UNIVERSITY OF CALIFORNIA

SANTA CRUZ

STUDYING ERYTHRAZOLE A-C BIOSYNTHESIS IN PURSUIT OF DISCOVERY AND CHARACTERIZATION OF NOVEL BIOCATALYSIS IN UNUSUALLY ASSEMBLED NATURAL PRODUCTS

A dissertation submitted in partial satisfaction of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

CHEMISTRY AND BIOCHEMISTRY

by

Rebecca S. Pelofsky

December 2023

The dissertation of Rebecca Pelofsky is approved:

_____
John B. MacMillan, Advisor

_____
Professor Laura M. Sanchez

_____
Professor Shaun McKinnie

_____
Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of contents

List of Figures

Chapter one

Chapter two

Chapter three

Chapter four

Abstract

Studying erythrazole biosynthesis in pursuit of discovery and characterization of novel biocatalysis in unusually assembled natural products

By Rebecca S. Pelofsky

Natural products are structurally optimized by evolution and are often complex and have a wide range of bioactivity. The organisms which produce these compounds use enzymatic machinery which can perform reactions beyond the scope of what is currently possible by purely synthetic routes. Studying these organisms and their natural products can inform scientists of novel biosynthetic mechanisms and potentially novel biocatalysts. The erythrazoles are produced by SNB-035, a Gram negative alphaproteobacterial species, and harbor unique biosynthetic assembly. These structural features include being derived from an unusual hybrid of the methylerythritol phosphate pathway and shikimate pathway, in addition to containing two amino acids. Additionally, erythrazole A and B contain a benzothiazole moiety and show a modified terpene. Their terpene chain ends in a carboxylic acid and erythrazole B has an unusual chain length of 22 carbons. How these molecules arise is of great interest to us as they have negligible biosynthetic precedent in literature. Additionally, these were some of the first metabolites ever discovered from this

genus of bacteria and some of the very few known naturally occurring benzothiazoles.

We used isotopic labeling to label the amino acids of the erythrazoles which determined that the benzothiazole of erythrazoles A-B is derived from a β-carbon cleavage of cysteine, showcasing the first known example of this mechanism. Stable isotope labeling studies also led us to discover that interconversion does not occur between erythrazoles A and B. This indicates that they arise separately and their terminal carboxylic acid is formed through oxidative cleavage rather than an acetate extension of erythrazole A to form erythrazole B, which we originally hypothesized was the source of the unusual C22 terpene of erythrazole B. We also used various methods to propose four putative biosynthetic gene clusters, which were investigated through transcriptional repression using CRISPRi. Though no gene cluster was identified, these experiments provided further insight into erythrazole biosynthesis and showed evidence for the relationship between the erythrazoles and ubiquinone. The work described in this dissertation shines light not only on the erythrazoles but on naturally occurring benzothiazoles.

Dedication

I dedicate this thesis to my parents; Mark & Kari and grandparents;

Brenda & Joel.

Acknowledgments

First and foremost, I want to thank John MacMillan, my PhD supervisor. You took me into the lab even though I came in with no natural products or chemistry background. Thank you for supporting me through a difficult project and allowing me to work on biosynthesis even though it is slightly outside of the lab's wheelhouse. These past 5 years have truly been transformative and I will always be grateful through the challenges and successes of my time in the lab. I'd also like to thank my committee members, Shaun McKinnie and Laura Sanchez. Thank you, Professor Shaun McKinnie for all of the guidance you gave me and your willingness to sit down and talk about my research. I've learned so much from you whether it was in our meetings, while taking and TAing your class, and from the journal clubs you welcomed me to join. Thank you, Professor Laura Sanchez for stepping in to chair my committee these last few years and for all of your mass spectrometry help and advice. I always appreciated you making time to meet with me and for sending me helpful papers regarding the questions I was having.

To mom, dad and Hannah: thank you for always supporting me and pushing me to start new chapters even when I don't want to. You've always seen my potential and have pushed me beyond what I think I can do. You never held me back from anything, always supporting me and letting me know that I will always have your support and love unconditionally. Even when I was a difficult teenager and

made everyone's lives miserable you never gave up on me or made me feel like I couldn't be myself or accomplish whatever I set my mind to. I genuinely don't think I could have asked for a better family and I will always feel lucky to have you as my family.

