised additional questions about the function of enzymes in INLP biosynthesis. We have previously shown that the NRPS, ScoA, requires a ScoB-bound substrate for the subsequent amide bond forming condensation reactions. Our *in vitro* ScoE system however utilized only **11**, which after forming isonitrile product **12**, would need to be activated and loaded onto ScoB again for ScoA to function (**Fig 3-1**). We then tested whether the acyl-ACP ligase, ScoC, could function again after ScoE to activate **12**. We first conducted the ScoE *in vitro* reaction and after a short incubation period, added ScoB, ScoC, and ATP. The LC-

HRMS analysis showed a strong signal for the formation of **9**, which was absent when **8** was directly used as a substrate for the reaction of ScoE (**Fig 3-6B**). This result indicated that ScoC functions twice in the pathway, to first activate and load crotonic acid, and subsequently to activate and load **5** onto ScoB to provide a preferred isonitrile substrate for ScoA (**Fig 3-1**). To further support the proposed functions of enzymes in INLP biosynthesis, we next performed an *in vitro* total enzymatic synthesis of INLP using purified enzymes including ScoA, ScoB, ScoC and ScoE. After incubating enzymes with **11,** ATP, Lys, and NADPH, the expected product **10** was successfully produced as compared to a chemical standard based on LC-HRMS analysis (**Fig 3-6C**).

## 3.3 Discussion

In summary, with the aid of a high-resolution crystal structure, we were able to circumnavigate inhibitory circumstances and biochemically reconstitute the activity of ScoE, a non-heme iron(II)-dependent enzyme, for isonitrile synthesis for the first time. We demonstrated that ScoE catalyzes the formation of isonitrile via the oxidative decarboxylation of the free acid substrate, **11**. We also provided evidence for a revised pathway for INLP synthesis with the second role of a promiscuous acyl-ACP ligase, ScoC, which activates the isonitrile product of ScoE before the NRPS-promoted INLP formation. This revision is expected to be applicable to other homologous INLP biosynthetic pathways found in Actinobacteria. This work paves the way to elucidate the enigmatic enzymatic mechanism for isonitrile formation using a non-heme iron(II)-dependent enzyme, of which homologues are conserved and critical for the virulence of pathogenic mycobacteria, including *M. tuberculosis*.

## 3.4 Materials and Methods

### 3.4.1 Overexpression and purification of ScoE for crystallization

Strain construction, protein solubility, and yield was described previously.[110] Expression and purification of ScoE with a His6-tag was conducted as follows: cells were grown at 37 °C in 1 L of LB with kanamycin to an OD600 of 0.5. The cells were then cooled on ice for 10 min and induced with 0.12 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 16 h at 16 °C. Subsequently, the cells were harvested by centrifugation (6371 × g, 15 min, 4 °C), resuspended in 30 mL lysis buffer (25 mM HEPES, pH 8, 500 mM NaCl, 5 mM imidazole), and lysed by homogenization on ice. Cellular debris was removed by centrifugation (27216 × g, 1 h, 4 °C). Ni-NTA agarose resin was added to the supernatant (1.5 mL/L of culture), and the solution was nutated at 4 °C for 1 h. The protein-resin mixture was loaded onto a gravity flow column, and proteins were eluted with increasing concentrations of imidazole in 50 mM HEPES, pH 8.0, 100 mM NaCl. Purified proteins were concentrated using Amicon Ultra spin filters to yield 3 mL of protein. Proteins were dialyzed at 4 °C using 10 kDa Slide-A-Lyzer Dialysis Cassettes in 1 L dialysis buffer (25 mM HEPES, pH 8.0, 100 mM NaCl, 1 mM EDTA). The dialysis buffer was changed twice over 9 h and then proteins were dialyzed overnight in the same buffer. After dialysis, proteins were desalted to remove EDTA using a GE Pd10 and eluted into 25 mM HEPES, pH 8.0, 100 mM NaCl. Proteins were concentrated using Amicon Ultra spin filters until the protein concentration reached 40 mg/mL. Proteins were flash frozen in liquid nitrogen and stored at -80 °C.

### 3.4.2 Overexpression and purification of ScoE for *in vitro* assays

Cells were grown at 37 °C in 1 L of LB with kanamycin to an OD600 of 0.5. Cells were pelleted by centrifugation (6371 × g, 10 min, 4 °C), and resuspended in 1 L M9 minimal medium containing kanamycin and 0.12 mM IPTG for induction. Cells were grown for 16 h at 16°C. Subsequently, the cells were harvested by centrifugation (6371 × g, 15 min, 4 °C), resuspended in 30 mL lysis buffer (25 mM HEPES, pH 8, 500 mM NaCl, 5 mM imidazole), and lysed by homogenization on ice. Cellular debris was removed by centrifugation (27216 × g, 1 h, 4 °C). Ni-NTA agarose resin was added to the supernatant (1.5 mL/L of culture), and the solution was nutated at 4 °C for 1 h. The protein-resin mixture was loaded onto a gravity flow column, and proteins were eluted with increasing concentrations of imidazole in 50 mM HEPES, pH 8.0, 100 mM NaCl. Purified proteins were concentrated using Amicon Ultra spin filters to yield 3 mL of protein. Proteins were dialyzed at 4 °C using 10 kDa Slide-A-Lyzer Dialysis Cassettes in 1 L dialysis buffer (25 mM HEPES, pH 8.0, 100 mM NaCl, 1 mM EDTA). The dialysis buffer was changed twice over 9 h and then proteins were dialyzed overnight in the same buffer. After dialysis, proteins were concentrated using Amicon Ultra spin filters. 10% v/v glycerol was added and proteins were flash frozen in liquid nitrogen and stored at -80 °C.

### 3.4.3 Reconstitution of ScoE to form the holo-enzyme

Immediately before each assay, a Bio-Rad Bio-Gel P-6 gel column was equilibrated in 25 mM HEPES, pH 8.0, 100 mM NaCl. 20-75 uL ScoE was desalted to remove EDTA following the manufacturer's protocol. ScoE was reconstituted by adding the following in the order listed: 1 mM DTT and 0.75 mM ammonium iron(II) sulfate hexahydrate.

### 3.4.4 Crystallization of ScoE

Crystals of metal chelated ScoE were obtained by the sitting drop vapor diffusion technique at 18 ºC. A 120 nL aliquot of protein solution (8 mg/mL ScoE in Buffer A, 50 mM HEPES pH 8 and 100 mM NaCl) was added to 120 nL of precipitant solution (200 mM sodium acetate pH 7, 100 mM Tris pH 8.5, 30% (w/v) PEG 4,000) in a 96-well sitting drop tray using a PHOENIX liquid handling robot (Art Robbins Instruments). Crystals formed within 1 wk and reached a maximum size after approximately 2 wk. Paraffin oil was used as a cryoprotectant during crystal harvesting. Crystals were manually looped and streaked through paraffin oil before being flash-frozen in liquid $N_2$.

### 3.4.5 Data collection and processing of ScoE crystals

Data for ScoE were collected at the Advanced Photon Source (Argonne, Illinois, USA) on beam line 24ID-C using a Pilatus 6M pixel detector at a temperature of 100 K. Crystals of ScoE belong to space group $P4_32_12$ and data were collected at 0.9791 Å in a single 360º wedge. Data extending to 1.8 Å-resolution were processed in XDS and scaled in XSCALE.[120]

### 3.4.6 Structure determination and refinement

The structure of ScoE was determined by molecular replacement using Phaser[121] with a homology model of TauD from *Pseudomonas putida* KT2440 (PDB 3PVJ)

generated using Sculptor.[122] Molecular replacement was restricted to data extending to 3 Å-resolution and yielded one ScoE protomer within the asymmetric unit.

Following molecular replacement, data was extended to 1.8 Å-resolution and AutoBuild[123] was run to assist in initial model building. To produce a final model, iterative rounds of model building and refinement were performed in Coot[124] and Phenix,[125] respectively. First rounds of refinement included simulated annealing to mitigate model bias from molecular replacement. Waters were added automatically using Phenix. Final ScoE refinements in Phenix yielded a working R-factor of 18% and free R-factor of 21%. The final model was confirmed using simulated annealing composite omit electron density maps generated in Phenix. The ScoE structure contains residues 28-323 (of 326 residues), a Zn(II) in the iron binding site, an acetate molecule at the α-ketoglutarate binding position on the metal, a $Cl^-$ in the presumed substrate-binding pocket, and two conformations of a choline molecule in the substrate-binding pocket. The identities of metals were assigned based on their local environment and coordination geometry. The identity of choline was determined by trial-and-error of various *E. coli* metabolites and evaluation of the fit to omit electron density maps. Restraint files for choline were generated on the Grade web-server[126] and each conformation was refined at half occupancy. The ScoE structure was deposited using the PDB identifier 6DCH.

### 3.4.7 General NMR and purification of synthetic materials

NMR spectra were recorded with a Bruker AVANCE spectrometer at 900 MHz or 800 MHz ($^1$H NMR), and 226 MHz or 201 MHz ($^{13}$C NMR). (*R*)-3-aminobutanoic acid was purchased from Ark Pharm Inc.; bromoacetic acid-2-$^{13}$C and NMR solvents were purchased from Cambridge Isotope Laboratories Inc. Other chemicals, solvents were purchased from Sigma-Aldrich. HPLC purification was carried out using Agilent 1200 systems and a Grace Alltima C18 column (150 x 10 mm). LC-HRMS was conducted using an Agilent Technologies 6510 Accurate-Mass Q-TOF.

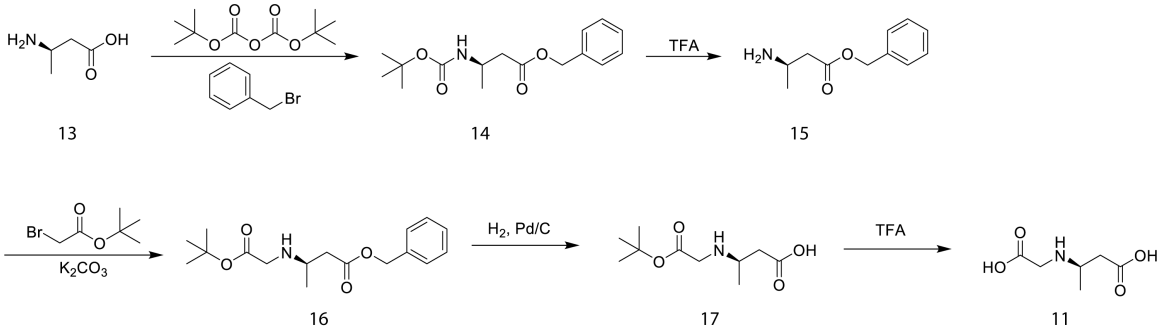



**Figure 3-11.** Chemical synthesis of (*R*)-3-((carboxymethyl)amino)butanoic acid **11.**

### 3.4.8 Synthesis of 11

(*R*)-3-aminobutanoic acid (**13**) (0.52 g, 5 mmol) was dissolved in a solution of water (10 ml) and *tert*-butyl alcohol (15 ml), then sodium carbonate (1.59 g, 15 mmol) and di-*tert*-butyl dicarbonate (1.64 g, 7.5 mmol) were added. The reaction was stirred and refluxed for 2 h. The mixture was cooled to 0 °C in an ice bath and acidified with 6 N HCl to pH 1.0. Alcohol was removed *in vacuo* and the aqueous phase was extracted with

EtOAc 3 times (25 ml each). The organic fractions were combined and washed with water, brine and then dried over $Na_2SO_4$, and solvent was removed to yield *N*-boc-(*R*)-3 aminobutanoic acid, a colorless oily liquid.[127] The crude product was dissolved in acetone (10 ml), then triethylamine (1.4 ml, 10 mmol) and benzyl bromide (0.7 ml, 5.5 mmol) were added sequentially. The reaction was stirred overnight at room temp. Water was added to quench the reaction and acetone was removed *in vacuo*. The aqueous phase was extracted with EtOAc 3 times (25 ml each). The organic fractions were combined and washed with water, brine and then dried over $Na_2SO_4$, and solvent was removed to yield benzyl *N*-boc-(*R*)-3 aminobutanoic acid (**7**) as a colorless oily liquid. The product was purified by flash silica chromatography. The overall yield was 40%. **7**: calculated for $C_{16}H_{24}NO_4$ $[M+H]^+$ 293.1; found 293.1. $^1$H NMR (900 MHz, $CD_3OD$) δ 7.26 – 7.21 (m, 5H), 5.02 – 4.97 (m, 2H), 3.89 (m, 1H), 2.44 (dd, *J* = 15.1, 6.9 Hz, 1H), 2.34 (dd, *J* = 15.1, 6.7 Hz, 1H), 1.30 (s, 9H), 1.03 (d, *J* = 6.7 Hz, 3H); $^{13}$C NMR (226 MHz, $CD_3OD$) δ 172.82, 157.62, 137.65, 129.64, 129.36, 129.30, 80.11, 67.43, 45.12, 42.30, 28.91, 20.99. The physical spectrum date matches previous data.

**14** (290 mg, 1 mmol) was dissolved in DCM (10 ml) and cooled to 0 °C in an ice bath, and trifluoroacetic acid (TFA, 5 ml) was added dropwise. The mixture was allowed to warm to room temp and stirred for 2 h and monitored by TLC. Solvent was removed *in vacuo* to yield benzyl (*R*)-3 aminobutanoic acid (**15**) as a colorless oily liquid. Yield was 100%. **15**: calculated for $C_{11}H_{16}NO_2$ $[M+H]^+$ 194.1176; found 194.1156. $^1$H NMR (800 MHz, $CDCl_3$) δ 7.36 – 7.27 (m, 5H), 5.10 (dd, *J* = 32.0, 12.2 Hz, 2H), 3.67 (m, 1H), 2.78 (dd, *J* = 17.5, 7.8 Hz, 1H), 2.66 (dd, *J* = 17.5, 4.8 Hz, 1H), 1.33 (d, *J* = 6.7 Hz, 3H); $^{13}$C NMR (201 MHz, $CDCl_3$) δ 171.54, 135.16, 128.86, 128.78, 128.55, 67.46, 45.06, 37.86, 18.39.

**15** (193 mg, 1 mmol) was dissolved in a mixture of acetonitrile (5 ml) and 2 M sodium diphosphate buffer (2.5 ml, pH 8.0), then *tert*-butyl-bromoacetate (234 mg, 1.2 mmol) was added. The reaction was stirred overnight at room temp. The organic solvent was removed *in vacuo* and the aqueous phase was extracted with EtOAc 3 times (25 ml each). The organic fractions were combined and washed with water, brine and then dried over $Na_2SO_4$, and solvent was removed to yield crude benzyl (*R*)-3-((2-(*tert*-butoxy)-2-oxoethyl)amino)butanoate (**16**). The crude product was purified using HPLC (a linear gradient of 2 to 95% acetonitrile over 30 min in water with 0.1% TFA) to yield the final product, a colorless oily liquid. The yield was 25%. **16**: calculated for $C_{17}H_{26}NO_4$ $[M+H]^+$ 308.1856; found 308.1836.

**16** (307 mg, 1 mmol) was dissolved in ethanol and 10% Pd/C (15 mg) was added. The reaction was stirred under a hydrogen atmosphere (1 atm) at room temp for 2 h. The solvent was removed *in vacuo* to yield (*R*)-3-((2-(*tert*-butoxy)-2-oxoethyl)amino)butanoate (**17**), a colorless oily liquid. **17:** calculated for $C_{10}H_{20}NO_4$ $[M+H]^+$ 218.1; found 218.1. $^1$H NMR (900 MHz, DMSO-$d_6$) δ 3.45 (d, *J* = 17.0 Hz, 1H), 3.40 (d, *J* = 17.0 Hz, 1H), 3.06 (m, 1H), 2.31 (dd, *J* = 15.8, 6.9 Hz, 1H), 2.24 (dd, *J* = 15.8, 6.0 Hz, 1H), 1.43 (s, 9H), 1.05 (d, *J* = 6.4 Hz, 3H); $^{13}$C NMR (226 MHz, DMSO) δ 173.22, 169.78, 81.05, 49.64, 47.17, 39.76, 27.73, 18.80.

**17** was dissolved in DCM (10 ml) and cooled to 0 °C in an ice bath, and trifluoroacetic acid (5 ml) was added dropwise. The mixture was allowed to warm to room temp and stirred for 2 h and monitored by TLC. Solvent was removed *in vacuo* to yield **11,** a colorless oily liquid. The yield was 98%. **11:** calculated for $C_6H_{12}NO_4$

$[M+H]^+$ = 162.0761, found: 162.0762. $^1$H NMR (900 MHz, DMSO-$d_6$) δ 3.77 (s, 2H), 3.51 (dd, *J* = 13.2, 6.6 Hz, 1H), 2.81 (dd, J = 16.6, 4.4 Hz, 1H), 2.48 (s, 1H), 1.23 (d, *J* = 6.6 Hz, 3H); $^{13}$C NMR (226 MHz, DMSO) δ 171.70, 168.51, 50.16, 49.94, 45.23, 16.30 (**Appendix H**).

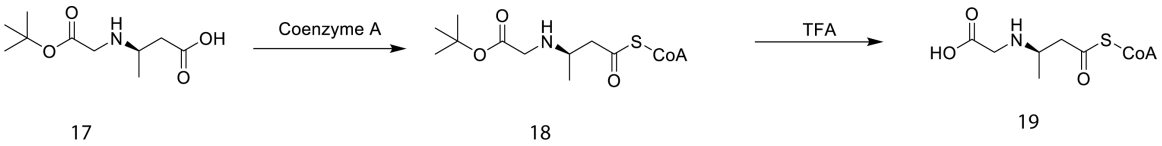



**Figure 3-12.** Chemical synthesis of CABA-CoA **(19).**

### 3.4.9 Synthesis of 19

**17** (21.7 mg, 0.1 mmol), CoA sodium salt (76.7 mg, 0.1 mmol), PyBOP (104.0 mg, 0.2 mmol) and potassium carbonate (55.3 mg, 0.4 mmol) were added to 4 ml of a tetrahydrofuran water solution (1:1). The mixture was stirred for 2 h at room temp. The reaction was purified by HPLC (a linear gradient of 2 to 95% acetonitrile over 30 min in water with 0.1% TFA). Fractions containing (*R*)-3-((2-(*tert*-butoxy)-2-oxoethyl)amino)butanyl-CoA (**18**) were dried *in vacuo* to yield the product, a white amorphous powder. **18:** calculated for $C_{31}H_{55}N_8O_{19}P_3S$ $[M+H]^+$ = 967.2433, found 967.2424. **18** was dissolved in 4 ml of a tetrahydrofuran water solution (1:1) and cooled to 0 °C in an ice bath, and trifluoroaetic acid (1 ml) was added dropwise. The reaction was stirred for 2 h at room temp and purified by HPLC as described above. (*R*)-3-((carboxymethyl)amino)butanyl-CoA (**19**) was dried *in vacuo* to yield a white amorphous powder.[128] **19:** calculated for $C_{27}H_{47}N_8O_{19}P_3S$ $[M+H]^+$ 911.1807; found 911.1807 (**Appendix I**).

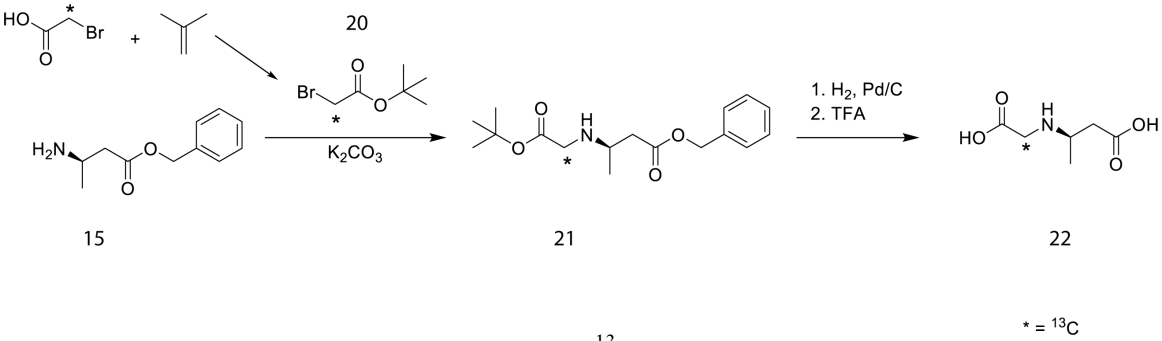



**Figure 3-13.** Chemical synthesis of CABA-5-$^{13}$C (**22**).

### 3.4.10 Synthesis of 22

An autoclave tube charged with bromoacetic acid-2-$^{13}$C (1 g, 6.8 mmol) and Amberlyst-15 resin (5 mg) was cooled to -78 °C and precooled isobutene (2 ml, 21 mmol) was added. The autoclave tube was sealed and allowed to slowly warm to room temp and the reaction was stirred for 24 h. Amberlyst-15 resin was filtered and unreacted isobutene was evaporated to yield *tert*-butyl bromoacetate-2-$^{13}$C (**20**), a colorless liquid.[129] The yield was 98%. $^1$H NMR (900 MHz, $CDCl_3$) δ 3.67 (s, 2H), 1.39 (S, 9H); $^{13}$C NMR (226 MHz, $CDCl_3$) δ 166.40, 166.20, 81.05, 27.77, 27.68. The synthetic procedure of CABA-

5-$^{13}$C (**22**) was identical to that of **11**, described above, except that **20** was used. **22:** calculated for $C_6H_{12}NO_4$ (5-$^{13}$C) $[M+H]^+$ 163.0794, found: 163.0794 (**Appendix L**).

### 3.4.11 Enzymatic synthesis and LC-HRMS analysis of 8 and coupled reaction with ScoE

**8** was synthesized enzymatically immediately before using in additional assays or in a coupled reaction with ScoE. Synthesis of **1** was performed in 50 µL of 50 mM HEPES (pH 8.0) containing 2 mM $MgCl_2$, 4 mM CABA-CoA, 100 µM apo-ScoB, 25-65 µM Sfp. The reaction was mixed gently and incubated for 15 min at room temp. When coupled with ScoE reaction, the following were included in the reaction mixture 0.5 mM alpha-ketoglutarate, 4 mM ascorbate, and 100 µM holo-ScoE. The coupled reaction was incubated for 15 min-overnight (sampling at short time intervals). After incubation, 450 uL of water was added and then the mixture was filtered (0.2 µm). LC-HRMS analysis was performed using an Agilent Technologies 6510 Accurate-Mass Q-TOF LC-MS instrument and a Phenomenex Aeris widepore XBC18 column (250 x 2.1 mm) with a linear gradient of 30 to 98% acetonitrile (v/v) over 30 min in $H_2O$ with 0.1% (v/v) formic acid, at a flow rate of 0.15 mL min$^{-1}$. The data were analyzed using Agilent MassHunter Qualitative Analysis software using the maximum entropy deconvolution feature.

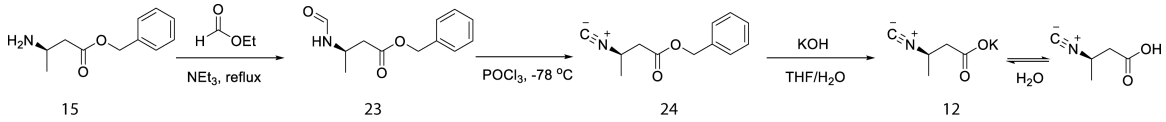



**Figure 3-14.** Chemical synthesis of **12** (highly toxic)**.**

### 3.4.12 Synthesis of 12 standard

A round flask with a magnetic stirrer was charged with 2 mmol **15**, 20 mL ethyl formate and 3 mmol $NEt_3$. The reaction mixture was refluxed overnight with a condenser. After cooling to room temp, the solvent was removed by rotatory evaporation and 20 mL DCM was added to the residue. The organic layer was washed with $H_2O$ and brine twice, respectively. Then the organic layer was dried with anhydrous $Na_2SO_4$ and concentrated by rotatory evaporation to get the crude product. The pure product was separated by silica gel with DCM:MeOH (100:1, Rf = 0.2) to yield benzyl (*R*)-3-formamidobutanoate (**23**), a colorless liquid (65% yield). **23:** Calculated for $C_{12}H_{15}NO_3$: 221.1, found: 221.0 (Varian GC-MS).