To Matt – I couldn't have imagined these past 5 years without you. You have been my rock through it all and I'm so grateful to have had you by my side. Coming home after a hard day to you makes everything seem like it's going to be okay. Thank you for always loving me and being my voice of reason. Goose – I know you're going to read and ponder every word of this thesis. You turned me into a crazy dog mom and I don't regret one minute of it because you taught me to love dogs and I know in my heart that you are the best dog in the whole world. Taking you out and seeing you happy when I get home is the best part of my day.

To Neve and Logan, I still can't believe how lucky I was to have answered that facebook ad when you were searching for a roommate. Living together my first 3 years of grad school and going through the pandemic together was truly such an incredible experience. Despite the terrors of the world around us we got to take a break from life, try some new hobbies, and grow our friendships even closer. I even learned to cook more than eggs. You quickly became like family to me and I know we will be in each others' lives forever.

To the past MacMillan lab grad students - you all really set the bar high. You taught me how to be a good lab mate and you taught me so much about natural

products chemistry. Duy and Victor, you guys really brought a special energy into the lab and it hasn't been the same without you. Spending Christmas breaks commiserating together, watching romcoms, and drinking fancy liquor are memories I'll always cherish. Whether it was a pastry left on my desk from Duy or an impromptu drink with Victor, you guys always made the lab feel a little warmer. Jocelyn you always set the best example. I knew at whatever hour I came into lab I'd see you on the HPLC and you always motivated me to work harder. I had your thesis up on my computer pretty much the entire time I wrote mine. It's even open right now to see how you formatted your acknowledgements. Thank you all for being role models as friends and also scientists.

To the current lab mates; Leah, Kaitlyn, Riley, Sahar, Akshar, Anam and Aswad: thanks for always putting up with my chattiness and occasional moody days. Having coworkers that are also friends makes going to work every day so much easier and I'm really going to miss working with you all and seeing you every day. Special thanks to my best friend and lab mate, Leah. It's truly how incredible to have had a best friend who I work side by side with every day. It's hard to believe that we won't be working together now and I know I'm going to be somewhat lost with you. We know every detail of each other's lives whether its what's happening at work or at home. We've been through so much together and you've supported me through all the ups and downs. I love that we can talk about literally everything whether its something ridiculous, an embarrassing story, or the challenges we're having with our

science. You are truly a remarkable human, friend and scientist and I know we will be life-long friends. I can't wait to see what you accomplish next. Kaitlyn – thank you for being my "bio side" buddy and always being willing to commiserate or celebrate. We both are hard-headed people and even when we disagreed, I always knew we would move on and laugh about it in the future. To Sahar – thank you for being such a genuine and a good-hearted person. You have been through so much but I know you are resilient and will accomplish whatever you set your mind to. Aswad – thank you for always sharing good food, recipes, and coffee. I loved going camping with your family in the redwoods. Thank you to Riley for all of your help with NMR. I know with your confidence and motivation you will do great things in the lab and beyond. Anam "nams" - I love watching trash tv with you and Leah and venting. Thank you for helping me with my biology and coming with me to set up my qPCR whether I needed your help or just wanted company. And finally, Akshar, thank you so much for all of your computational help. You always made time to help me and I could not have done a lot of the work here without you.

Thank you to the entire McKinnie lab. I loved having you all next door and you often felt like the extended family of the lab. You all have helped me with research in one way or another and I always appreciated you taking the time to help someone outside of the lab in addition to being able to just come by to hang out or vent with you all.

To the Thera-squad, you all were my first real lab family. Thank you for introducing me to the world of industry while also sheltering me and keeping me optimistic. I feel so lucky to have ended up working with you all and you set up me up for success going forward. Michelle, thank you for always being such an amazing friend. I know you will always be there for me and I know I can tell you anything. Thank you for always answering my questions whether it was about RNA or career advice. Our girls' days out on the weekends and vacations really kept me sane during grad school and I always look forward to them. I loved evolving from our early 20's on Polk street to our homebody dog mom era together.