A round flask with a magnetic stirrer was charged with 1.0 mmol **23**, 2.7 mmol $NEt_3$ and 3 mL DCM. The mixture was stirred at -78 $^oC$ under a low flow of $N_2$. A solution of phosphoryl chloride (1.2 mmol) and DCM (0.5 mL) was added dropwise over 10 min. The reaction mixture was stirred at -78 $^oC$ for 1.5 h and monitored by TLC. The reaction was quenched by dropwise addition of saturated $Na_2CO_3$ solution (15 mL) and stirred for 15 min. The mixture was extracted with DCM three times and the organic layer was dried, concentrated and separated by silica gel with pure DCM as the eluant (Rf = 0.5). The compound benzyl (*R*)-3-isocyanobutanoate (**24**) was obtained as a colorless liquid (74% yield). This product can be kept at -80 $^oC$ for storage. Since product **12** is unstable after generation and highly toxic, product **24** was kept as the precursor of final product, and **24** could be hydrolyzed to make a freshly prepared solution of **12** when needed (for

experiments or HILIC-HRMS analysis). **24:** calculated for $C_{12}H_{13}NO_2$: 203.1, found: 203.1 (Varian GC-MS). $^1$H (400 MHz; $CDCl_3$) δ 7.39-7.34 (m, 5H), 5.17 (s, 2H), 4.16-4.08(m, 1H), 2.34-2.78 (m, 1H), 2.64-2.57 (m, 1H), 1.45-1.42 (m, 3H); $^{13}$C NMR (100 MHz; $CDCl_3$) δ 169.2, 156.2, 135.3, 128.8, 128.7, 128.5, 67.1, 46.5, 41.7, 21.5 (**Appendix M**).

Compound **24** was immediately subjected to the hydrolysis reaction by using KOH. A round flask equipped with a magnetic stirrer was charged with 0.5 mmol **24**, 0.5 mmol KOH, 3 mL THF and 0.6 mL $H_2O$. The mixture was stirred at room temp for 4-5 hrs. The organic solvent was removed *in vacuo* and the residue was redissolved in 10 mL DCM and 10 mL water. The water layer was separated and washed with DCM three times. The water layer was then frozen and lyophilized to yield the salt of product **12** (~55% yield)**.** Product **5** can be easily hydrolyzed into its acid form when dissolving in water, which could be detected by HILIC-HRMS. However, the lower stability of product **12** prevented its pure NMR characterization. **5:** calculated for $C_5H_8NO_2$ $[M+H]^+$ 114.0550; found 114.0548 (**Appendix K**).

### 3.4.13 ScoE *in vitro* reaction and HILIC-HRMS analysis

Assays were performed in 50 µL of 50 mM HEPES (pH 8.0) containing, 4 mM ascorbate, 0.5 mM alpha-ketoglutarate, 2 mM **11** or 6-$^{13}$C-CABA, and 100 µM holo-ScoE. The reaction was mixed gently and incubated for 15 min at room temp. After incubation, the reaction was quenched by adding 100 µL cold methanol, vortexed briefly, and centrifuged to remove protein residues. HILIC-HRMS analysis was performed using an Agilent Technologies 6510 Accurate-Mass Q-TOF LC-MS instrument and an Atlantis HILIC Silica 5 µm column (150 x 4.6 mm) with an isocratic mobile phase (80% acetonitrile (v/v) with 40 mM Ammonium formate, pH 4.5) for 30 min at a flow rate of 1 mL $min^{-1}$.

### 3.4.14 Determination of kinetic parameters of ScoE toward 11

Assays were performed in triplicate in 50 µL of 50 mM HEPES (pH 8.0) containing, 4 mM ascorbate, 10 mM alpha-ketoglutarate, 50 µM holo-ScoE. The reaction was initiated by adding **11** in the following concentrations: 63 µM, 127 µM, 400 µM, 638 µM, 1.27 mM, and 2 mM. After incubation, the reaction was quenched by adding 50 µL cold methanol, vortexed briefly, and centrifuged to remove protein residues. Time points were taken at 30 sec, 1 min, 2 min, 5 min, and 15 min to determine initial velocity of product formation. HILIC-HRMS analysis was performed using an Agilent Technologies 6510 Accurate-Mass Q-TOF LC-MS instrument and an Atlantis HILIC Silica 5 µm column (150 x 4.6 mm) with an isocratic mobile phase (80% acetonitrile (v/v) with 40 mM Ammonium formate, pH 4.5) for 20 min at a flow rate of 1 mL $min^{-1}$. Product concentration was calculated by comparing to synthetic **12**. Kinetic parameters were determined and plotted using Prism.

### 3.4.15 Determination of kinetic parameters of ScoE toward alpha ketoglutarate

Assays were performed in triplicate in a Corning half area flat bottom transparent 96 well plate in 50 µL containing, 4 mM ascorbate, 1 mM **11**, 2 µM holo-ScoE and the components of the Megazyme succinate assay kit. The reaction was initiated by adding alpha ketoglurate in the following concentrations: 10 µM, 20 µM, 50 µM, 100 µM, 200

μM, 500 μM. Formation of succinate was determined by monitoring NADH oxidation at 340 nm at 15 second intervals for 20 min using a Tecan M1000 plate reader. Kinetic parameters were determined and plotted using Prism.

### 3.4.16 ScoBCE coupled *in vitro* reaction and LC-HRMS analysis

Assays were performed in 30 μL of 50 mM HEPES (pH 8.0) containing, 4 mM ascorbate, 0.5 mM alpha-ketoglutarate, 2 mM **11**, and 100 μM holo-ScoE. The reaction was mixed gently and incubated for 15 min at room temp. After incubation, the reaction was mixed with 20 μL: 2 mM $MgCl_2$, 5 mM ATP, 50 μM holo-ScoB, and 50 μM ScoC. The reaction was gently mixed and incubated for 15 min at room temp. After incubation, 450 uL of water was added and then the mixture was filtered (0.2 μm). LC-HRMS analysis was performed using an Agilent Technologies 6510 Accurate-Mass Q-TOF LC-MS instrument and a Phenomenex Aeris widepore XBC18 column (250 x 2.1 mm) with a linear gradient of 30 to 98% acetonitrile (v/v) over 30 min in $H_2O$ with 0.1% (v/v) formic acid, at a flow rate of 0.15 mL $min^{-1}$. The data were analyzed using Agilent MassHunter Qualitative Analysis software using the maximum entropy deconvolution feature.

### 3.4.17 ScoABCE coupled in vitro reaction and LC-HRMS analysis

Assays were performed in 25 μL of 50 mM HEPES (pH 8.0) containing, 4 mM ascorbate, 0.5 mM alpha-ketoglutarate, 2 mM **11**, and 100 μM holo-ScoE. The reaction was mixed gently and incubated for 15 min at room temp. After incubation, the reaction was mixed with 25 uL: 2 mM $MgCl_2$, 5 mM ATP, 0.5 mM lysine, 4 mM NADPH, 50 μM ScoA, 50 μM holo-ScoB, and 20 μM ScoC. The reaction was gently mixed and incubated for 2 h at room temp. After incubation, the reaction was quenched by adding 100 μL cold methanol, vortexed briefly, and centrifuged to remove protein residues. LC-HRMS analysis was performed using an Agilent Technologies 6510 Accurate-Mass Q-TOF LC-MS instrument and an Eclipse Plus C18 column (100 x 4.6 mm). A linear gradient of 10-50% acetonitrile (vol/vol) over 12 min in $H_2O$ with 0.1% formic acid (vol/vol) at a flow rate of 0.5 $mL^{-1}$ was used. The product, **10,** was characterized by NMR previously.[110]

### 3.4.18 Marfey's analysis of 12

After ScoE + **11** *in vitro* reaction described above, protein was precipitated by adding MeOH to a final concentration of 50% and removed by centrifugation. Solvent was dried under $N_2$ and remaining residue was redissolved in 1 N HCl in MeOH to hydrolyze isonitrile moiety to a primary amine, forming β-aminobutyric acid (BABA).[130] After stirring at room temp for 3 h, the reaction was quenched by adding 1 M NaOH and the pH was adjusted to 7.0. After removing the solvent under $N_2$, the remaining residue was dissolved in 200 μl water and reacted with Marfey's reagent as described previously.[131] In brief, 50 μl sample was mixed with 200 μl Marfey's reagent (1% solution in acetone), and then 50 μl $NaHCO_3$ (1 M in water) was added. The reaction was incubated at 40 °C for 1 h and quenched by adding 25 μl HCl (2 N stock). The reaction mixture was diluted with 0.5 ml MeOH and analyzed by LC-MS. EIC traces (extracted at *m/z*= 356.1, calculated $[M+H]^+$=356.1) were used to analyze the data. Derivatized standards were prepared from the authentic *R*- and *S*-BABA and analyzed following an identical procedure.

### 3.4.19 Detection of the isonitrile moiety of 12 using tetrazine click reaction

Synthesized **12** (0.02 mmol, 2.26 mg, 1.0 eq) was dissolved in 100 μL THF and then 3,6-di(pyridin-2-yl)-1,2,4,5-tetrazine (0.04 mmol, 9.44 mg, 2.0 eq) was added.[132] The reaction was stirred at room temp for 1 h and quenched by adding 100 μL $H_2O$. ScoE + **11** *in vitro* reaction was carried out as described above. The protein was removed using Amicon Ultra centrifugal filters (cutoff MW 3000 Dalton), and the aqueous phase was dried under $N_2$. The remaining material was redissolved in 100 μL THF and any insoluble material was removed by centrifugation. Then 3,6-di(pyridin-2-yl)-1,2,4,5-tetrazine (0.04 mmol, 9.44 mg) was added. The reaction was stirred at room temp for 1 h and quenched by adding 100 μL $H_2O$. Click chemistry product was analyzed on LC-MS. 3,5-di(pyridin-2-yl)-1H-pyrazol-4-amine $[M+H]^+$ observed: 238.1068, calculated 238.1087.

# Chapter 4. Towards the Discovery of Isonitrile in Mycobacteria

Parts of this chapter have been adapted from the following with permission:
Harris, N. C., Sato, M., Herman, N. A., Twigg, F., Cai, W., Liu, J., Zhu, X., Downey, J., Khalaf, R., Martin, J., Koshino, H., Zhang, W.
Biosynthesis of isonitrile lipopeptides by conserved nonribosomal peptide synthetase gene clusters in Actinobacteria. *Proc Natl Acad Sci* **114**(27) 7025-7030 (2017).

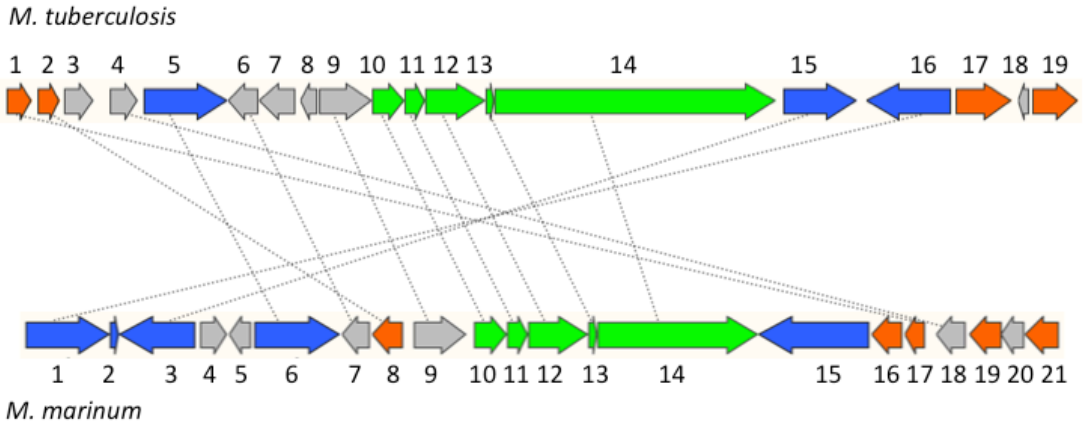
## 4.1 Introduction

Chapter 2 and 3 provided evidence that ScoE and MmaE mediate the formation of isonitrile when heterologously expressed in *E. coli*. However, at the conclusion of the aforementioned research, we had not observed the production of INLPs by any native producers, raising the possibility that the products observed in *E. coli* are not the same as those produced by native producers. Fortunately, direct evidence for isonitrile formation by ScoE and MmaA homologs was recently established. A gene cluster in *Staphylococcus aureus* containing *mmaB-E* homologs was shown to produce an isocyano-epoxyketone with antimicrobial activity;[133] and additional research confirmed that ScoA-E homologs in *Streptomyces sp.* are responsible for the biosynthesis of INLP **1** and **2** as well as the previously characterized INLP, SF2768.[44] SF2768 is structurally similar to INLP **2** with the Lys having been cyclized. Using these purified INLPs Wang *et al.* provided evidence for Cu(I) chelation by all INLPs tested with two INLP molecules coordinating a single Cu(I) atom. Interestingly, despite Cu(I) chelation by all three INLPs, only SF2768 was recognized for copper import. It appears that the cyclization of Lys is necessary to activate the chalkophore and to enable recognition by its cognate receptor, thus enabling metal transport from the extracellular environment into the cell.

It is not surprising that INLPs are implicated in metal transport. Indeed the isonitrile functionality is known to behave as an electron-rich analog of carbon monoxide and form coordination complexes with most transition metals.[41,134] Metals are required in many life processes, and bacteria have addressed the challenge of deficiencies in essential metals or high concentrations of toxic metal cations in their niche by evolving metal homeostasis mechanisms that in pathogenic bacteria are frequently associated with virulence.[27,135–141] The operon of *Rv0096-0101* has been known to be critical for the virulence of *M. tuberculosis* and essential for the survival of this pathogen in macrophages and mice.[83,85,86] Based on the metal binding ability of isonitrile and the recent evidence that INLPs from *Streptomyces sp.* coordinate Cu(I), we postulated that this conserved biosynthetic gene cluster plays a role in metal transport and homeostasis in Mycobacteria. *In silico* analysis demonstrated that putative heavy metal translocating P-type ATPases such as CtpA and CtpB are encoded in close proximity to the INLP biosynthetic genes in the *M. tuberculosis* and *M. marinum* genomes (**Fig 4-1**), and the expression of *ctpB* was shown to be co-regulated with *Rv0096-0101* in M. tuberculosis by transcriptomic analysis.[142] In addition, it has been reported that the expression of the *Rv0096- 0101* gene cluster was normally repressed in many synthetic media including

7H9 where metal concentrations are high, but was constantly induced in Sauton's medium where metal concentrations are relatively low,[82] suggesting a metal uptake function of the cluster.

These data encouraged us to pursue two related paths of research. The first goal, was to isolate the native compound produced by *Mycobacteria sp.* This proved to be more difficult than expected and research is currently ongoing; however this chapter will detail the efforts made to identify such a product. The second goal was to look for an association between production of the native MmaA-E INLP and metal acquisition in *M. marinum* and to assign biological function to this INLP. Unfortunately, we recently acquired additional data that suggests the *mmaA-E* operon is not expressed in planktonic growth conditions and is only active during biofilm growth. Being that the experiments detailed in this chapter were performed under planktonic growth conditions, it is highly likely that most of the data from the native producer are artifacts. Nonetheless, the tools developed during this exploration and the intention behind each experiment may still prove to be useful for future studies and so they will be discussed in this chapter.




| Gene Number | Locus | Putative Function |
|---|---|---|
| Mtb 1 | Rv0088 | Polyketide cyclase/dehydratase |
| Mtb 2 | Rv0089 | O-methyltransferase |
| Mtb 3 | Rv0090 | Membrane protein |
| Mtb 4 | Rv0091 | MTA/SAH nucleosidase |
| Mtb 5 | Rv0092 | CtpA Cation-transporting P-type ATPase |
| Mtb 6 | Rv0093 | Anti-sigma factor |
| Mtb 7 | Rv0094c | Membrane protein |
| Mtb 8 | Rv0095 | Conserved hypothetical protein |
| Mtb 9 | Rv0096 | PPE family protein |
| Mtb 10 | Rv0097 | Oxidase |
| Mtb 11 | Rv0098 | Thioesterase |
| Mtb 12 | Rv0099 | Acyl ACP ligase |
| Mtb 13 | Rv0100 | Acyl carrier protein |

| | | |
|---|---|---|
| Mtb 14 | Rv0101 | Non-ribosomal peptide synthetase |
| Mtb 15 | Rv0102 | CtaG cytochrome C oxidase assembly factor |
| Mtb 16 | Rv0103c | CtpB Cation-transporting P-type ATPase |
| Mtb 17 | Rv0104 | cAMP dependent protein kinase |
| Mtb 18 | Rv0105c | L28P Family of ribosomal proteins |
| Mtb 19 | Rv0106 | Nitrile hydratase |

| **Gene Number** | **Locus** | **Putative Function** |
|---|---|---|
| Mma1 | Mmar_0269 | CtpB Cation-transporting P-type ATPase |
| Mma2 | Mmar_0268 | Metal Chaperone |
| Mma3 | Mmar_0267 | CtaG cytochrome C oxidase assembly factor |
| Mma4 | Mmar_0266 | Secreted protein |
| Mma5 | Mmar_0265 | SigC regulatory element |
| Mma6 | Mmar_0264 | CtpA Cation-transporting P-type ATPase |
| Mma7 | Mmar_0263 | Putative Zinc Finger; anti-sigma factor |
| Mma8 | Mmar_0262 | O-methlytransferase |
| Mma9 | Mmar_0261 | PPE family protein |
| Mma10 | MmaE | Oxidase |
| Mma11 | MmaD | Thioesterase |
| Mma12 | MmaC | Acyl ACP ligase |
| Mma13 | MmaB | Acyl carrier protein |
| Mma14 | MmaA | Non-ribosomal peptide synthetase |
| Mma15 | Mmar_0255 | mmpl5 MFS |
| Mma16 | Mmar_0254 | Ketoacyl reductase |
| Mma17 | Mmar_0253 | SRPBCC Superfamily (putative polyketide cyclase/dehydratase) |
| Mma18 | Mmar_0251 | MTA/SAH nucleosidase |
| Mma19 | Mmar_0250 | Desaturase |
| Mma20 | Mmar_0249 | TetR transcriptional regulator |
| Mma21 | Mmar_0248 | Short-chain type dehydrogenase/reductase |

**Figure 4-1.** Bioinformatics analysis of *M. marinum* strain M and *M. tuberculosis* H37Rv biosynthetic gene clusters and surrounding genes. ABCDE are colored green, genes putatively related to metal transport are colored blue, putative fatty acid modification genes are colored orange, and genes with other functions are colored gray. The genes are numbered and putative functions are listed in the table. Homologous genes between the two clusters are cross-linked in the figure.

## 4.2 Results

### 4.2.1 Potential role of INLP in Metal Transport

During LC-HRMS analysis of *E. coli* cultures expressing *mmaA-E*, we consistently noticed the presence of two additional metabolites with molecular formulas $C_{28}H_{48}CuN_4O_4$ (calculated for $C_{28}H_{48}CuN_4O_4^+$: 567.2966; 567.2978) and $C_{56}H_{96}CuN_8O_8$ (calculated for $C_{56}H_{96}CuN_8O_8^+$: 1071.6642; found 1071.6632) (**Fig 4-2**). The MS/MS spectra of these two new metabolites indicated that they are copper chelated forms of INLP **6** consisting of a single copper atom coordinated with either one or two INLP **6** molecules (**Fig 4-3**). High collision energy (CE 200) MS/MS produced a product ion (62.9) consistent with the presence of copper (**Fig 4-4**).

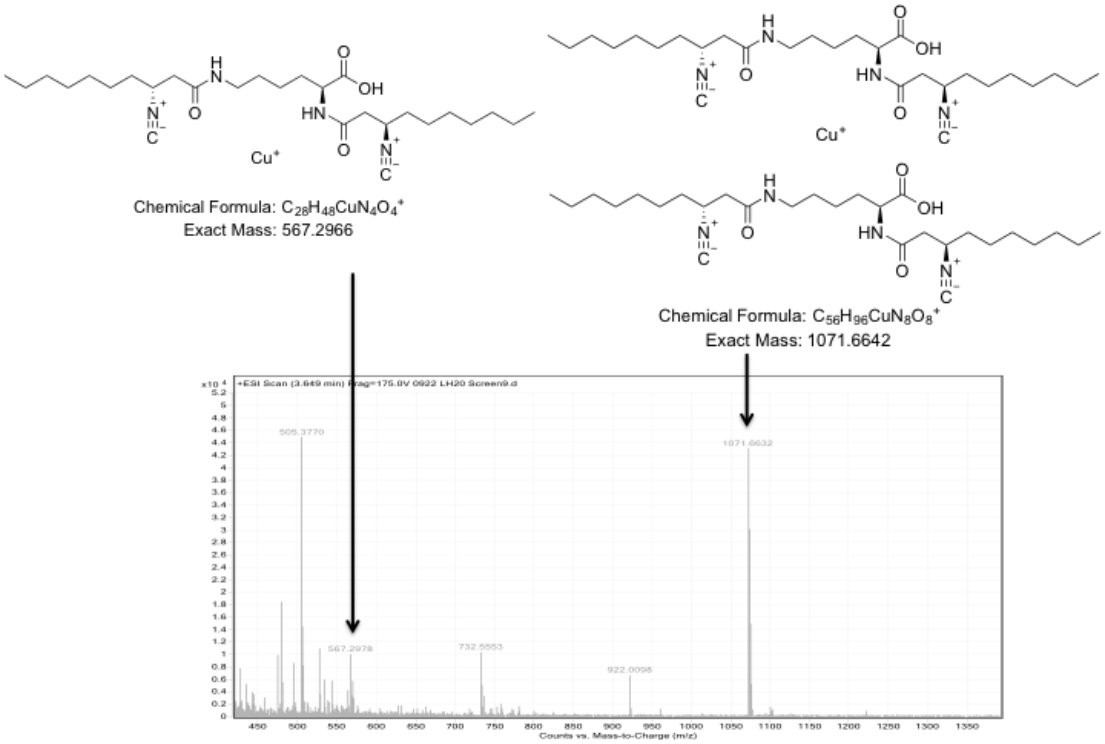



**Figure 4-2.** HRMS spectra of INLP **6**-Cu(I) complexes.

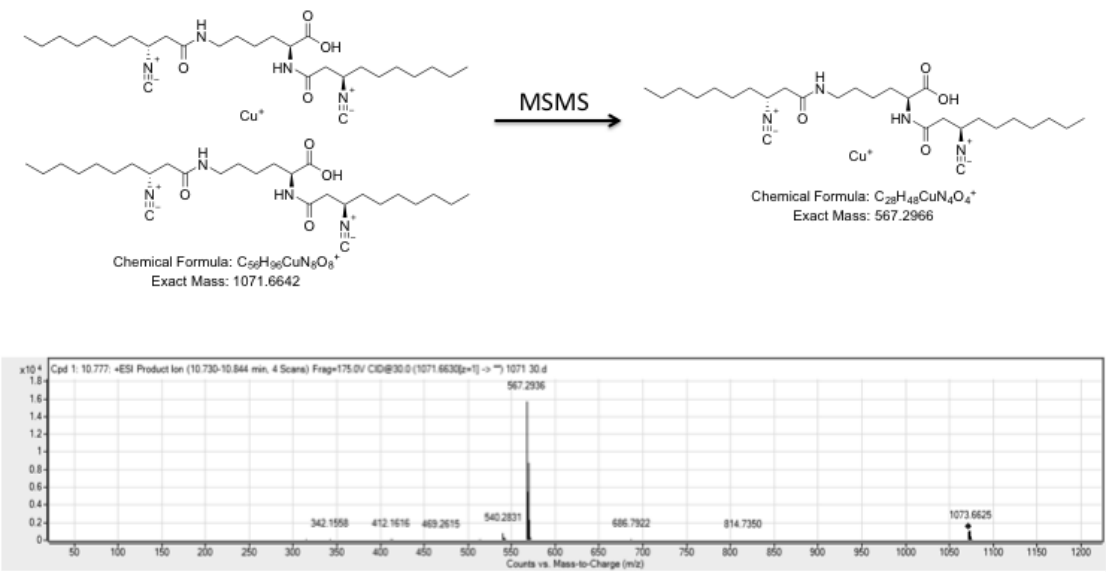



**Figure 4-3.** HRMS/MS spectra of the INLP **6**-Cu(I) complex. Two INLP **6** molecules can coordinate a single Cu(I) and MS/MS of this complex yields a major product ion representing a single INLP 6 molecule coordinated to a single Cu(I) atom.

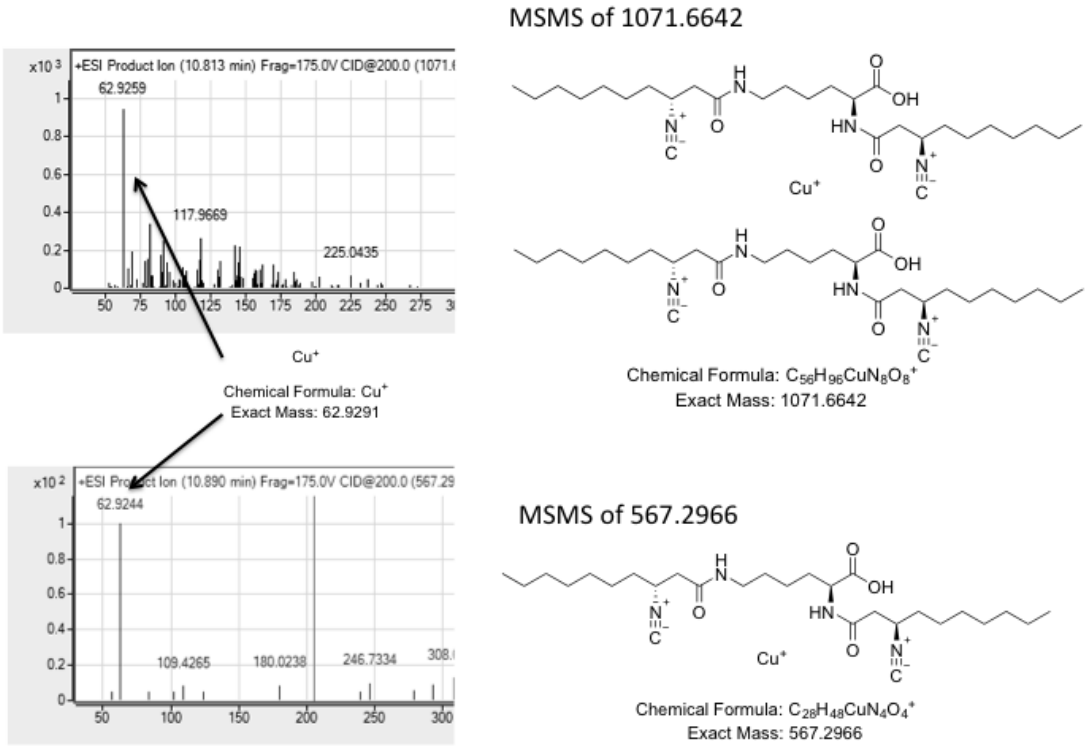



**Figure 4-4.** High collision energy MS/MS of INLP **6** metal complexes results in the release of Cu(I). A collision energy of 200V was used.

This finding as well as the general theme of isonitrile involvement in metal transport in other systems, the recent discovery that related INLPs from *Streptomyces sp.* are

chalkophores, and the presence of metal transport genes in the vicinity of the native *mmaA-E* operon, supported the hypothesis that the mycobacterial MmaA-E product is a metallophore with a role in metal transport. To further probe this hypothesis, we constructed a mutant strain of *M. marinum* devoid of *mmaA-E* through allelic exchange according to the reported method[143] and compared the intracellular metal content of the *M. marinum* wild-type and *ΔmmaA-E* cells. Deletion of the gene cluster caused a significant decrease in the intracellular accumulation of zinc in Sauton's medium as shown by inductively coupled plasma-optical emission spectrometry (ICP-OES) analysis (**Fig 4-5**). This is consistent with the finding from microarray analysis that the transcription of *Rv0097* was down-regulated in *M. tuberculosis* exposed to excess zinc.[138] However, increased accumulation of zinc was not consistent with our finding that Cu(I) is coordinated by INLP **6**. Although, because INLP **6** is produced via heterologous expression in *E. coli*, we reasoned that the native INLP produced by *M. marinum* may coordinate with additional or different transition metal ions.

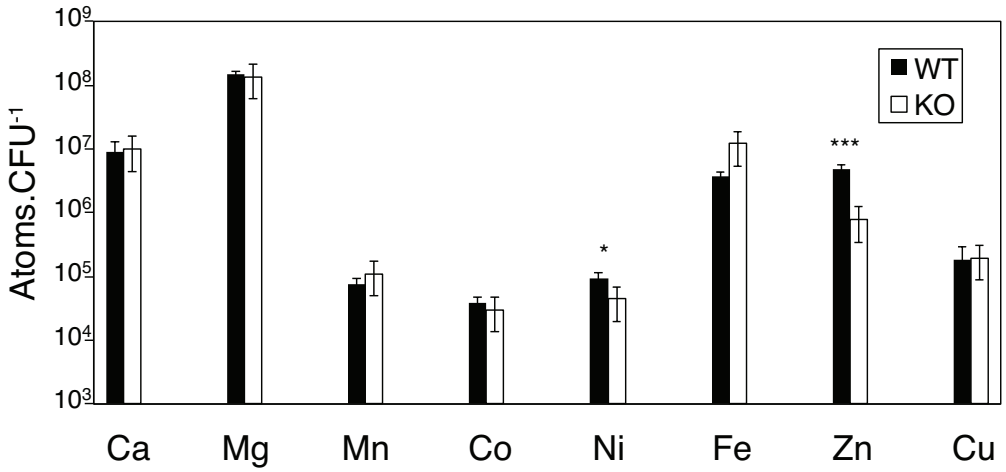



**Figure 4-5.** Effect of *mmaA-E* mutation in metal content. Intracellular metal content of the *M. marinum* wild-type (Left) and *ΔmmaA-E* (Right) strains grown in Sauton's medium was determined by ICP-OES. Error bars, mean ± SD. *P ≤ 0.05 and ***P ≤ 0.001.

### 4.2.2 Metabolomics analysis of Mycobacteria to identify native INLP

Heterologous expression of *mmaA-E* in *E. coli* allows for the production of INLPs **6** and **7**. Although these products are consistently produced in *E. coli*, we could not rule out that the native MmaA-E product, produced by *Mycobacteria sp.*, differs in structure, perhaps being modified by additional tailoring enzymes. Presumably the native product still maintains the basic structural architecture of INLPs **6** and **7**, having a central Lys molecule with two C10-C12 acyl chains bound via amide linkages from the activity of MmaABC and isonitrile moieties from the activity of MmaDE. However, after numerous attempts to grow both *M. marinum* and *M. tuberculosis* in various media with substrate feeding and subsequent solvent extraction and LC-HRMS analysis, none of the INLPs produced by *E. coli* cultures expressing *mmaA-E* or *Rv0097-0101* were observed. This

negative result indicated that either the gene cluster was silent under the culturing conditions and the compound(s) were not produced, or that the products were modified to an extent that we could not predict. Due to the technical difficulty of working with BSL2-3 organisms, RNA extraction and expression analysis routinely failed (positive and negative controls failed), so we turned to comparative metabolomic analysis using targeted gene disruption mutants.

Wild-type H37Rv *M. tuberculosis* and strains harnessing transposons in *Rv0099* and *Rv0101* were grown in triplicate in Sauton's and 7H9 media. Media was either prepared without additions or supplemented with different combinations of Gly, Lys, and with either 2,3-decenoic acid, or 2,3-dodecenoic acid. Cultures were extracted with organic solvents, concentrated, and analyzed by LC-HRMS followed by untargeted metabolomics analysis using XCMS[94,95] for the determination of metabolic profile differences and the identification of metabolites. The resulting metabolite pools were very similar and any positive hits were discredited with future analysis, being that with successive experimentation no compounds were produced solely by wild-type cells and absent in mutant pools.

We then switched systems to see if *M. marinum* would yield more promise. Plasmids were constructed with strong mycobacterial constitutive promoters driving expression of either *mmaA-E*, *sigC* (a positive regulator of *Rv0096-0101* in *M. tuberculosis*),[142] or *Rv0097-0101*, as well as plasmids containing strong tetracycline inducible promoters for tunable expression of the same. Wild-type *M. marinum* strains harnessing overexpression plasmids and *ΔmmaA-E* strains were grown in various types of media with or without feeding additional substrates. Cultures were extracted with organic solvents, concentrated, and analyzed by LC-HRMS followed by untargeted metabolomics analysis. One new metabolite **25** with a proposed molecular formula $C_{62}H_{99}CuN_4O_8$ (calculated for $C_{62}H_{99}CuN_4O_8^+$: 1090.6759; found 1090.6750) was consistently identified in every culture with the exception of the *ΔmmaA-E* strain (**Fig 4-6**). However, the titer was not increased when overexpressing pathway genes or *sigC* as compared to wild-type *M. marinum*.

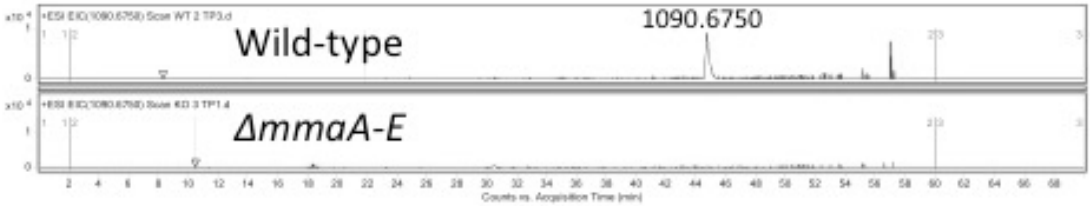
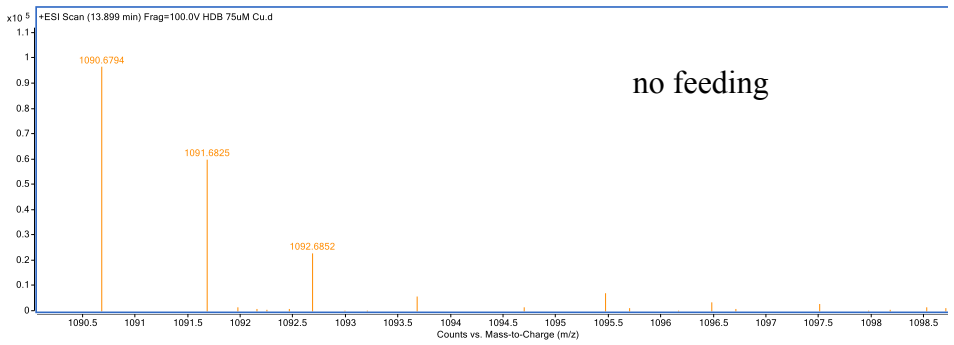
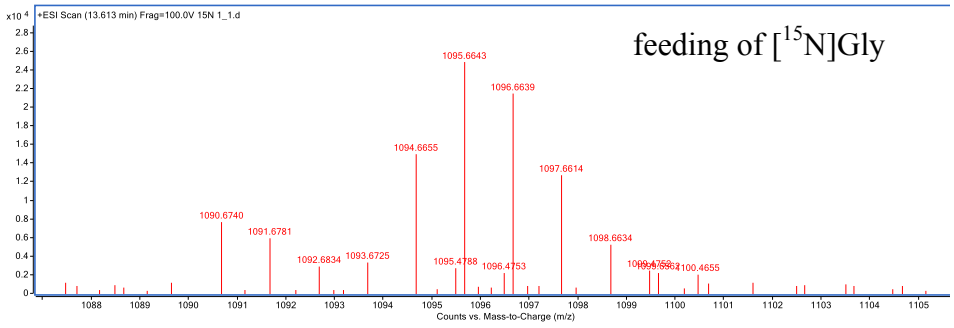


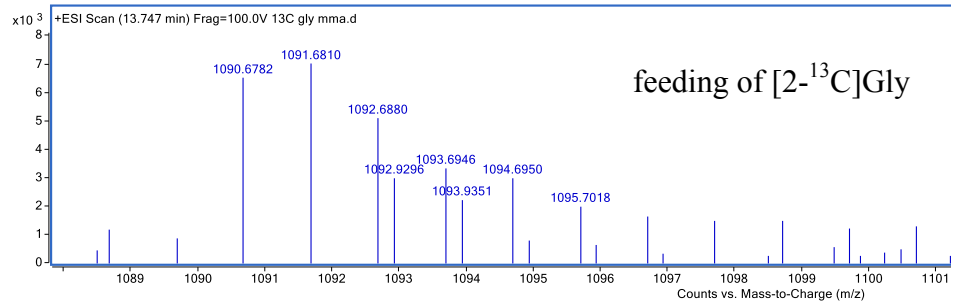

**Figure 4-6.** Extractive ion chromatograms of compound **25** comparing wild-type *M. marinum* and *ΔmmaA-E* strain production. **25** is produced by wild-type *M. marinum* but absent in the *ΔmmaA-E* strain. A 10 ppm mass-error tolerance was used.

To further verify that this metabolite was a product of *mmaA-E*, we separately fed labeled [4,4,5,5-d4]Lys, [2- $^{13}$C]Gly, and [$^{15}$N]Gly in addition to supplementation with 2,3-decenoic acid to wild-type *M. marinum* cultures. Feeding of [4,4,5,5-d4]Lys, [2-$^{13}$C]Gly, and [$^{15}$N]Gly resulted production of **25** having mass shifts of +10, +2, and +5,

respectively (**Fig 4-7**). The mass spectral shifts indicate that these two amino acid substrates are incorporated into **25**, however these data did not clearly define how this new metabolite is related to previously identified INLPs. We thus proceeded with further chemical characterization first attempting to purify **25** with successive rounds of HPLC, however due to the nature of growing *M. marinum* and the safety concerns associated with growing large culture volumes, we were unable to purify a sufficient quantity for NMR analysis. Therefore we used MS/MS analysis to match the fragmentation profile with that of other INLPs, however few similarities were identified. Both unlabeled and labeled substrate feeding were used before subsequent extraction and MS/MS analysis.



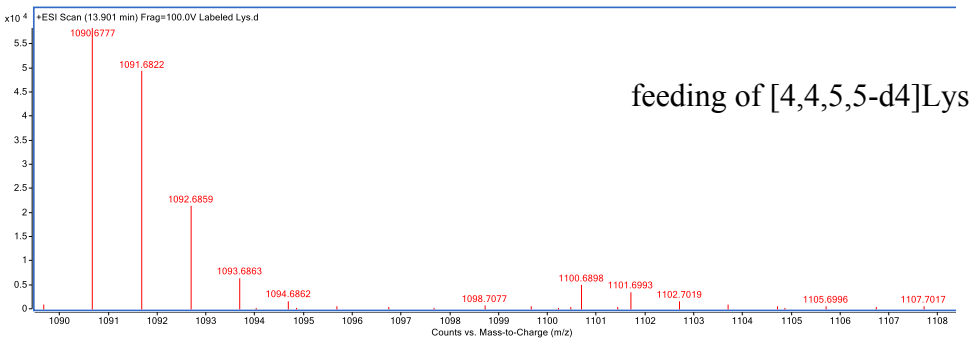



**Figure 4-7.** Mass spectra of **25** resulting from feeding of labeled substrates to a *M. marinum* cultures. Labeled Lys and Gly are retained however it is not clear how these substrates have been incorporated into the final **25** product.

Since we presumed **25** is an INLP, we hypothesized that it should chelate transition metals. We therefore used high collision energy MS/MS to release any bound metal ions; however no metal ions were detected as product ions. Despite this result, we probed whether we could increase production of **25** by modulating the metal concentration in the media. We reasoned if **25** is an INLP and involved with metal homeostasis, limiting or saturating conditions of various metals may lead to cellular reprogramming and increased flux through the MmaA-E pathway. Media was treated with Chelex to remove all transition metals and then specific metals were supplemented at varying concentrations. Different combinations were used to keep either one or multiple metals limiting or added in excess, with concentrations varying between 0 and 100 μM. No differences in product titer were observed (**Fig 4-8**), suggesting that **25** does not play a role in metal homeostasis or at minimum, that its biosynthesis is not in response to metal availability.

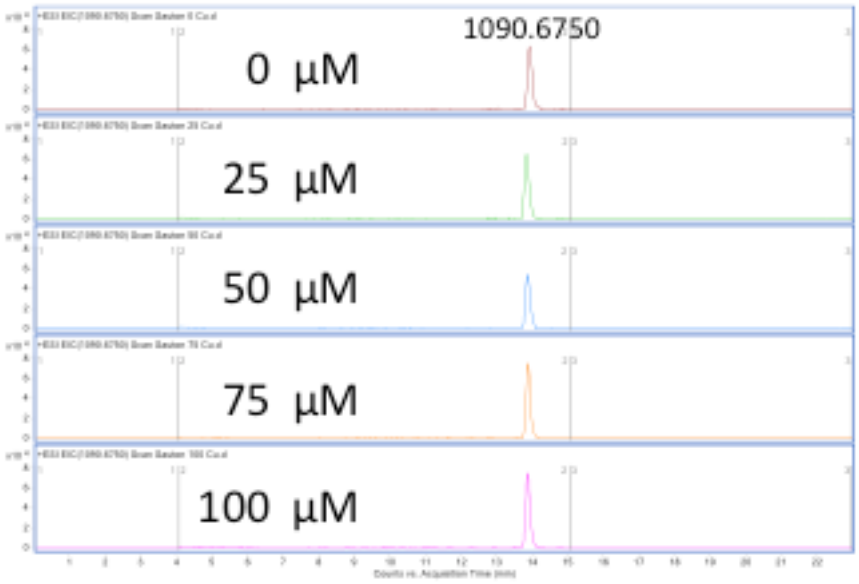



**Figure 4-8.** Extracted ion chromatograms of **25** from *M. marinum* cultures supplemented with various concentrations of Cu(II). Titer of 25 is not influenced by Cu(II) concentration. A 10 ppm mass-error tolerance was used.

Directly downstream of the *mmaA-E* gene cluster is a major facilitator superfamily type protein (MFS) encoded by *mmar0255*. MFS are often associated with transport of lipids into the cell[144] or the secretion of small molecule secondary metabolites including siderophores and virulence factors out of the cell.[145] Due to the proximity of the MFS to the INLP biosynthetic genes we posited that the MFS might actively secrete the MmaA-E product. Based on the assumption that **25** is produced by MmaA-E, we constructed a strain deficient of the MFS to probe whether it is necessary for import of the **25** substrate pool or for secretion of the final product. Wild-type *M. marinum* and *Δmma0255* strains were grown and the cellular and extracellular fractions separated prior to solvent extraction and subsequent LC-HRMS analysis. The *Δmma0255* mutant showed decreased total production of **25** but the percentage of intra- and extra-cellular product was unchanged as compared to wild-type (**Fig 4-9**). The decrease in titer of **25** indicated that the MFS encoded by *mma0255*, is functionally important, either directly or indirectly, for the import of the **25** substrate pool.

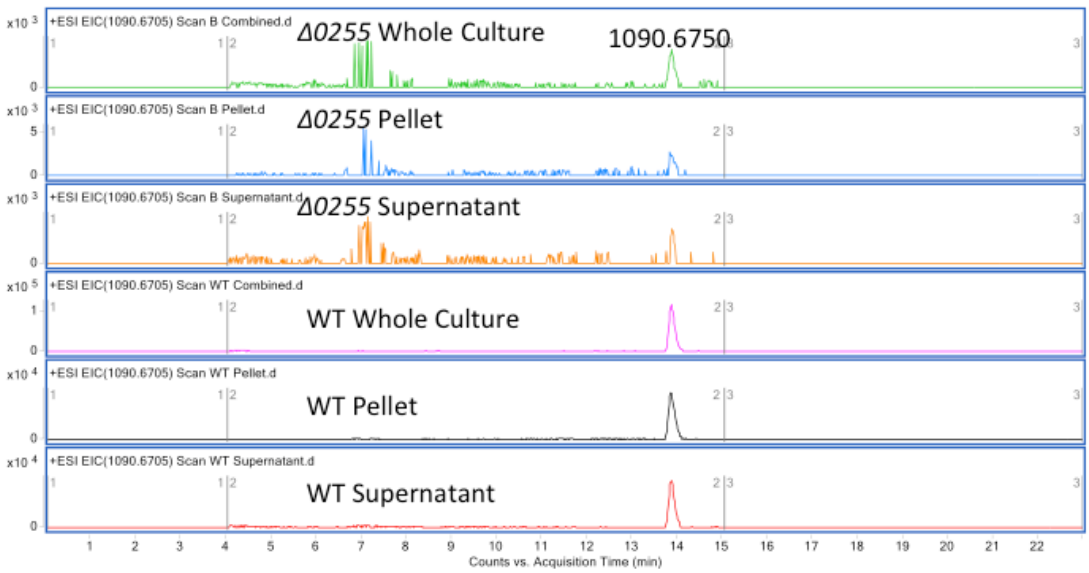



**Figure 4-9.** Extracted ion chromatograms of **25** from wild-type *M. marinum* and a *Δmma0255* strain. The cultures were fractionated to separate the cellular and extracellular fractions prior to compound extraction and subsequent LC-HRMS analysis. Disruption of the MFS encoded by mma0255 results in a decreased titer of **25**. A 10 ppm mass-error tolerance was used.

## 4.3 Discussion

Preliminary data suggests that *M. marinum* produces a secondary metabolite, **25,** that is not produced in strains with *mmaA-E* disrupted. Production of this metabolite is also decreased when the MFS encoded by *mma0250* is disrupted, indicating that the MFS is involved in substrate import. Due to the difficulty of working with a human pathogen, we were unable to purify a sufficient quantity of **25** for NMR analysis. Feeding of labeled substrates to the culture led to the incorporation of Lys and Gly into the final **25** product, however further HRMS and MS/MS analysis was inconclusive, and it is currently not known how these amino acid substrates are utilized in the biosynthesis of **25**. Additionally, based on high collision energy MS/MS analysis, **25** does not appear to coordinate metal ions, which differs from the tendency of INLPs to chelate Cu(I). Along these lines, the titer of **25** is not influenced by metal availability as determined by modulating metal concentrations in prepared media. Perhaps contradicting these data, a *ΔmmaA-E* strain accumulates less zinc than wild-type *M. marinum*.

Although **25** is not produced by the *ΔmmaA-E* strain, it is not possible at this time to confirm that **25** is the product of MmaA-E enzymes; more structural information is needed to predict how **25** is biosynthesized and additional experiments are required to assign a biological function to **25**. For instance using chemical complementation during infection of macrophages with a *ΔmmaA-E* strain may prove useful for determining if **25** can recapitulate a wild-type virulence phenotype. Additionally, the observed metabolic phenotypes observed for *ΔmmaA-E* and strain *Δmma0250* should be genetically complemented to rule out the possibility of polar effects.

After completing the experiments detailed above, we discovered that *mmaA-E* is only expressed during biofilm growth and not expressed during planktonic growth. As the growth conditions used during this study were only conducted during planktonic growth, it is likely that the assignment of **25** to the *mmaA-E* gene cluster is not warranted. That being said, **25** is observed in a *M. marinum* strain harnessing a plasmid with constitutive expression of *mmaA-E*; genetic complementation of the *ΔmmaA-E* strain with this overexpression plasmid and subsequent LC-HRMS analysis would help clarify this quandary. Additionally, our ICP-OES data is likely an artifact. Currently, we cannot definitively say that wild-type *M. marinum* cells accumulate more zinc than a strain lacking *mmaA-E* until this phenotype is recapitulated with genetic or chemical complementation. Future experiments should also be conducted during biofilm forming conditions to accurately assign a metabolite and biological role to the *mmaA-E* gene cluster.

## 4.4 Materials and Methods

### 4.4.1 Construction of Plasmids for Overexpression in *M. marinum*

Individual genes were PCR amplified from genomic DNA and cloned into pMK1 (tetracycline inducible promoter, $Hyg^r$), or pMV261(strong constitutive promoter, $Zeo^r$) by Gibson Assembly (New England Biolabs). All primers used are listed in **Table S3**. Plasmids were isolated using a QIAprep Spin Miniprep Kit (Qiagen) and confirmed by DNA Sequencing (UC Berkeley DNA Sequencing Facility).

**Table 4-1.** Plasmids used in this study.

| |
|---|
| pNCH2 *Rv0097-0101* on pMK1. |
| pNCH13 *Rv0097-0101* on pMV261. |
| pNCH8 *mmaA-E* on pMK1. |
| pNCH12 *mmaA-E* on pMV261. |
| pNCH11 *mma_sigC* on pMV261. |
| pNCH20 *mma_sigC* on pMK1. |

**Table 4-2.** Primers used in this study.

| | |
|---|---|
| NCH2 mtbA-E overlap fwd | CTAAAGAGGAGAAAGGATCTATGACGCTTAAGGTCAAAGG |
| NCH2 mtbA-E piece 1 rev | TGCCGCCGCTGCAAGCGAGGC |
| NCH2 mtbA-E piece 2 fwd | GCCTCGCTTGCAGCGGCGGCA |
| NCH2 mtbA-E piece 2 rev | GCCACCAGCACATGTTCGTCGTCG |
| NCH2 mtbA-E piece 3 fwd | CGACGACGAACATGTGCTGGTGGC |
| NCH2 mtbA-E overlap rev | cctggagatccttactcgaGCTACAGCAGTCCGAGCAGTTG |
| NCH13 mtbA-E overlap fwd | cggatccagctgcagaattcATGACGCTTAAGGTCAAAGG |
| NCH13 mtbA-E piece 1 rev | TGCCGCCGCTGCAAGCGAGGC |
| NCH13 mtbA-E piece 2 fwd | GCCTCGCTTGCAGCGGCGGCA |

| | |
|---|---|
| NCH13 mtbA-E piece 2 rev | GCCACCAGCACATGTTCGTCGTCG |
| NCH13 mtbA-E piece 3 fwd | CGACGACGAACATGTGCTGGTGGC |
| NCH13 mtbA-E overlap rev | gtcgacatcgataagcttcCTACAGCAGTCCGAGCAGTTG |
| NCH8 mmaA-E overlap fwd | CTAAAGAGGAGAAAGGATCTatgacgctcaacgtgaaagg |
| NCH8 mmaA-E piece 1 rev | gatacagcgcggggtcgttg |
| NCH8 mmaA-E piece 2 fwd | caacgaccccgcgctgtatc |
| NCH8 mmaA-E overlap rev | caagacccgcactggcctagCtcgagtaaggatctccagg |
| NCH12 mmaA-E overlap fwd | cggatccagctgcagaattcATGACGCTTAAGGTCAAAGG |
| NCH12 mmaA-E piece 1 rev | TGCCGCCGCTGCAAGCGAGGC |
| NCH12 mmaA-E piece 2 fwd | caacgaccccgcgctgtatc |
| NCH12 mmaA-E overlap rev | cgtcgacatcgataagcttcctaggccagtgcgggtcttg |
| NCH11 sigC overlap fwd | cggatccagctgcagaattcatgaccacgtcaccggacga |
| NCH11 sigC overlap rev | cgtcgacatcgataagcttctcagccggttagatcgtcgc |
| NCH20 sigC overlap fwd | ctaaagaggagaaaggatctatgaccacgtcaccggacga |
| NCH20 sigC overlap rev | cctggagatccttactcgagtcagccggttagatcgtcgc |

### 4.4.2 Gene Disruption in *M. marinum* and Mutant Analysis

Deletion of the *Mmar_0256-0260 (mmaA-E)* and *Mmar_0255* loci in *Mycobacterium marinum* strain M was performed by cloning 800-bp of the start of *Mmar_0260* and end of *Mmar_0256,* and the start and end of *Mmar_0255*, respectively, on either side of the hygromycin resistance gene on pMSG360 and using homologous recombination via transduction with the temperature-sensitive mycobacteriophage phAE87, as described previously. The phage was incubated with *M. marinum* cells at an MOI of 10 for 4 h at 39ºC before plating on hygromycin 7H10 plates and growing at 30ºC. The gene disruption, via hygromycin insertion, was verified by Sanger sequencing.

### 4.4.3 Identification of *E. coli-mmaA-E* metabolite-Cu(I) complexes and high collision energy MS/MS

*E. coli* BAP1 strain harnessing plasmids containing *mmaA-E* was grown as 30 mL cultures at 37ºC in LB containing appropriate antibiotics until $OD_{600}$ 0.5 and induced with 0.5 mM IPTG. For 2,3-decenoic acid and Gly was fed to a final concentration of 1 mM and 10 mM, respectively, at the time of induction. After induction, the temperature was decreased to 20ºC, and compound production was allowed to proceed for approximately 48 h. The entire culture (both pellet and supernatant) was extracted with an equal amount of chloroform. The organic extract was dried and redissolved in methanol and analyzed via liquid chromatography-high resolution mass spectroscopy (LC-HRMS). LC-HRMS analysis was normally performed using an Agilent Technologies 6520 Accurate-Mass Q-TOF LC-MS instrument and an Agilent Eclipse Plus C18 column (4.6 x 100 mm). A linear gradient of 50-98% $CH_3CN$ (vol/vol) over 12 min in $H_2O$ with 0.1% (vol/vol) formic acid at a flow rate of 0.5 mL/min was used. HRMS/MS analysis was conducted using targeted MS/MS with collision energy of 5-200V, with 200V used for high collision energy MS/MS experiments.

### 4.4.4 Identification of *M. marinum* metabolite through untargeted, comparative metabolomics

Wild-type *M. marinum* or strains harnessing overexpression plasmids and a control *ΔmmaA-E* strain were grown in triplicate. 50 mL cultures were grown at 30ºC with shaking at 105 rpm in either 7H9, Sula's, Sauton's, Dubo's, Proskauer-Beck, Mueller-Hinton, or GAS media containing appropriate antibiotics for 1 week. Cultures were supplemented with 2,3-decenoic acid or dodecenoic acid and Gly and Lys using a final concentration of 1 mM for all substrates. Samples were taken for compound production analysis every 24 hours. The entire culture (both pellet and supernatant) was extracted with an equal amount of chloroform or using a ratio of 2:1 chloroform:methanol. The organic extract was dried and redissolved in methanol and analyzed via LC-HRMS. LC-HRMS analysis was normally performed using an Agilent Technologies 6520 Accurate-Mass Q-TOF LC-MS instrument and an Agilent Eclipse Plus C18 column (4.6 x 100 mm). A linear gradient of 2-98% $CH_3CN$ (vol/vol) over 45 min in $H_2O$ with 0.1% formic acid (vol/vol) at a flow rate of 0.5 mL/min was used. XCMS software was used for untargeted metabolomic comparisons of extracts from wild-type and *ΔmmaA-E* strains. In a typical small-scale experiment for **25** production analysis, media was either untreated or treated with Chelex to remove metals. For media treated with Chelex, metals (Zn, Fe, K, Mg, Ca, Cu, Co, Mn, Ni, or no metals added) were supplemented at a concentration of 0-100 μM and in various combinations to keep one or multiple metals either limiting or in excess. The entire culture or fractions containing either the cell pellet or supernatant were extracted and prepared for LC-HRMS analysis. For these experiments, a linear gradient of 10 to 98% $CH_3CN$ (vol/vol) over 30 min followed by 10 min at 98% in $H_2O$ with 0.1% formic acid (vol/vol) at a flow rate of 0.5 mL/min was used. HRMS/MS analysis was conducted using targeted MS/MS with collision energy of 5-200V, with 200V used for high collision energy MS/MS experiments.

### 4.4.5 ICP-OES Analyses.

Wild-type and *ΔmmaA-E* were grown in 50 mL Sauton's medium at 30ºC and 105 rpm in triplicate. 15 mL of culture was harvested at late exponential phase and washed 3x with cold 1 mM EDTA. Cell pellets were dried at 100ºC overnight and resuspended in 70% trace metal grade nitric acid. The cell suspensions were boiled overnight in acid-washed Pyrex Erlenmeyer flasks. Dried samples were redissolved in 5 mL 2% nitric acid prior to analysis. Standard solutions were dissolved in the same matrix. The concentration of the trace elements was determined using a Perkin Elmer 5300 DV ICP-OES at College of Natural Resources, University of California at Berkeley. The analytical precision of the measurements is better than $\pm$2% RSD.

# Chapter 5. Conclusion

Tuberculosis remains a public health threat with millions of deaths occurring annually. Over evolutionary history, *M. tuberculosis* has developed many effective mechanisms to dodge and manipulate the host immune system. At the root of this strategy is the secretion of virulence factors that trigger and rewire phagocytosis to allow safe passage into a comfortable niche from where the bacilli thrive. The bacterium lives within the macrophage and stealthily scavenges key nutrients from the host. It's a game of patience and *M. tuberculosis* can avoid immune surveillance for up to decades waiting for the host to become immunocompromised. And when the time is right, the bacterium triggers massive events of autophagy, which lead to severe inflammation and the progression of the disease state. Necrosis of the lungs causes uncontrolled bouts of coughing and the contagious spread to additional hosts. It is this necrosis that inevitably leads to death.

It could be argued that no other pathogen so effectively manipulates the human immune system the way that *M. tuberculosis* is able to. This pathogenic achievement is owed to the extraordinarily complex array of virulence factors produced during the progression of infection. This dissertation described the discovery of a new family of INLPs that may serve as additional virulence factors in the already large quiver. The results described add another piece to the puzzle in this complex game of cat and mouse. In order to more effectively combat this awful pathogen, we need to continue dissecting the chemical logic that dictates pathogenesis. There has been a recent rise in the number of *M. tuberculosis* strains possessing resistance to most or even all of the antibiotics currently used to treat the infection. We need to continue to innovate and establish new public health initiatives to slow the spread of disease, and the quick development of new drugs is needed to win the evolutionary battle. Many of the produced virulence factors are necessary for survival in the host and therefore their biosynthetic machinery may be an effective target for new drugs. In order to develop drugs that effectively inhibit biosynthetic pathways, we need to fully understand the activities of each implicated enzyme in these pathways. Once we understand how the machinery works, we can then develop strategies to disrupt it. My hope is that the information included in this dissertation will add to the collective knowledge of the chemical logic used by *M. tuberculosis*, which in turn will help move us closer to ending the epidemic.

Additionally, a new mechanism for isonitrile biosynthesis was described. The isonitrile is a bioactive warhead that has been incorporated into numerous natural products. It is often stitched onto molecular scaffolds to provide a means for metal chelation, which relates to many life processes extending beyond virulence. Outside of their biological environment, many of these isonitrile containing natural products have medicinal value as antibiotics and anticancer drugs. Additionally, while the isonitrile is an important moiety for natural products, its biosynthesis could also be tailored to support the growing field of chemical biology. The electron rich CN bond has a unique Raman and infrared excitation enabling quick identification. Bioengineering the incorporation of the isonitrile into other natural products could prove a useful tool which could allow for *in situ* molecular imaging and quantification using either stimulated Raman spectroscopy or using a click reaction to link a bio-orthogonal fluorophore to it. The isonitrile shows

promise as a chemical tool and increased understanding of its biosynthesis will enable us to take advantage of its use as such.

# References


[1] D. Romero, M. F. Traxler, D. López, R. Kolter, *Chem. Rev.* **2011**, *111*, 5492–505.
[2] J. O'Brien, G. D. Wright, *Curr. Opin. Biotechnol.* **2011**, *22*, 552–8.
[3] A. Camilli, B. L. Bassler, *Science* **2006**, *311*, 1113–6.
[4] A. L. Demain, *Adv. Appl. Microbiol.* **1959**, *1*, 23–47.
[5] H. Jeong, Y. M. Sim, H. J. Kim, D.-W. Lee, S.-K. Lim, S. J. Lee, *Genome Announc.* **2013**, *1*, e00408-13.
[6] S. Javed, S. Bukhari, I. Zovia, M. Meraj, *Curr. Pharm. Biotechnol.* **2014**, *15*, 422–427.
[7] S. Gagneux, *Nat. Rev. Microbiol.* **2018**, *16*, 202–213.
[8] D. Brites, S. Gagneux, *Infect. Genet. Evol.* **2012**, *12*, 678–685.
[9] P. J. Brennan, *Tuberculosis (Edinb).* **2003**, *83*, 91–7.
[10] H. Marrakchi, M.-A. Lanéelle, M. Daffé, *Chem. Biol.* **2014**, *21*, 67–85.
[11] K. Takayama, C. Wang, G. S. Besra, *Clin. Microbiol. Rev.* **2005**, *18*, 81–101.
[12] A. Bhatt, V. Molle, G. S. Besra, W. R. Jacobs, L. Kremer, *Mol. Microbiol.* **2007**, *64*, 1442–1454.
[13] M. Jain, C. J. Petzold, M. W. Schelle, M. D. Leavell, J. D. Mougous, C. R. Bertozzi, J. A. Leary, J. S. Cox, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 5133–8.
[14] M. A. Velasco-Velázquez, D. Barrera, A. González-Arenas, C. Rosales, J. Agramonte-Hevia, *Microb. Pathog.* **2003**, *35*, 125–31.
[15] C. Astarie-Dequeker, L. Le Guyader, W. Malaga, F.-K. Seaphanh, C. Chalut, A. Lopez, C. Guilhot, *PLoS Pathog.* **2009**, *5*, e1000289.
[16] J. Augenstreich, A. Arbues, R. Simeone, E. Haanappel, A. Wegener, F. Sayes, F. Le Chevalier, C. Chalut, W. Malaga, C. Guilhot, et al., *Cell. Microbiol.* **2017**, *19*, e12726.
[17] A. Aporta, A. Arbues, J. I. Aguilo, M. Monzon, J. J. Badiola, A. de Martino, N. Ferrer, D. Marinova, A. Anel, C. Martin, et al., *PLoS One* **2012**, *7*, e45213.
[18] O. A. Trivedi, P. Arora, A. Vats, M. Z. Ansari, R. Tickoo, V. Sridharan, D. Mohanty, R. S. Gokhale, *Mol. Cell* **2005**, *17*, 631–643.
[19] P. Brodin, Y. Poquet, F. Levillain, I. Peguillet, G. Larrouy-Maumus, M. Gilleron, F. Ewann, T. Christophe, D. Fenistein, J. Jang, et al., *PLoS Pathog.* **2010**, *6*, e1001100.
[20] A. K. Azad, T. D. Sirakova, L. M. Rogers, P. E. Kolattukudy, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 4787–92.
[21] J. Buglino, K. C. Onwueme, J. A. Ferreras, L. E. N. Quadri, C. D. Lima, *J. Biol. Chem.* **2004**, *279*, 30634–42.
[22] J. S. Cox, B. Chen, M. McNeil, W. R. Jacobs, *Nature* **1999**, *402*, 79–83.
[23] M. Jain, J. S. Cox, *PLoS Pathog.* **2005**, *1*, e2.
[24] P. Domenech, M. B. Reed, *Microbiology* **2009**, *155*, 3532–43.
[25] M. Sritharan, *J. Bacteriol.* **2016**, *198*, 2399–409.
[26] J. R. Chipperfield, C. Ratledge, *BioMetals* **2000**, *13*, 165–168.
[27] J. J. De Voss, K. Rutter, B. G. Schroeder, H. Su, Y. Zhu, C. E. Barry, *Proc. Natl.*

*Acad. Sci. U. S. A.* **2000**, *97*, 1252–7.
[28] M. Curto, C. Reali, G. Palmieri, F. Scintu, M. L. Schivo, V. Sogos, M. A. Marcialis, M. G. Ennas, H. Schwarz, G. Pozzi, et al., *Neurochem. Int.* **2004**, *44*, 381–92.
[29] S. Pelly, K. Winglee, F. F. Xia, R. L. Stevens, W. R. Bishai, G. Lamichhane, *Tuberculosis (Edinb).* **2016**, *99*, 70–80.
[30] M. D. McMahon, J. S. Rush, M. G. Thomas, *J. Bacteriol.* **2012**, *194*, 2809–18.
[31] E. Tatham, S. Sundaram Chavadi, P. Mohandas, U. R. Edupuganti, S. K. Angala, D. Chatterjee, L. E. N. Quadri, *BMC Microbiol.* **2012**, *12*, 118.
[32] B. Boll, T. Taubitz, L. Heide, *J. Biol. Chem.* **2011**, *286*, 36281–90.
[33] K. J. Esquilín-Lebrón, T. O. Boynton, L. J. Shimkets, M. G. Thomas, *J. Bacteriol.* **2018**, *200*, e00346-18.
[34] S. J. Sasindran, J. B. Torrelles, *Front. Microbiol.* **2011**, *2*, 2.
[35] M. J. Marakalala, L. M. Graham, G. D. Brown, *Clin. Dev. Immunol.* **2010**, *2010*, 567571.
[36] R. Lang, *Front. Immunol.* **2013**, *4*, 5.
[37] K. M. Backus, M. A. Dolan, C. S. Barry, M. Joe, P. McPhie, H. I. M. Boshoff, T. L. Lowary, B. G. Davis, C. E. Barry, *J. Biol. Chem.* **2014**, *289*, 25041–53.
[38] A. A. Elamin, M. Stehr, R. Spallek, M. Rohde, M. Singh, *Mol. Microbiol.* **2011**, *81*, 1577–1592.
[39] L. Favrot, D. H. Lajiness, D. R. Ronning, *J. Biol. Chem.* **2014**, *289*, 25031–40.
[40] T. Warrier, M. Tropis, J. Werngren, A. Diehl, M. Gengenbacher, B. Schlegel, M. Schade, H. Oschkinat, M. Daffe, S. Hoffner, et al., *Antimicrob. Agents Chemother.* **2012**, *56*, 1735–43.
[41] H. S. La Pierre, J. Arnold, R. G. Bergman, F. D. Toste, *Inorg. Chem.* **2012**, *51*, 13334–13344.
[42] M. F. Clarke-Pearson, S. F. Brady, *J. Bacteriol.* **2008**, *190*, 6927–6930.
[43] J. M. Crawford, C. Portmann, X. Zhang, M. B. J. Roeffaers, J. Clardy, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 10821–6.
[44] L. Wang, M. Zhu, Q. Zhang, X. Zhang, P. Yang, Z. Liu, Y. Deng, Y. Zhu, X. Huang, L. Han, et al., *ACS Chem. Biol.* **2017**, *12*, 3067–3075.
[45] M. J. Garson, J. S. Simpson, *Nat. Prod. Rep.* **2004**, *21*, 164.
[46] A. Park, R. E. Moore, G. M. L. Patterson, *Tetrahedron Lett.* **1992**, *33*, 3257–3260.
[47] S. F. Brady, J. Clardy, *Angew. Chemie - Int. Ed.* **2005**, *44*, 7063–7065.
[48] W. Rothe, *Pharmazie* **1950**, *5*, 190.
[49] B. J. Burreson, P. J. Scheuer, J. Finer, J. Clardy, *J. Am. Chem. Soc.* **1975**, *97*, 4763–4764.
[50] C. W. J. Chang, A. Patra, D. M. Roll, P. J. Scheuer, G. K. Matsumoto, J. Clardy, *J. Am. Chem. Soc.* **1984**, *106*, 4644–4646.
[51] B. J. Burreson, C. Christophersen, P. J. Scheuer, *Tetrahedron* **1975**, *31*, 2015–2018.
[52] L. Minale, R. Riccio, G. Sodano, *Tetrahedron* **1974**, *30*, 1341–1343.
[53] B. Di Blasio, E. Fattorusso, S. Magno, L. Mayol, C. Pedone, C. Santacroce, D. Sica, *Tetrahedron* **1976**, *32*, 473–478.
[54] F. Cafieri, E. Fattorusso, S. Magno, C. Santacroce, D. Sica, *Tetrahedron* **1973**, *29*, 4259–4262.

[55] S. Mo, A. Krunic, G. Chlipala, J. Orjala, *J. Nat. Prod.* **2009**, *72*, 894–899.
[56] K. Stratmann, R. E. Moore, G. M. L. Patterson, R. Bonjouklian, J. B. Deeter, S. Shaffer, T. A. Smitka, C. D. Smith, *J. Am. Chem. Soc.* **1994**, *116*, 9935–9942.
[57] F. Y. Lim, T. H. Won, N. Raffa, J. A. Baccile, J. Wisecaver, A. Rokas, F. C. Schroeder, N. P. Keller, *MBio* **2018**, *9*, e00785-18.
[58] S. F. Brady, J. Clardy, *Angew. Chemie - Int. Ed.* **2005**, *44*, 7045–7048.
[59] J. B. Gloer, G. K. Poch, D. M. Short, D. V. McCloskey, *J. Org. Chem.* **1988**, *53*, 3758–3761.
[60] M. J. Schnermann, R. A. Shenvi, *Nat. Prod. Rep.* **2015**, *32*, 543–577.
[61] J. Emsermann, U. Kauhl, T. Opatz, J. Emsermann, U. Kauhl, T. Opatz, *Mar. Drugs* **2016**, *14*, 16.
[62] *,† Emiliano Manzo, † M. Letizia Ciavatta, † Margherita Gavagnin, † Ernesto Mollo, *,‡ and Yue-Wei Guo, G. Cimino†, **2004**, DOI 10.1021/NP0400961.
[63] M. L. Hillwig, Q. Zhu, X. Liu, *ACS Chem. Biol.* **2014**, *9*, 372–377.
[64] M. S. C. Pedras, E. E. Yaya, *Org. Biomol. Chem.* **2012**, *10*, 3613.
[65] E. J. Drake, A. M. Gulick, *J. Mol. Biol.* **2008**, *384*, 193–205.
[66] U. Qaisar, C. J. Kruczek, M. Azeem, N. Javaid, J. A. Colmer-Hamood, A. N. Hamood, *J. Microbiol.* **2016**, *54*, 573–581.
[67] N. L. Carty, N. Layland, J. A. Colmer-Hamood, M. W. Calfee, E. C. Pesci, A. N. Hamood, *Mol. Microbiol.* **2006**, *61*, 782–794.
[68] A. N. Hamood, J. A. Colmer, U. A. Ochsner, M. L. Vasil, *Mol. Microbiol.* **1996**, *21*, 97–110.
[69] M. S. C. Pedras, P. B. Chumala, W. Jin, M. S. Islam, D. W. Hauck, *Phytochemistry* **2009**, *70*, 394–402.
[70] M. S. C. Pedras, M. R. Park, *Phytochemistry* **2016**, *132*, 26–32.
[71] H. Kan, C.-H. Kim, H.-M. Kwon, J.-W. Park, K.-B. Roh, H. Lee, B.-J. Park, R. Zhang, J. Zhang, K. Söderhäll, et al., *J. Biol. Chem.* **2008**, *283*, 25316–23.
[72] I. Eleftherianos, P. J. Millichap, R. H. ffrench-Constant, S. E. Reynolds, *Dev. Comp. Immunol.* **2006**, *30*, 1099–107.
[73] P. G. Becher, S. Keller, G. Jung, R. D. Süssmuth, F. Jüttner, *Phytochemistry* **2007**, *68*, 2493–7.
[74] T. A. Smitka, R. Bonjouklian, L. Doolin, N. D. Jones, J. B. Deeter, W. Y. Yoshida, M. R. Prinsep, R. E. Moore, G. M. L. Patterson, *J. Org. Chem.* **1992**, *57*, 857–861.
[75] S. Mo, A. Krunic, B. D. Santarsiero, S. G. Franzblau, J. Orjala, *Phytochemistry* **2010**, *71*, 2116–2123.
[76] R. E. Moore, C. Cheuk, X. Q. G. Yang, G. M. L. Patterson, R. Bonjouklian, T. A. Smitka, J. S. Mynderse, R. S. Foster, N. D. Jones, J. K. Swartzendruber, et al., *J. Org. Chem.* **1987**, *52*, 1036–1043.
[77] C. J. R. Fookes, M. J. Garson, J. K. MacLeod, B. W. Skelton, A. H. White, *J. Chem. Soc. Perkin Trans. 1* **1988**, *0*, 1003.
[78] M. J. Garson, *J. Chem. Soc. Chem. Commun.* **1986**, *0*, 35.
[79] P. Karuso, P. J. Scheuer, *J. Org. Chem.* **1989**, *54*, 2092–2095.
[80] G. H. Hur, C. R. Vickery, M. D. Burkart, *Nat. Prod. Rep.* **2012**, *29*, 1074–98.
[81] L. E. Quadri, J. Sello, T. A. Keating, P. H. Weinreb, C. T. Walsh, *Chem. Biol.* **1998**, *5*, 631–45.
[82] D. Giovannini, G. Cappelli, L. Jiang, C. Castilletti, A. Colone, A. Serafino, F.

Wannenes, L. Giacò, G. Quintiliani, M. Fraziano, et al., *Microb. Pathog.* **2012**, *53*, 135–46.
[83] N. Dhar, J. D. McKinney, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 12275–80.
[84] H. Zheng, L. Lu, B. Wang, S. Pu, X. Zhang, G. Zhu, W. Shi, L. Zhang, H. Wang, S. Wang, et al., *PLoS One* **2008**, *3*, e2375.
[85] G. S. Hotter, B. J. Wards, P. Mouat, G. S. Besra, J. Gomes, M. Singh, S. Bassett, P. Kawakami, P. R. Wheeler, G. W. de Lisle, et al., *J. Bacteriol.* **2005**, *187*, 2267–77.
[86] C. M. Sassetti, E. J. Rubin, *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 12989–94.
[87] G. Cappelli, E. Volpe, M. Grassi, B. Liseo, V. Colizzi, F. Mariani, *Res. Microbiol.* **2006**, *157*, 445–55.
[88] J. Rengarajan, B. Bloom, E. Rubin, *Proc Natl Acad Sci USA* **n.d.**, *102*, DOI 10.1073/pnas.0503272102.
[89] C. M. Sassetti, D. H. Boyd, E. J. Rubin, *Mol. Microbiol.* **2003**, *48*, 77–84.
[90] G. Lamichhane, M. Zignol, N. J. Blades, D. E. Geiman, A. Dougherty, J. Grosset, K. W. Broman, W. R. Bishai, *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 7213–8.
[91] W. Zhang, B. Ostash, C. Walsh, *Proc Natl Acad Sci USA* **n.d.**, *107*, DOI 10.1073/pnas.1011557107.
[92] Z. Liu, T. R. Ioerger, F. Wang, J. C. Sacchettini, *J. Biol. Chem.* **2013**, *288*, 18473–83.
[93] A. Chhabra, *Proc Natl Acad Sci USA* **n.d.**, *109*, DOI 10.1073/pnas.1118680109.
[94] R. Tautenhahn, G. Patti, D. Rinehart, G. Siuzdak, *Anal. Chem. (Washington, DC)* **n.d.**, *84*, 5035–5039.
[95] A. Saghatelian, S. A. Trauger, E. J. Want, E. G. Hawkins, G. Siuzdak, B. F. Cravatt, *Biochemistry* **2004**, *43*, 14332–9.
[96] R. Stephany, M. Bie, W. Drenth, *Org Magn Reson.* **n.d.**, *6*, 45–47.
[97] R. Bhushan, H. Brückner, *Amino Acids* **2004**, *27*, 231–47.
[98] S.-I. Amano, T. Sakurai, K. Endo, H. Takano, T. Beppu, K. Furihata, S. Sakuda, K. Ueda, *J. Antibiot. (Tokyo).* **2011**, *64*, 703.
[99] Y. Tabata, M. Hatsu, S. Amano, A. Shimizu, S. Imai, *Sci Rep Meiji Seika Kaisha* **n.d.**, *34*, 1–9.
[100] T. Sasaki, *Sci Rep Meiji Seika Kaisha* **n.d.**, *26*, 10–16.
[101] F. Wang, R. Langley, G. Gulten, L. Wang, J. Sacchettini, *Chem. Biol.* **n.d.**, *14*, 543–551.
[102] W. Huang, S. Kim, J. Liu, W. Zhang, *Org Lett* **n.d.**, *17*, 5344–5347.
[103] M. Clarke-Pearson, S. Brady, *J Bacteriol* **n.d.**, *190*, DOI 10.1128/jb.00801-08.
[104] M. L. Micallef, D. Sharma, B. M. Bunn, L. Gerwick, R. Viswanathan, M. C. Moffitt, *BMC Microbiol.* **2014**, *14*, DOI 10.1186/s12866-014-0213-7.
[105] K. Amagai, R. Takaku, F. Kudo, T. Eguchi, *ChemBioChem* **n.d.**, *14*, 1998–2006.
[106] K. Maity, P. Bajaj, N. Surolia, A. Surolia, K. Suguna, *J. Biomol. Struct. Dyn.* **2012**, *29*, 973–983.
[107] K. Haug, R. M. Salek, P. Conesa, J. Hastings, P. de Matos, M. Rijnbeek, T. Mahendraker, M. Williams, S. Neumann, P. Rocca-Serra, et al., *Nucleic Acids Res.* **2013**, *41*, D781–D786.
[108] T. Yamaguchi, Y. Miyake, A. Miyamura, N. Ishiwata, K. Tatsuta, *J. Antibiot. (Tokyo).* **2006**, *59*, 729–734.
[109] W. chen Chang, D. Sanyal, J. L. Huang, K. Ittiamornkul, Q. Zhu, X. Liu, *Org. Lett.*

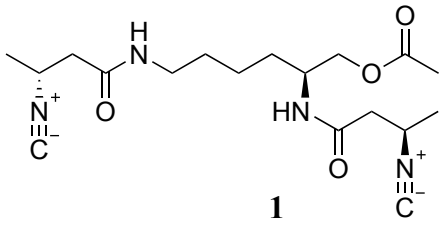
**2017**, *19*, 1208–1211.
[110] N. C. Harris, M. Sato, N. A. Herman, F. Twigg, W. Cai, J. Liu, X. Zhu, J. Downey, R. Khalaf, J. Martin, et al., *Proc. Natl. Acad. Sci.* **2017**, *114*, 7025–7030.
[111] K. Amagai, R. Takaku, F. Kudo, T. Eguchi, *ChemBioChem* **2013**, *14*, 1998–2006.
[112] T. Chisuga, A. Miyanaga, F. Kudo, T. Eguchi, *J. Biol. Chem.* **2017**, *292*, 10926–10937.
[113] L. Holm, L. M. Laakso, *Nucleic Acids Res.* **2016**, *44*, W351–W355.
[114] J. M. Elkins, M. J. Ryle, I. J. Clifton, J. C. Dunning Hotopp, J. S. Lloyd, N. I. Burzlaff, J. E. Baldwin, R. P. Hausinger, P. L. Roach, *Biochemistry* **2002**, *41*, 5185–92.
[115] J. R. O'Brien, D. J. Schuller, V. S. Yang, B. D. Dillard, W. N. Lanzilotta, *Biochemistry* **2003**, *42*, 5547–5554.
[116] L. Zimmermann, A. Stephens, S. Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A. N. Lupas, V. Alva, *J. Mol. Biol.* **2017**, DOI 10.1016/j.jmb.2017.12.007.
[117] J. Notredame, C., Higgins, D. G., & Heringa, C. Notredame, D. G. Higgins, J. Heringa, *J. Mol. Biol.* **2000**, *302*, 205–217.
[118] M. W. Eshoo, *J. Bacteriol.* **1988**, *170*, 5208–5215.
[119] P. K. Grzyska, E. H. Appelman, R. P. Hausinger, D. A. Proshlyakov, *Proc. Natl. Acad. Sci.* **2010**, *107*, 3982–3987.
[120] W. Kabsch, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 125–132.
[121] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Crystallogr.* **2007**, *40*, 658–674.
[122] G. Bunkóczi, R. J. Read, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67*, 303–312.
[123] T. C. Terwilliger, R. W. Grosse-Kunstleve, P. V. Afonine, N. W. Moriarty, P. H. Zwart, L.-W. Hung, R. J. Read, P. D. Adams, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2008**, *64*, 61–69.
[124] P. Emsley, K. Cowtan, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60*, 2126–2132.
[125] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, et al., *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 213–221.
[126] O. S. Smart, T. O. Womack, *Glob. Phasing Ltd.* **2011**.
[127] S. Lucarini, C. Tomasini, *J. Org. Chem.* **2001**, *66*, 727–732.
[128] W. Zhang, M. L. Bolla, D. Kahne, C. T. Walsh, *J. Am. Chem. Soc.* **2010**, *132*, 6402–6411.
[129] Y. Yamamoto, T. Shirai, N. Miyaura, *Chem. Commun.* **2012**, *48*, 2803.
[130] T. Buyck, Q. Wang, J. Zhu, *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 12714–8.
[131] W. Cai, X. Wang, S. I. Elshahawi, L. V. Ponomareva, X. Liu, M. R. McErlean, Z. Cui, A. L. Arlinghaus, J. S. Thorson, S. G. Van Lanen, *J. Nat. Prod.* **2016**, *79*, 2731–2739.
[132] H. Stöckmann, A. A. Neves, S. Stairs, K. M. Brindle, F. J. Leeper, *Org. Biomol. Chem.* **2011**, *9*, 7303.
[133] G. Pishchany, E. Mevers, S. Ndousse-Fetter, D. J. Horvath, C. R. Paludo, E. A. Silva-Junior, S. Koren, E. P. Skaar, J. Clardy, R. Kolter, *Proc. Natl. Acad. Sci.*

**2018**, *115*, 10124–10129.
[134] C. Mokhtarzadeh, *Inorg Chem* **n.d.**, *54*, 5579–5587.
[135] Z. Ma, F. Jacobsen, D. Giedroc, *Chem. Rev.* **n.d.**, *109*, 4644–4681.
[136] L. Remy, M. Carrière, A. Derré-Bobillot, C. Martini, M. Sanguinetti, E. Borezée-Durant, *Mol. Microbiol.* **2013**, *87*, 730–743.
[137] G. Ghssein, C. Brutesco, L. Ouerdane, C. Fojcik, A. Izaute, S. Wang, C. Hajjar, R. Lobinski, D. Lemaire, P. Richaud, et al., *Science (80-. ).* **2016**, *352*, 1105–1109.
[138] H. Botella, *Cell Host Microbe* **n.d.**, *10*, 248–259.
[139] F. Wolschendorf, D. Ackart, T. B. Shrestha, L. Hascall-Dove, S. Nolan, G. Lamichhane, Y. Wang, S. H. Bossmann, R. J. Basaraba, M. Niederweis, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 1621–6.
[140] K. Chaturvedi, C. Hung, J. Crowley, A. Stapleton, J. Henderson, *Nat Chem Biol* **n.d.**, *8*, 731–736.
[141] T. Johnstone, E. Nolan, *Dalt. Trans* **n.d.**, *44*, 6320–6339.
[142] T. Rustad, *Genome Biol* **n.d.**, *15*, 502.
[143] S. Bardarov, *PNAS* **n.d.**, *94*, 10961–10966.
[144] N. Yan, *Trends Biochem. Sci.* **2013**, *38*, 151–159.
[145] B. R. Wilson, A. R. Bogdan, M. Miyazawa, K. Hashimoto, Y. Tsuji, *Trends Mol. Med.* **2016**, *22*, 1077–1090.

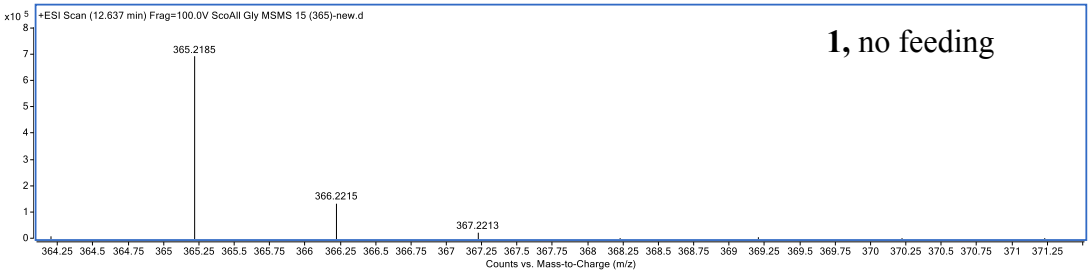
# Appendices

## A. HRMS analysis of 1 and 2.

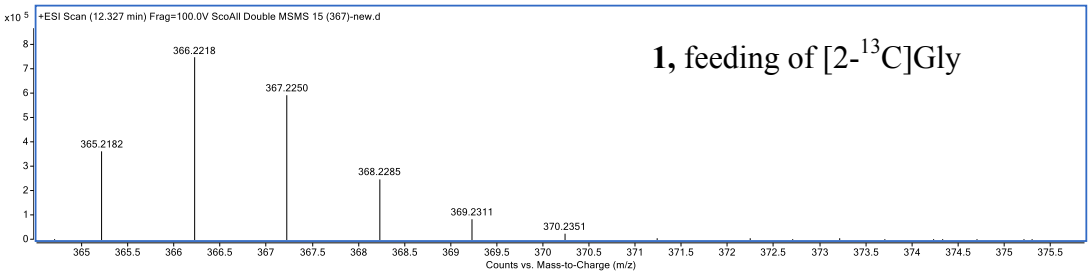
*m/z* $[M+H]^+$
Obs. 365.2185
Calc. 365.2183
Error: 0.5 ppm



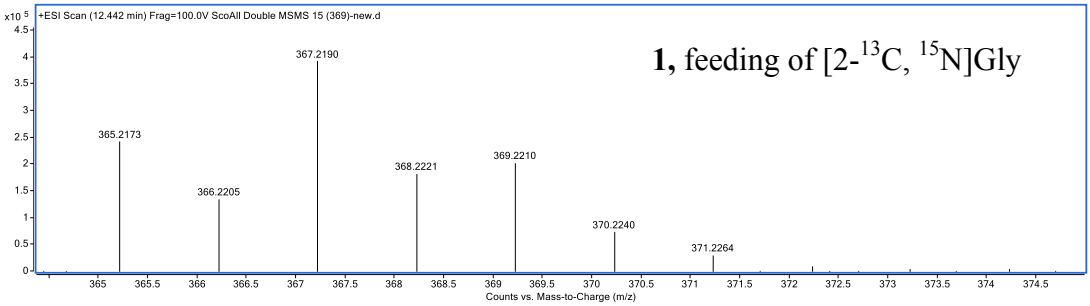
**1**



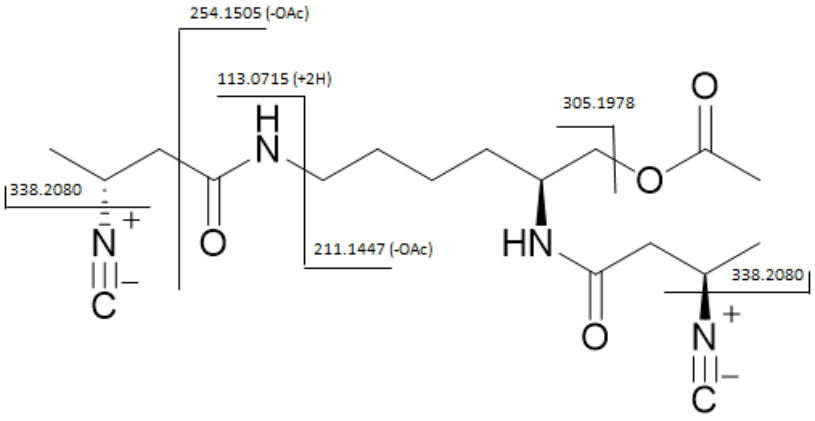

**1,** no feeding

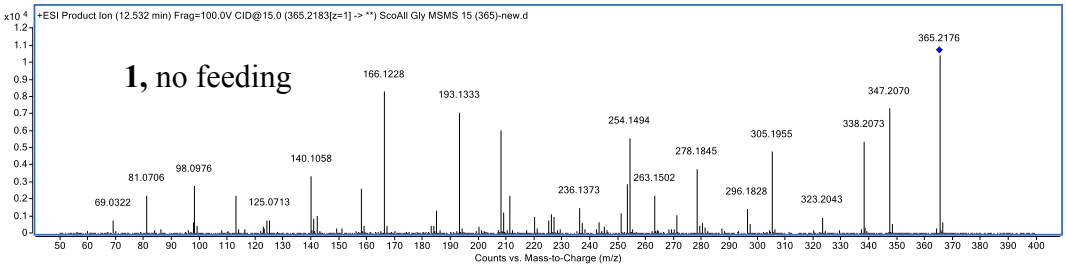


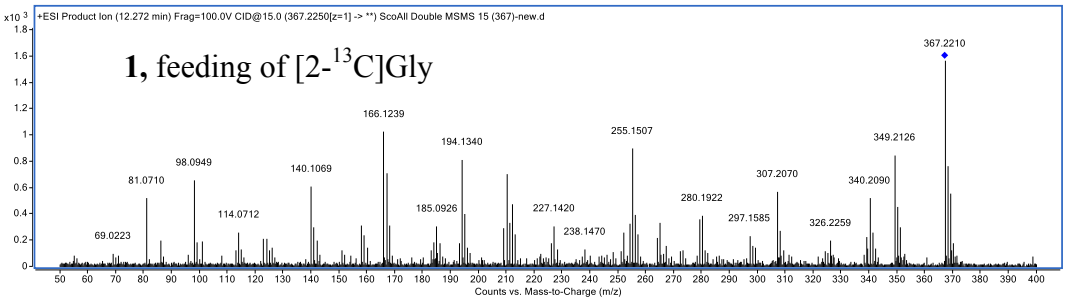

**1,** feeding of [2-$^{13}$C]Gly

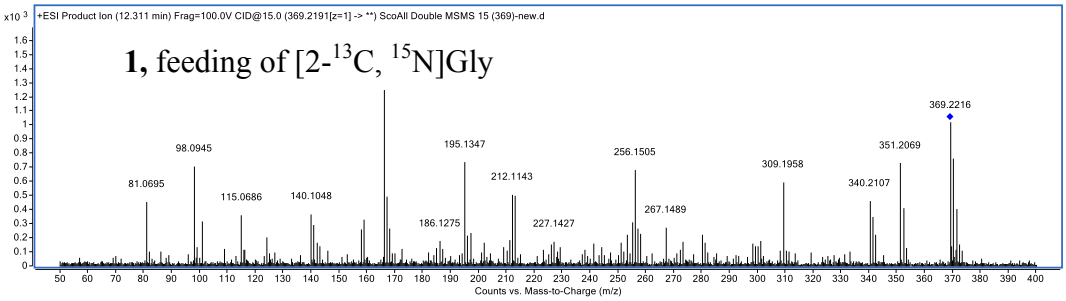



**1,** feeding of [2-$^{13}$C, $^{15}$N]Gly

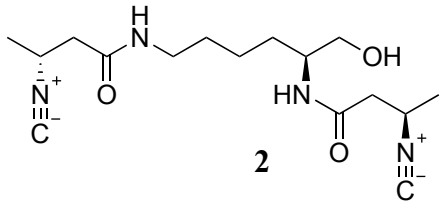

254.1505 (-OAc)
113.0715 (+2H)
305.1978
338.2080
211.1447 (-OAc)
338.2080


+ESI Product Ion (12.532 min) Frag=100.0V CID@15.0 (365.2183[z=1] -> **) ScoAll Gly MSMS 15 (365)-new.d
1, no feeding
365.2176
166.1228
193.1333
347.2070
254.1494
338.2073
305.1955
278.1845
140.1058
98.0976
81.0706
263.1502
236.1373
296.1828
69.0322
125.0713
323.2043
Counts vs. Mass-to-Charge (m/z)

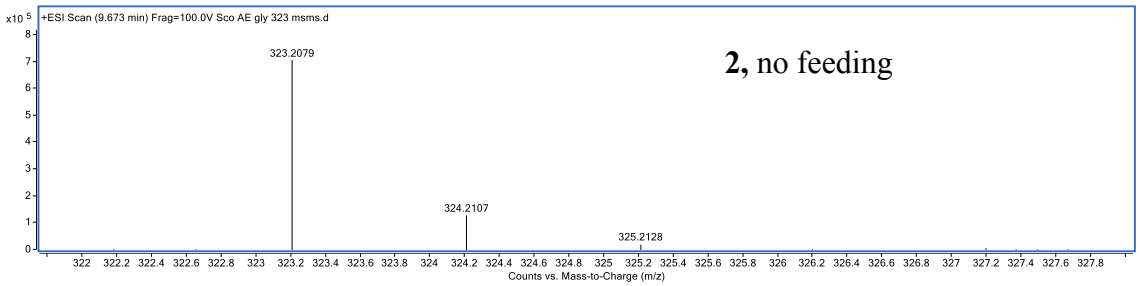

+ESI Product Ion (12.272 min) Frag=100.0V CID@15.0 (367.2250[z=1] -> **) ScoAll Double MSMS 15 (367)-new.d
1, feeding of [2-13C]Gly
367.2210
166.1239
255.1507
194.1340
349.2126
98.0949
140.1069
307.2070
81.0710
340.2090
280.1922
185.0926
227.1420
114.0712
297.1585
238.1470
326.2259
69.0223
Counts vs. Mass-to-Charge (m/z)


+ESI Product Ion (12.311 min) Frag=100.0V CID@15.0 (369.2191[z=1] -> **) ScoAll Double MSMS 15 (369)-new.d
1, feeding of [2-13C, 15N]Gly
369.2216
195.1347
256.1505
98.0945
309.1958
351.2069
212.1143
81.0695
340.2107
115.0686
140.1048
267.1489
186.1275
227.1427
Counts vs. Mass-to-Charge (m/z)

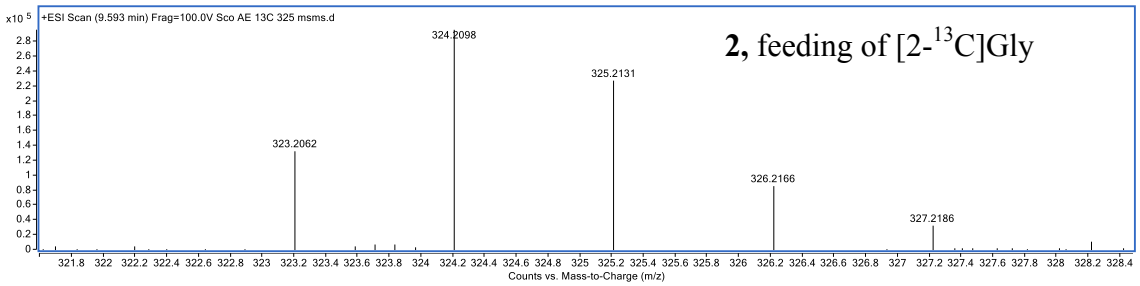
*m/z* $[M+H]^+$
Obs. 323.2079
Calc. 323.2078
Error: 0.3 ppm



**2**

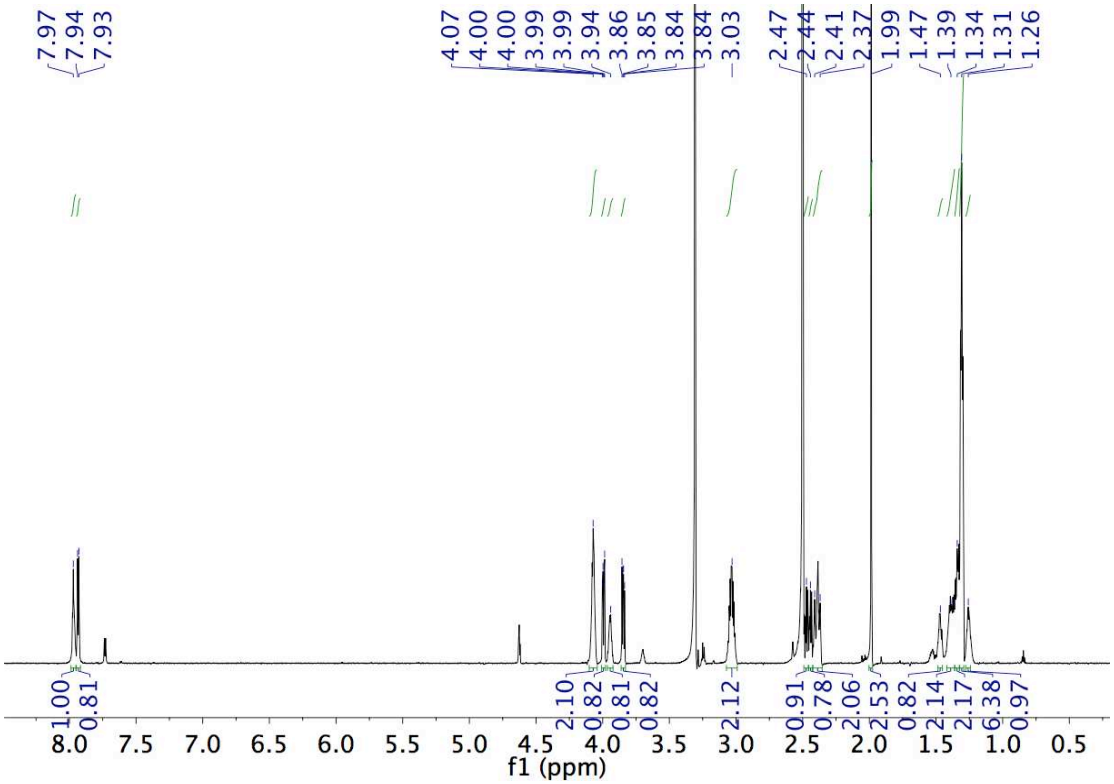



**2,** no feeding

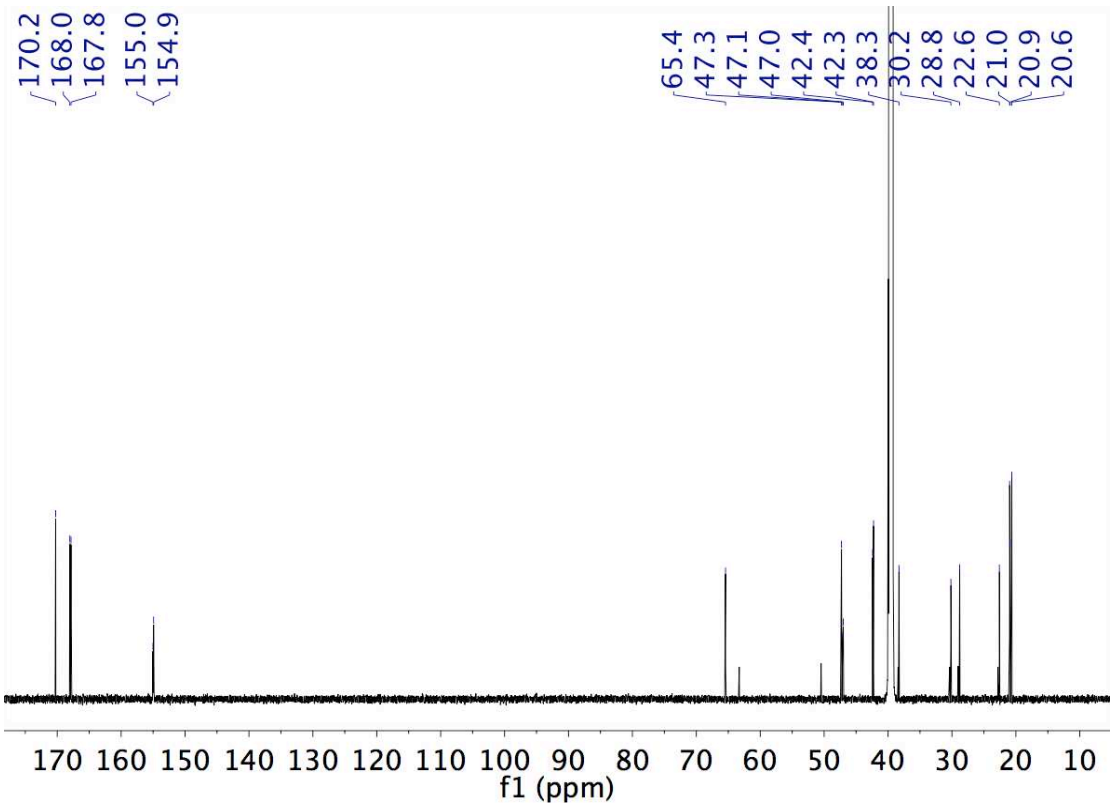



**2,** feeding of [2-$^{13}$C]Gly

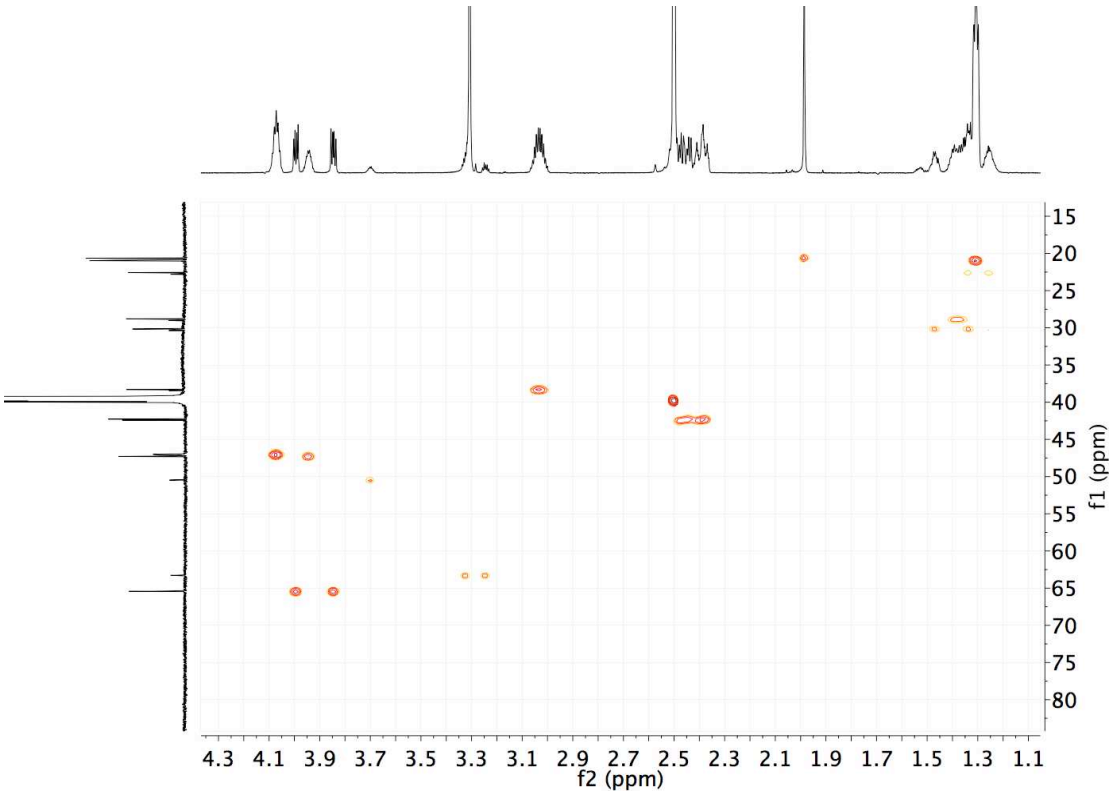
## B. NMR spectra of compound 1.



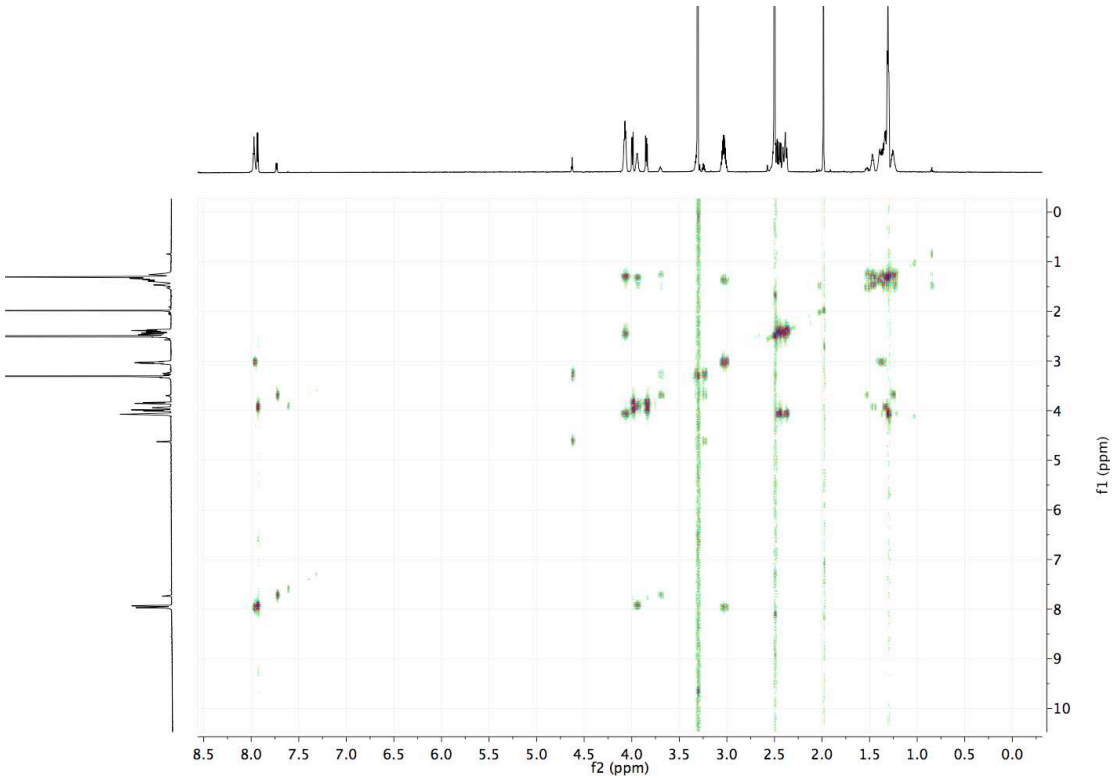

$^{1}$H NMR spectrum (DMSO-$d_6$, 900 MHz) of compound **1**.



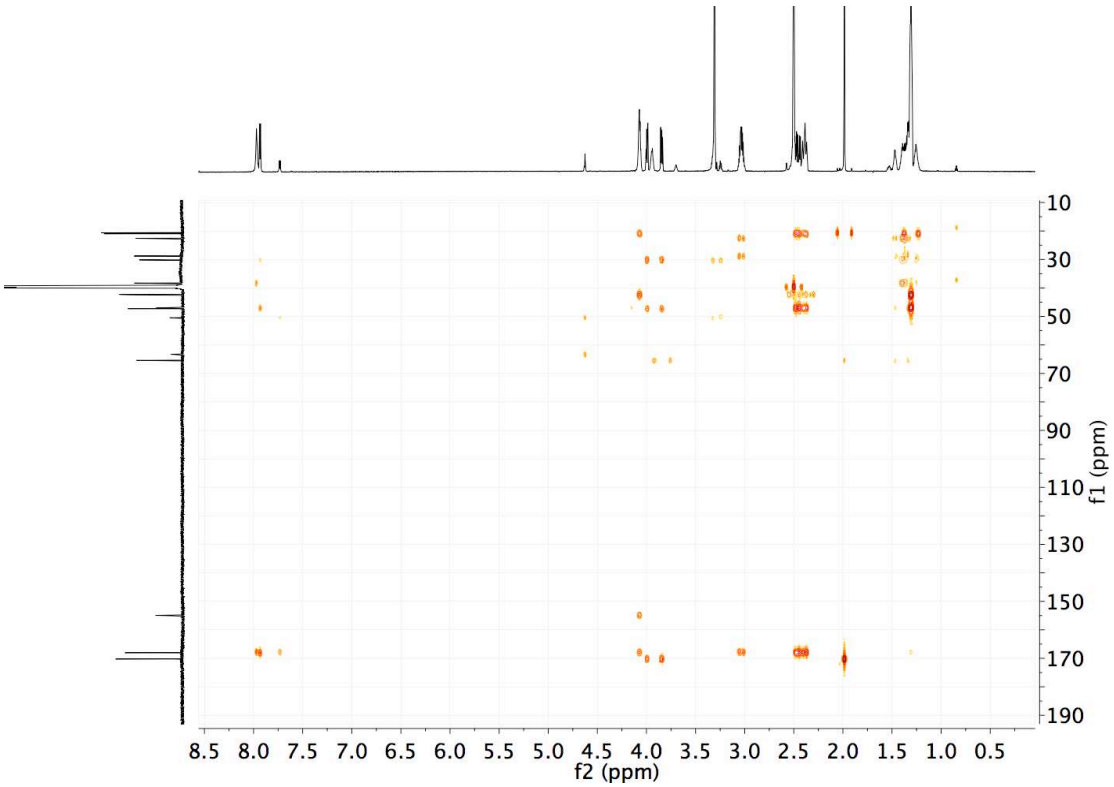

$^{13}$C NMR spectrum (DMSO-$d_6$, 225 MHz) of **1**



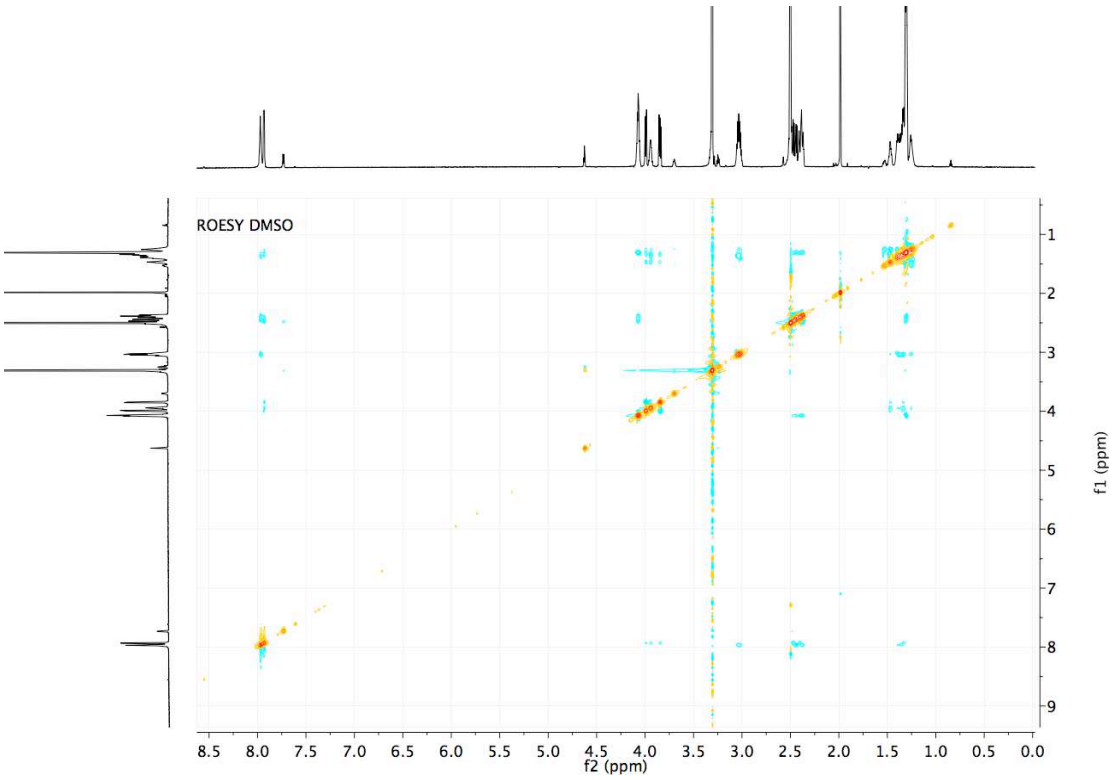

HSQC spectrum (DMSO-$d_6$, 900 MHz) of **1**.



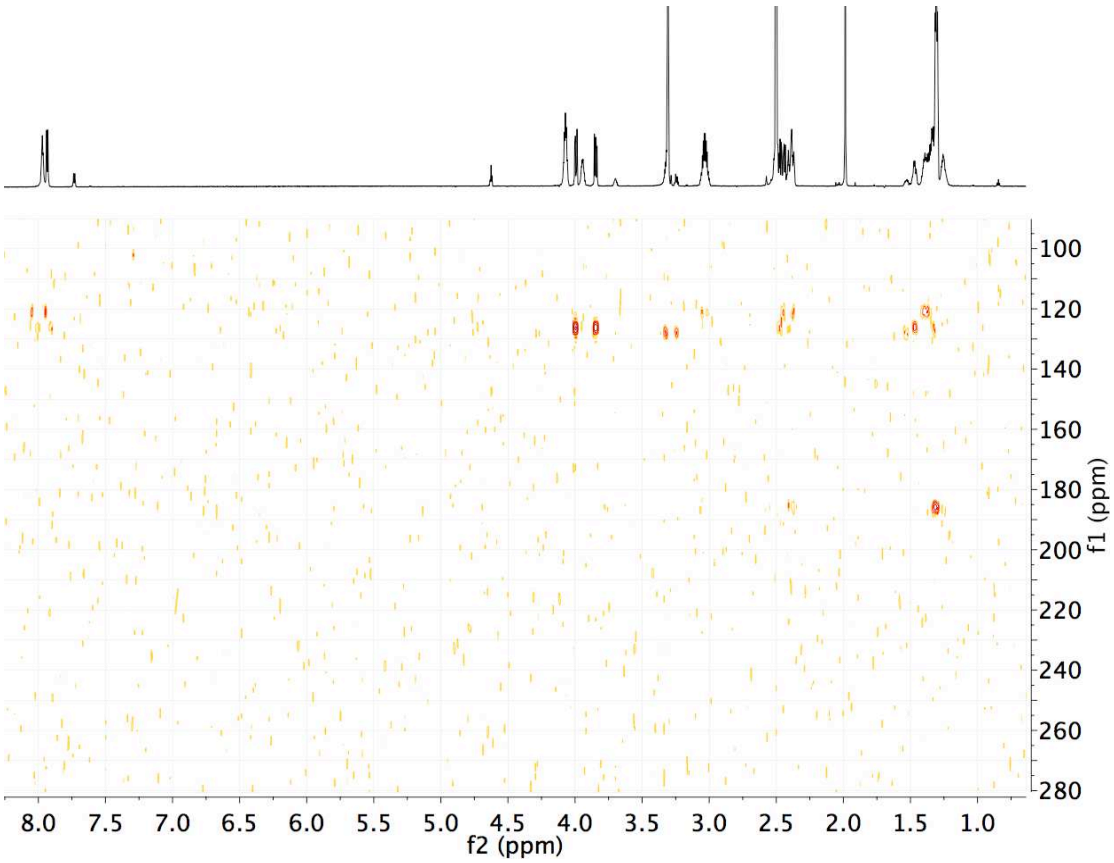

$^{1}$H, $^{1}$H-COSY spectrum (DMSO-$d_6$, 900 MHz) of **1.**




HMBC spectrum (DMSO-$d_6$, 900 MHz) of **1**.

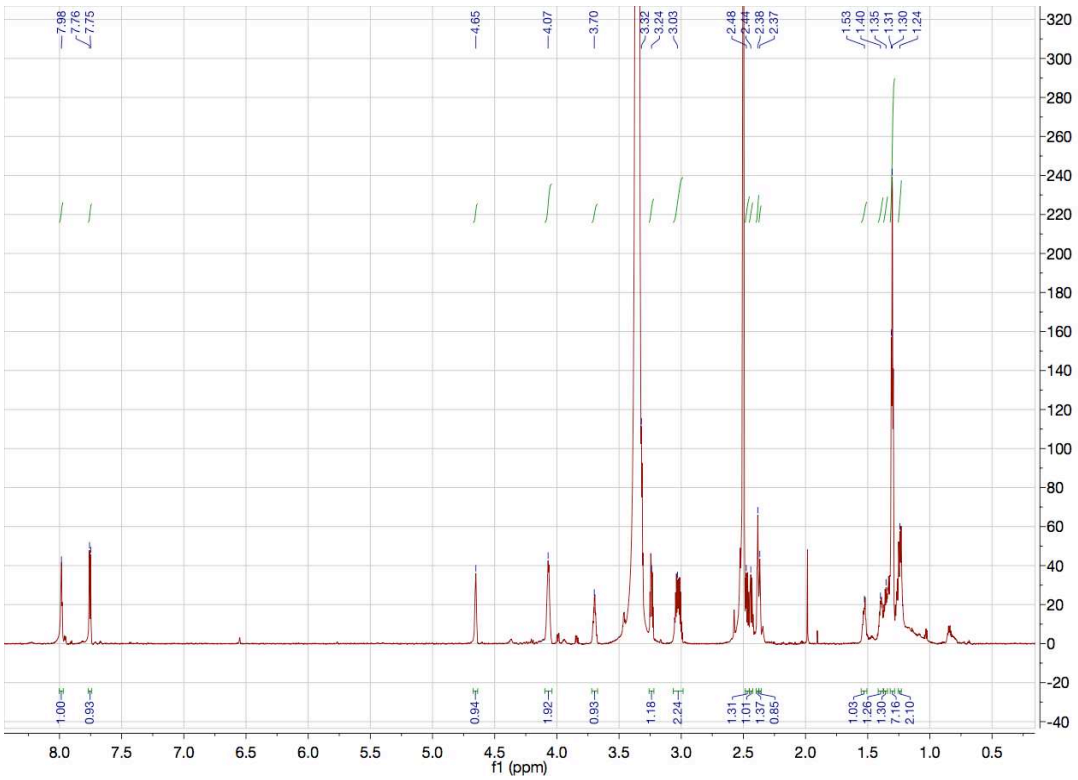



$^{1}$H, $^{1}$H-ROESY spectrum (DMSO-$d_6$, 900 MHz) of **1.**

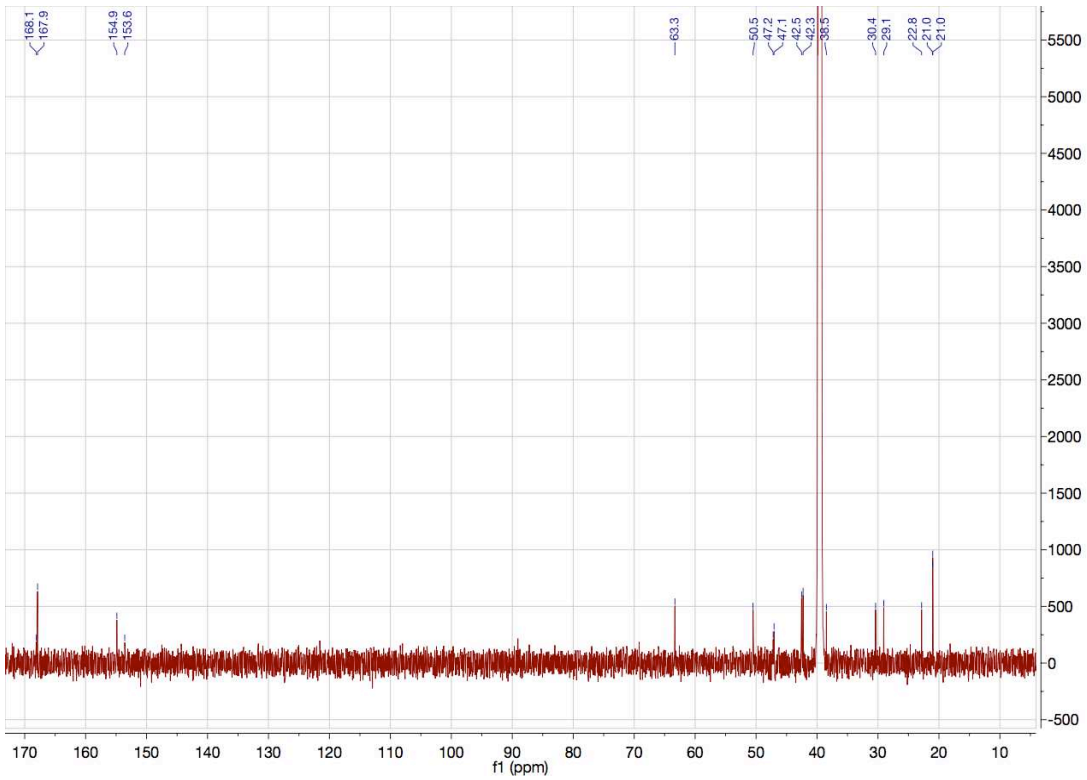



$^{1}H$, $^{15}N$-HMBC spectrum (DMSO-$d_6$, 900 MHz) of **1.**

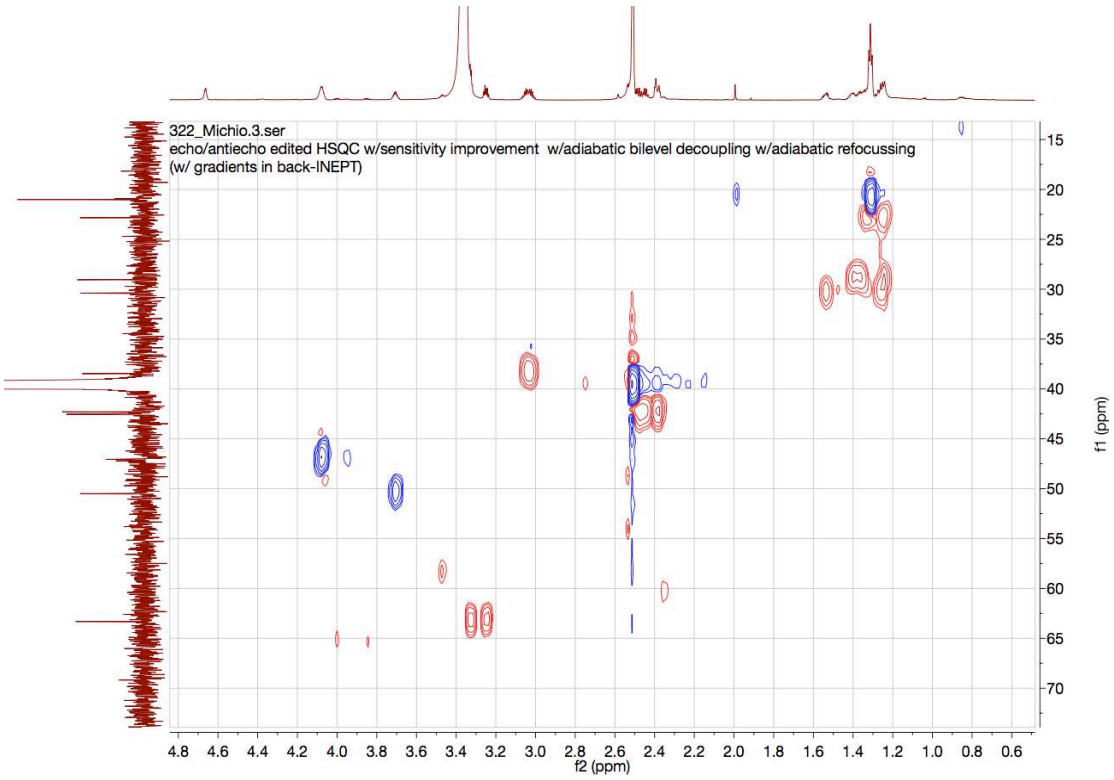
## C. NMR spectra of compound 2.



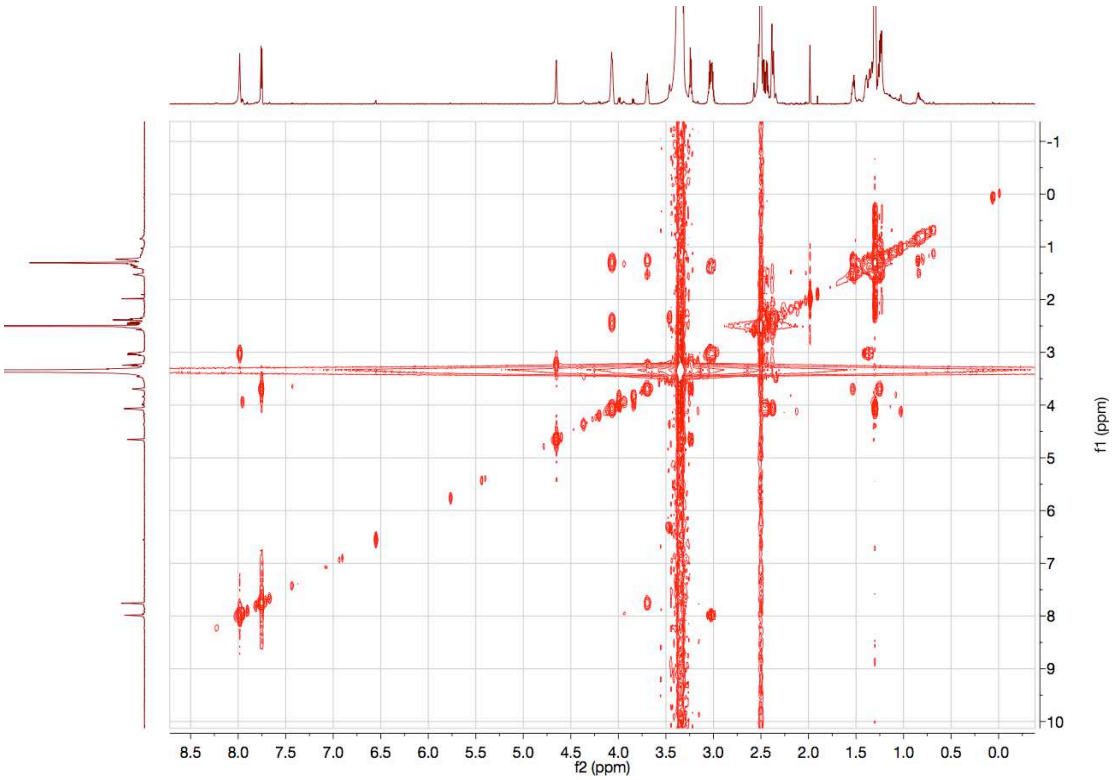

$^{1}$H NMR spectrum (DMSO-$d_6$, 900 MHz) of **2**.



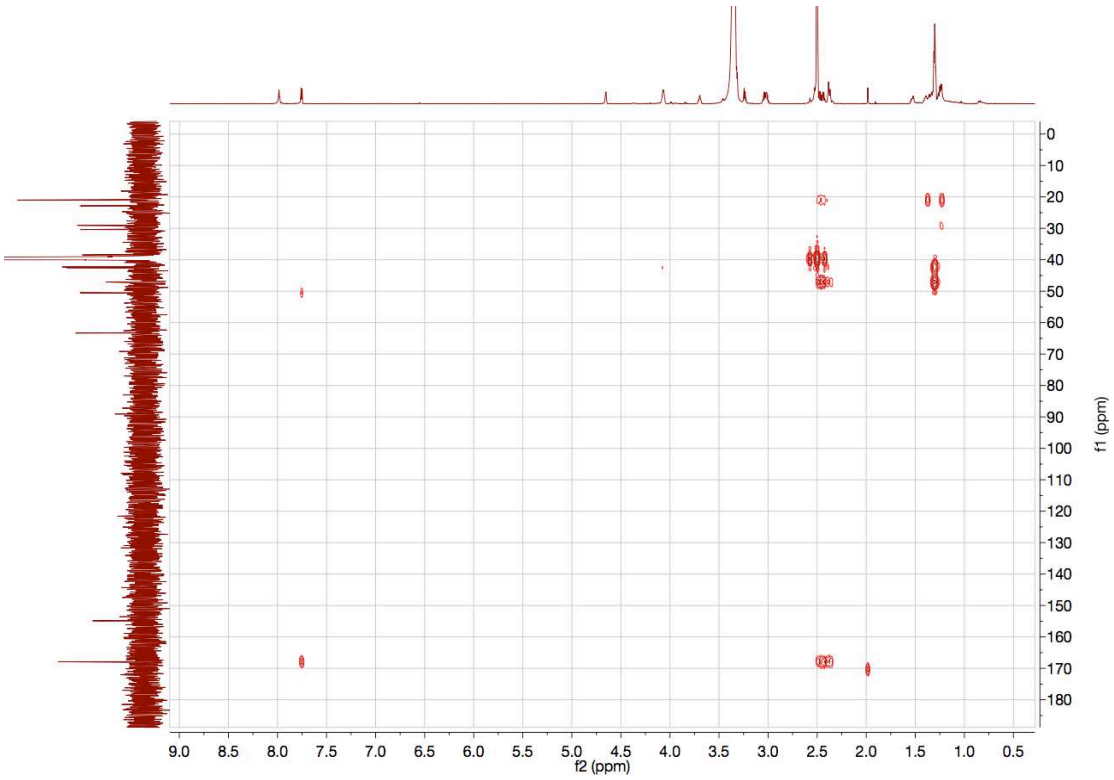

$^{13}$C NMR spectrum (DMSO-$d_6$, 225 MHz) of **2**.




HSQC spectrum (DMSO-$d_6$, 900 MHz) of **2**.



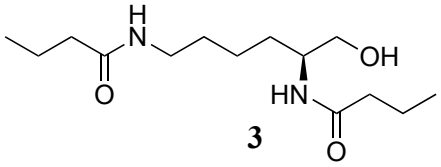

$^{1}$H, $^{1}$H-COSY spectrum (DMSO-$d_6$, 900 MHz) of **2.**



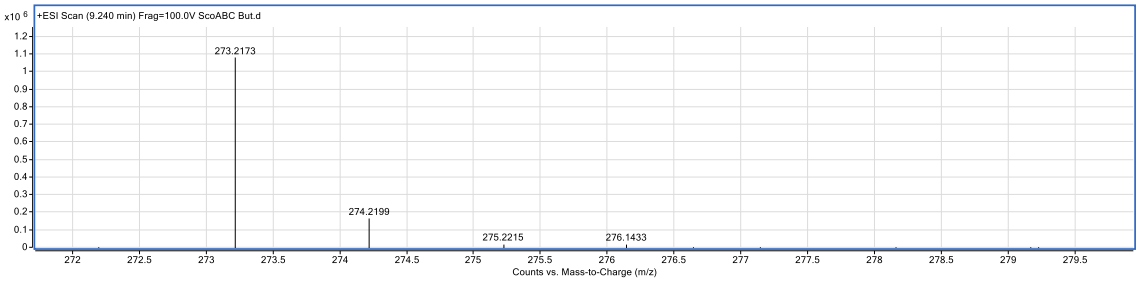

HMBC spectrum (DMSO-$d_6$, 900 MHz) of **2**.

## D. HRMS analysis of 3-5.

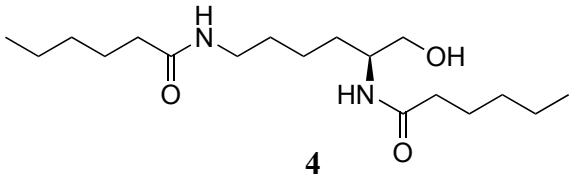
*m/z* $[M+H]^+$
Obs. 273.2173
Calc. 273.2173
Error: 0 ppm



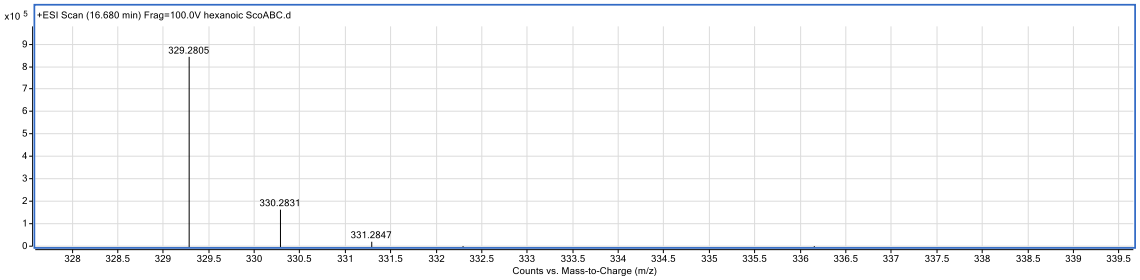

**3**



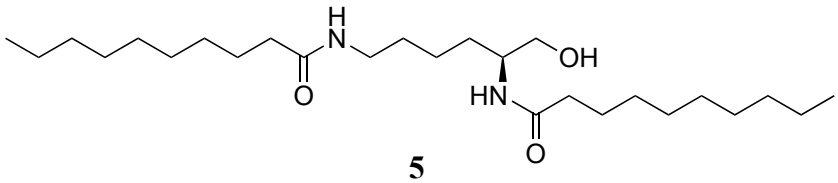

*m/z* $[M+H]^+$
Obs. 329.2805
Calc. 329.2799
Error: 1.8 ppm



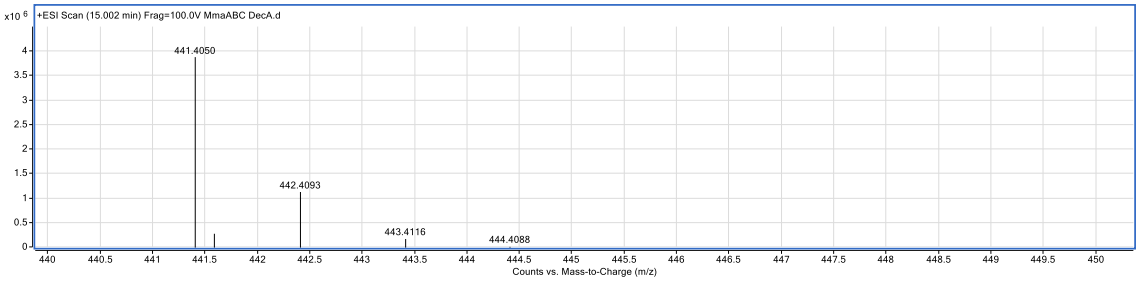

**4**



*m/z* $[M+H]^+$
Obs. 441.4051
Calc. 441.4050
Error: 0.2 ppm

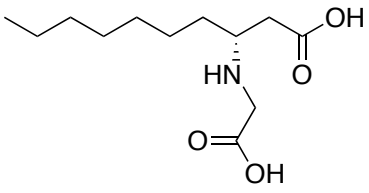


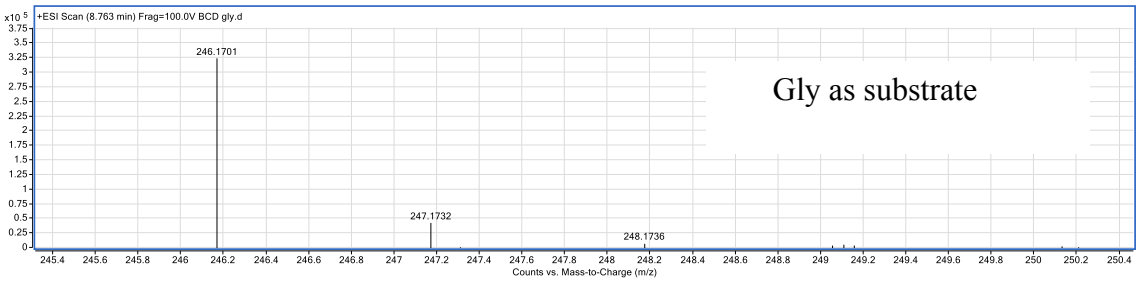

**5**

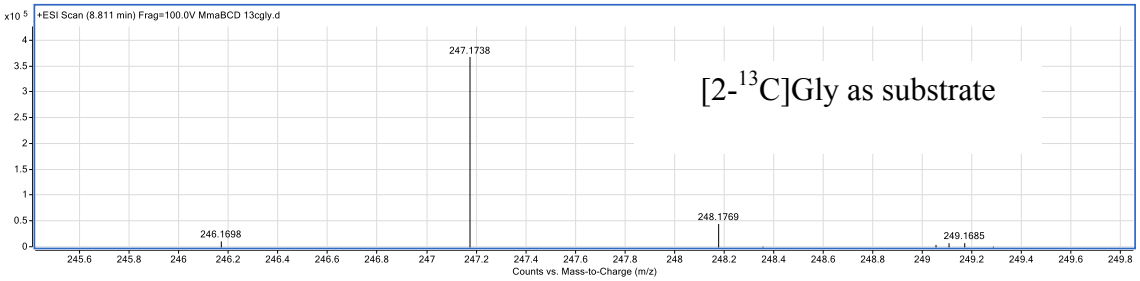


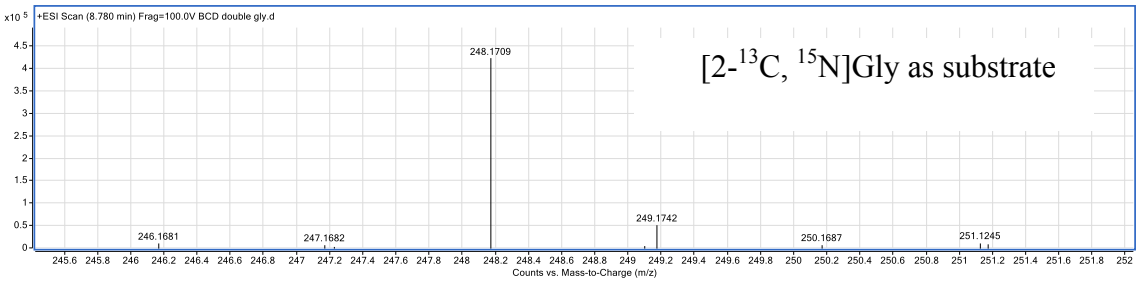

## E. HRMS analysis of MmaBCD in vitro reaction products.

*m/z* $[M+H]^+$
Obs. 246.1701
Calc. 246.1700
Error: 0.4 ppm



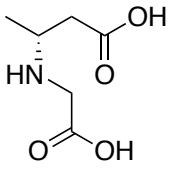



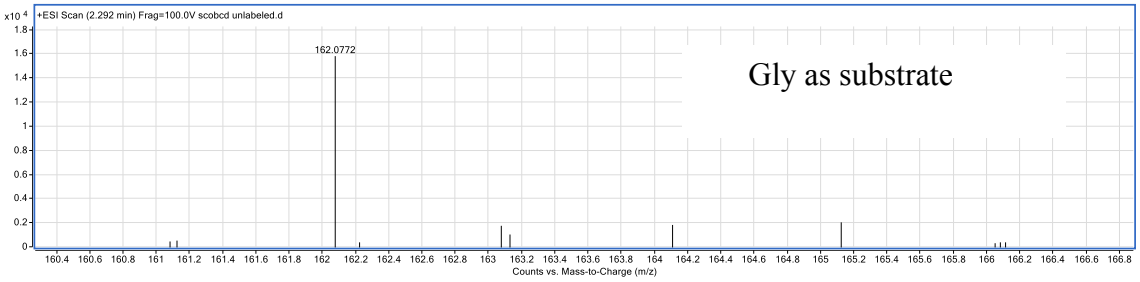


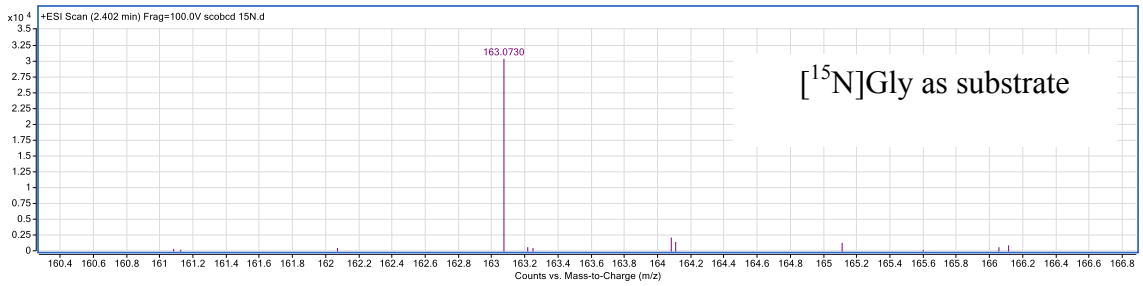




## F. HRMS analysis of ScoBCD *in vitro* reaction products.

*m/z* $[M+H]^+$
Obs. 162.0772
Calc. 162.0761
Error: 6.7 ppm

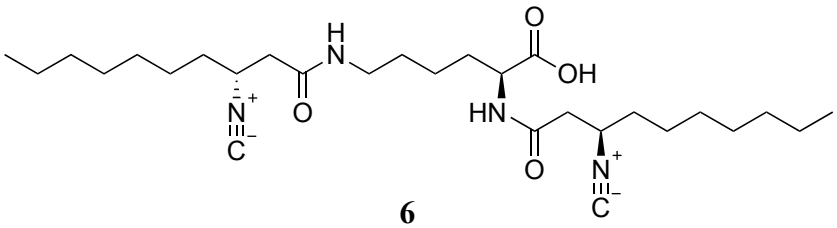


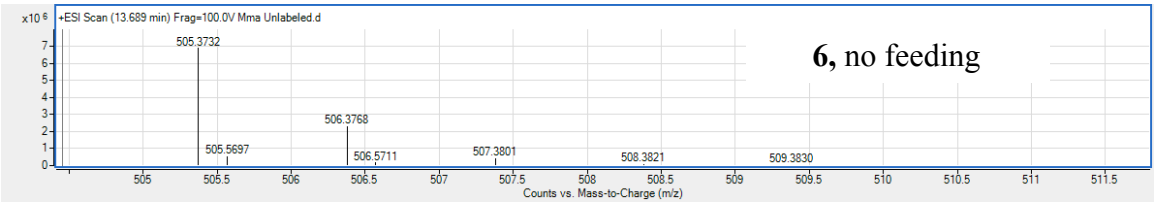


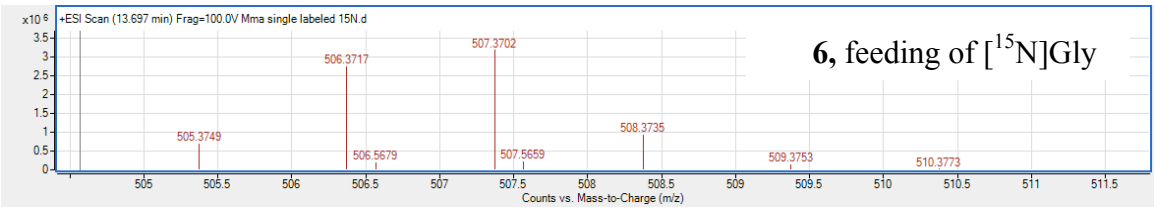

## G. HRMS analysis of 6 and 7.

*m/z* $[M+H]^+$
Obs. 505.3748
Calc. 505.3748
Error: 0 ppm

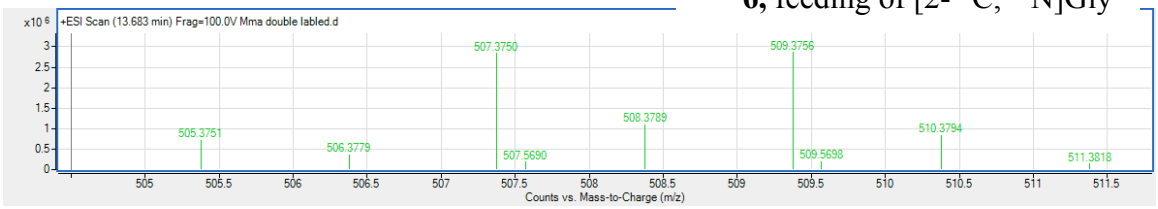


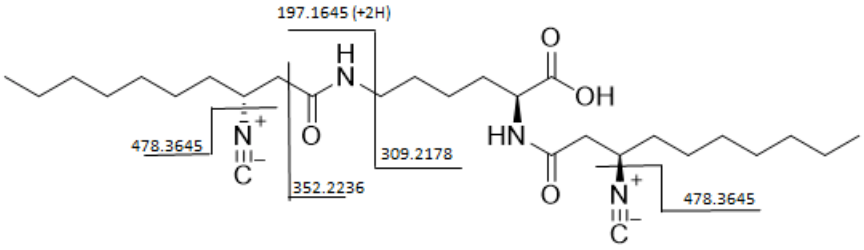
**6**

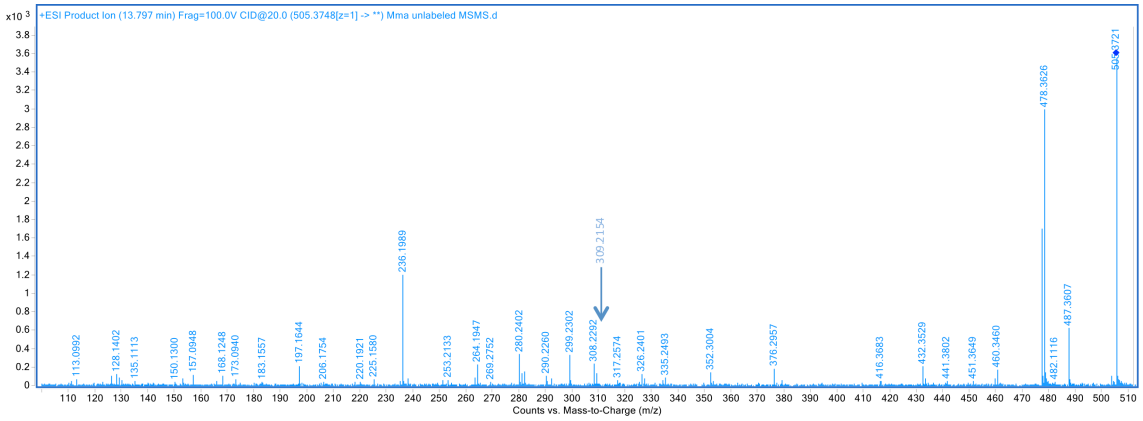


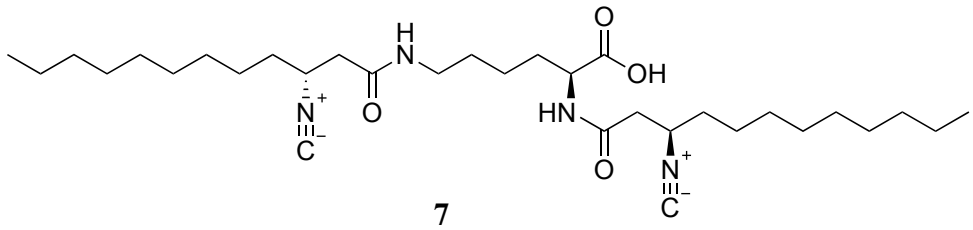
**6,** feeding of [2-$^{13}$C, $^{15}$N]Gly



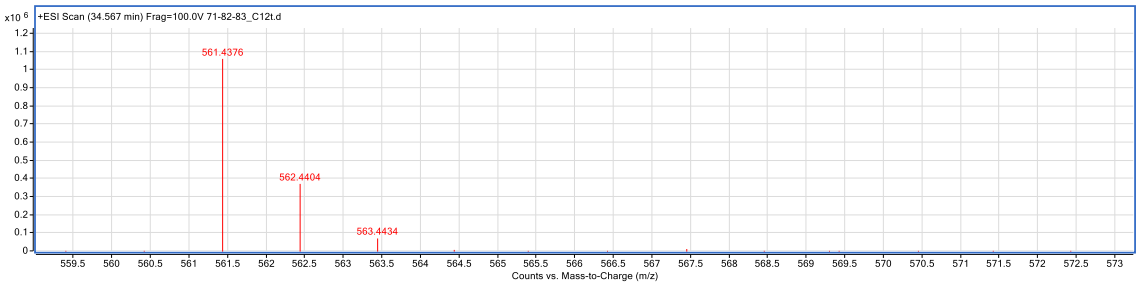



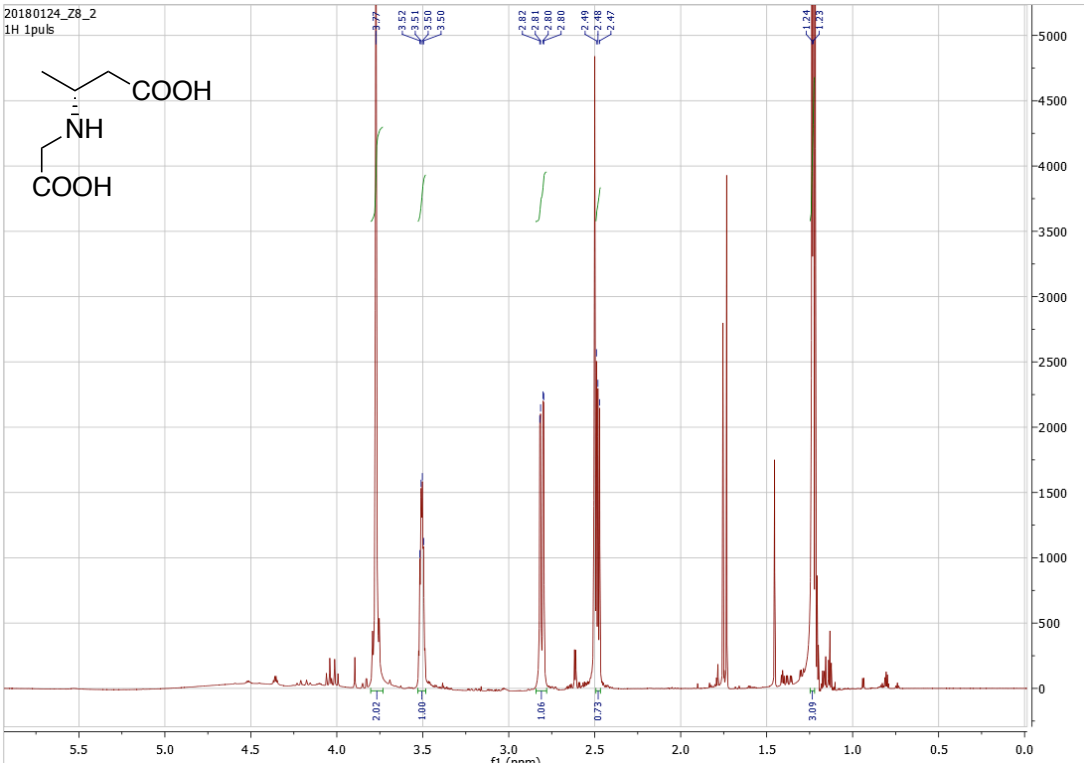
*m/z* $[M+H]^+$
Obs. 561.4376
Calc. 561.4374
Error: 0.4 ppm



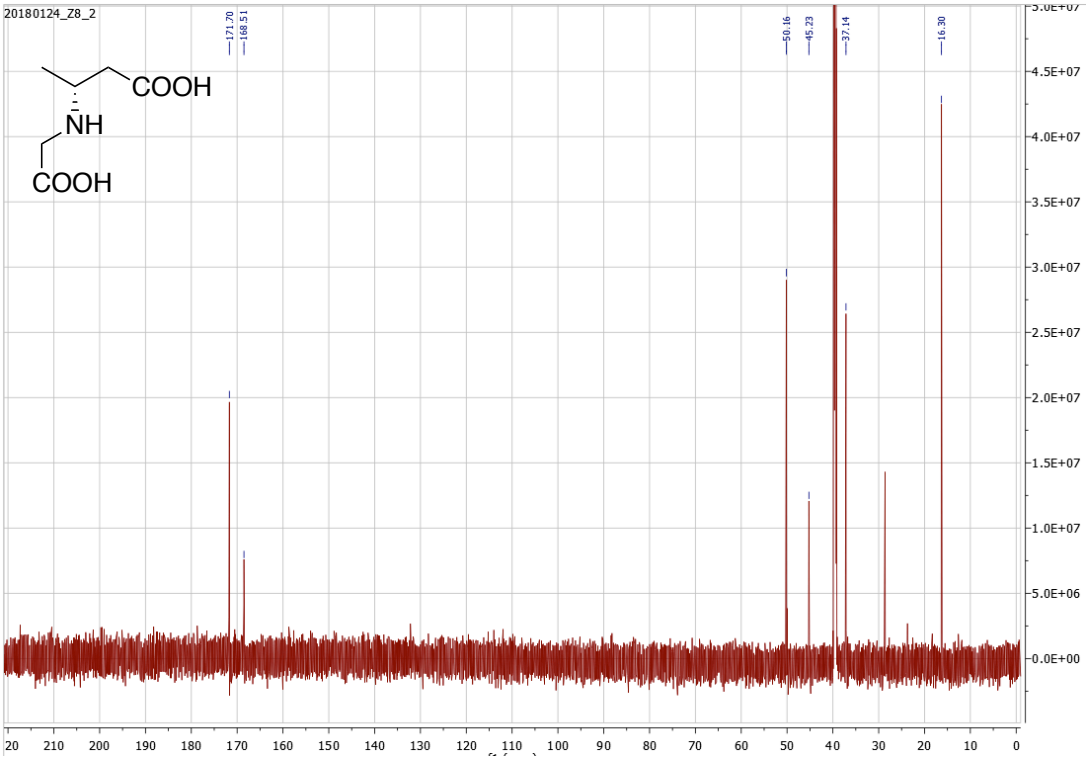
**7**




## H. $^{1}$H and $^{13}$C NMR spectra of 11.



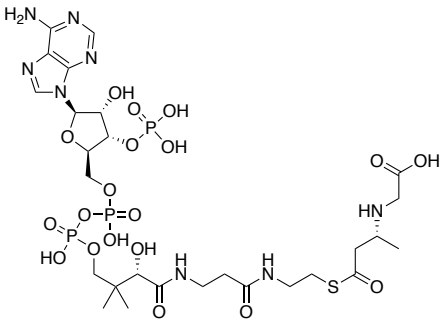

$^{1}$H NMR spectrum (DMSO-$d_6$, 900 MHz) of **11**.

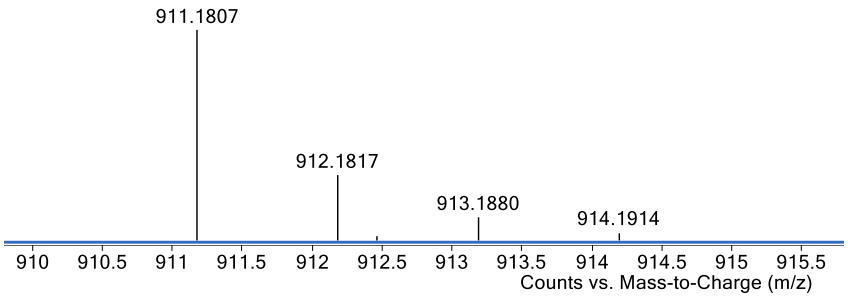



$^{13}$C NMR spectrum (DMSO-$d_6$, 225 MHz) of **11**.

# I. LC-HRMS analysis of 19.

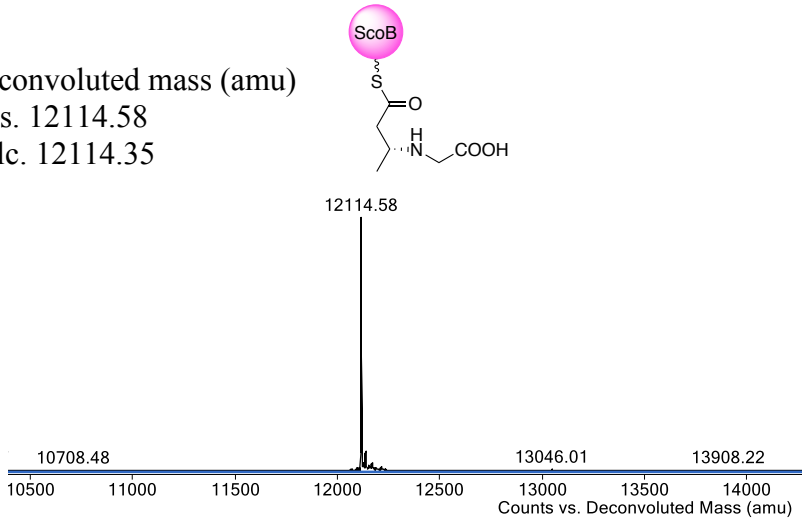
CABA-CoA
m/z $[M+H]^+$
Obs. 911.1807
Calc. 911.1807
Error: 0 ppm





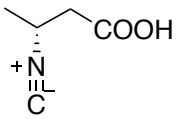
## J. LC-HRMS analysis of 8.

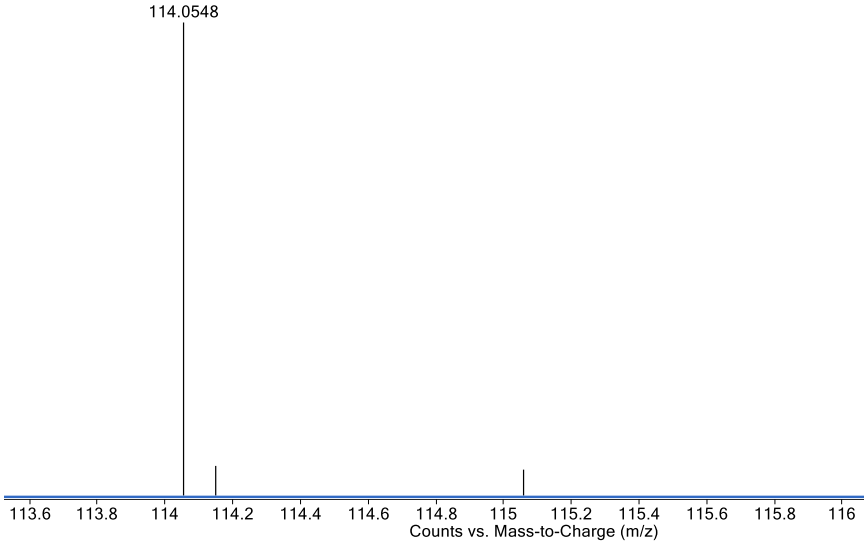
Deconvoluted mass (amu)
Obs. 12114.58
Calc. 12114.35



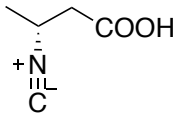
**K. HILIC-HRMS analysis of 12. Top spectrum is synthetic standard; bottom two spectra are from ScoE *in vitro* reactions. Bottom spectrum shows *in vitro* incorporation of labeled $^{13}C$.**

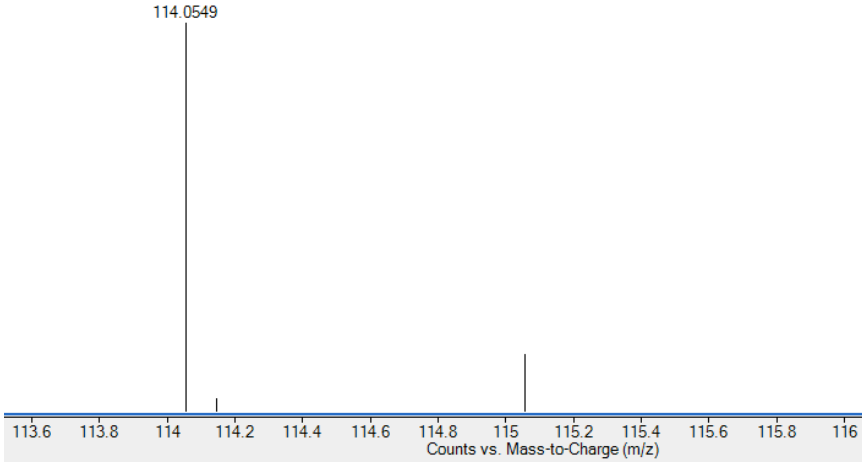
**12** synthetic standard
m/z $[M+H]^+$
Obs. 114.0548
Calc. 114.0550
Error: 1.75 ppm



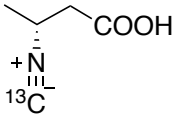



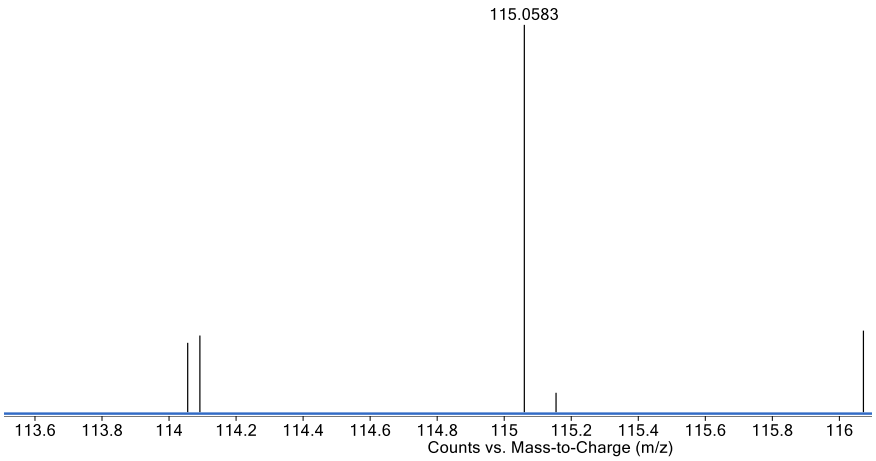
ScoE + **11** *in vitro* reaction
m/z $[M+H]^+$
Obs. 114.0549
Calc. 114.0550
Error: 0.88 ppm



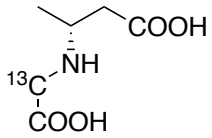

ScoE + **22** *in vitro* reaction
m/z $[M+H]^+$
Obs. 115.0583
Calc. 115.0583
Error: 0 ppm

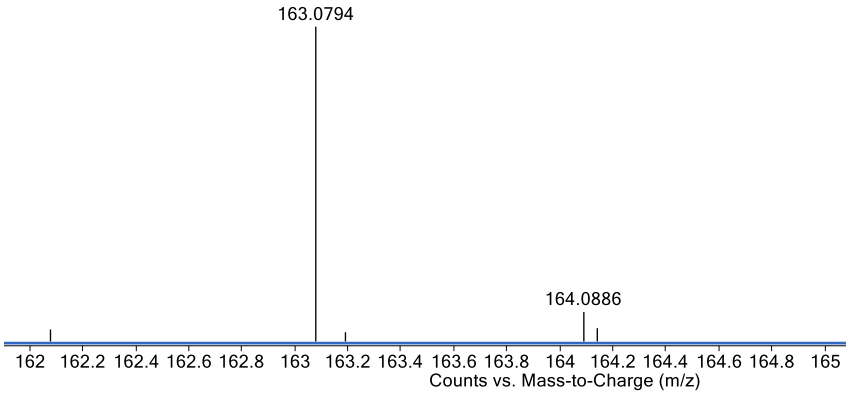


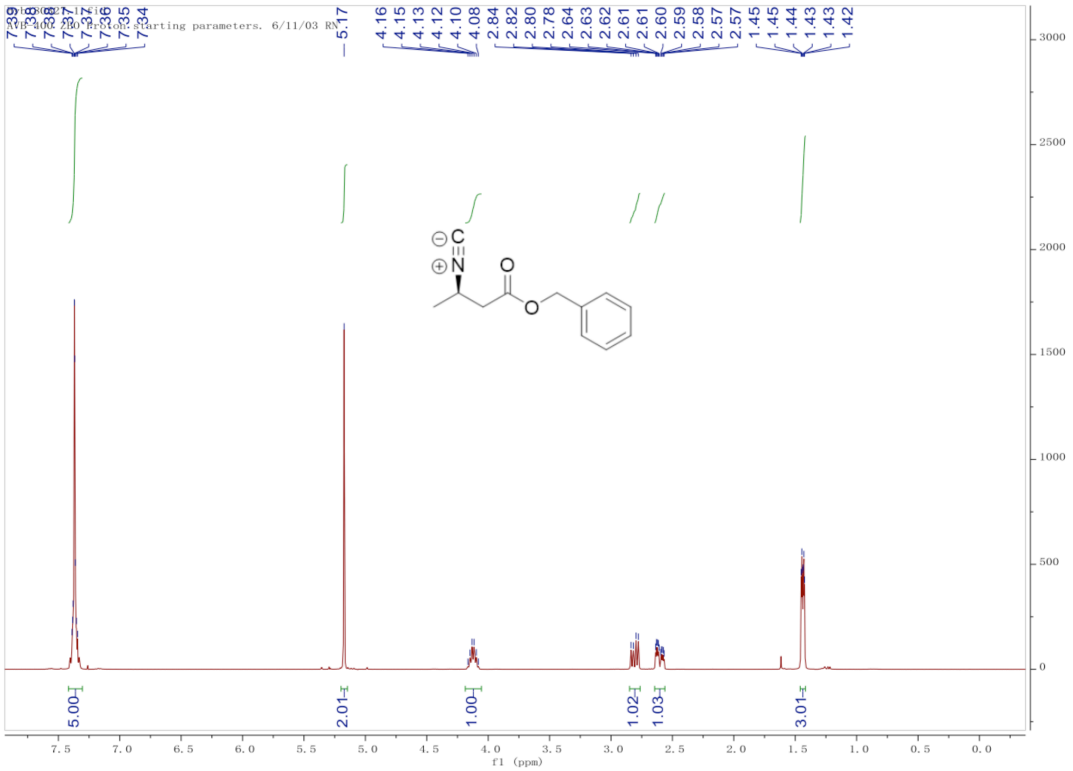
## L. HILIC-HRMS analysis of 22.

m/z $[M+H]^+$
Obs. 163.0794
Calc. 163.0794
Error: 0 ppm

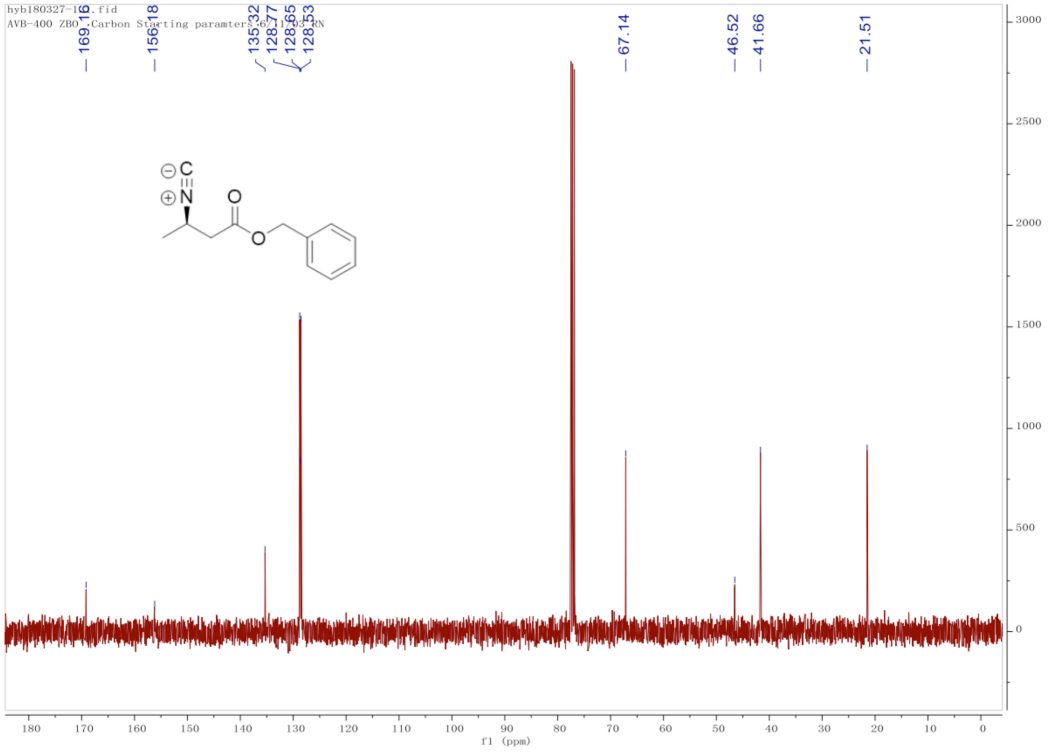



## M. $^1$H and $^{13}$C NMR spectra of 24.




$^1$H NMR spectrum ($CDCl_3$, 400 MHz) of **24**.




$^{13}C$ NMR spectrum ($CDCl_3$, 100 MHz) of **24**.