List of Abbreviations

°C – degrees Celsius

A1BFe+C – high nutrient, seawater based medium

ACN – acetonitrile

AhR – aryl hydrocarbon receptor

AT – acyl transferase

ATP – adenosine triphosphate

attTn7 – attachment site for the Tn7 transposon

AUC – area under the curve

BGC – biosynthetic gene cluster

CAM – chloramphenicol

CCD – carotenoid cleavage dioxygenase

CM – chorismate mutase

CoAB – phosphopantothenate-cysteine ligase

CoQ10 – coenzyme-Q10 (ubiquinone with 10 prenyl groups, or C50)

CRISPR – clustered regularly interspaced short palindromic repeats

CRISPRi – CRISPR interference

CYP450 – cytochrome P450

Da – daltons

DBU – 1,5-diazabicyclo[5.4.0]undecane

DDMQ – C1-demethyl-C6-demethoxy-Q

DMAPP – dimethylallyl pyrophosphate

DSS – dextran sodium sulfate

DXP – deoxy-D-xylulose 5-phosphate

EA – erythrazole A

EB – erythrazole B

EC – erythrazole C

EDC – 1-ethyl-3-(3-diamethylaminopropyl)carbodiimide

FPPS – farnesyl diphosphate synthase

GGPPS – geranylgeranyl pyrophosphate synthase

GlyLi – glycine ligase

GNPS – global natural product social molecular networking

GST – glutathione S-transferase

HEPES – 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HMM – hidden markov model

HMP1 – The NIH Human Mictobiome Project

HPLC – high performance liquid chromatography

I3P – indole-3-pyruvic acid

IBD – inflammatory bowel disorder

ICN – indole-3-carbonyl nitrile

IDS – isoprenoid diphosphate synthase

iHMP – Integrative Human Microbiome Project

IPP – isopentenyl pyrophosphate

IPTG – isopropyl β-D-1-thiogalactopyranoside

ITE – 1’H-indole-3’-carbonyl)- thiazole-4-carboxylic acid methyl ester

KEGG – Kyoto Encyclopedia of Genes and Genomes

LC-HRMS – liquid chromatography high resolution liquid chromatography mass spectrometry

LC-MS – liquid chromatography mass spectrometry

m/z – mass-to-charge ratio

MeOD – deuterated methanol

MeOH – methanol

MEP – methylerythritol phosphate

MEV – mevalonate

mRFP – monomeric red fluorescent protein

MS/MS – tandem mass spectrometry

MSA – multiple sequence alignment

NADH – nicotinamide adenine dinucleotide + hydrogen

NCBI – National Center for Biotechnology Information

NHase – nitrile hydratase

Ni-NTA – nickel-nitrilotriacetic acid

NMR – nuclear magnetic resonance

NRP – non-ribosomal polypeptides

NRPS – non-ribosomal polypeptide synthase

OD – optical density

OMT – *O*-methyltransferase

PCR – polymerase chain reaction

PK – polyketide

PKS – polyketide synthetase

PLP – pyridoxal 5'-phosphate

PQQ – pyrroloquinoline quinone

RiPP – ribosomally synthesized and post-translationally modified peptides

RNA-Seq – RNA sequencing

RPM – revolutions per minute

RT-PCR – reverse transcriptase polymerase chain reaction

SacB – gene which converts sucrose to levan, causing sucrose toxicity to host organism

SAM – S-adenosyl methionine

SDS-Page – sodium dodecyl-sulfate polyacrylamide gel electrophoresis

sfGFP – superfolder green fluorescent protein

sgRNA – single guide RNA

SIM – single ion monitoring

TCDD – 2,3,7,8- [3H]tetrachlorodibenzo-p-dioxin

TCG – a minimal medium composed of tryptone, casitone, and glucose

ThiG – thiazole synthase protein

ThsX – thiazole synthase knockout

TNBS – 2,4,6-trinitrobenzenesulfonic acid

***Trans*-AT-PKS** – *trans*-acyl transferase polyketide synthases

UbiG – ubiquinone biosynthesis *O*-methyltransferase

CHAPTER ONE

INTRODUCTION TO THE ERYTHRAZOLES, THEIR PRODUCER, AND WHY WE STUDY THEM

1.1 Introduction

Natural products, or secondary metabolites, comprise ~65% of all small molecule approved pharmaceuticals, which have a wide variety of bioactivity[1]. In addition to their activity, natural products can be structurally complex and diverse, leading them to be of interest in both academia and industry. Compared to synthetic compounds, natural products often have increased complexity, with complex ring systems, stereocenters, and densities of functional groups[2–5].

Erythrazoles A-C were isolated by the MacMillan lab from a marine, alphaproteobacterial species, SNB-035. While alphaproteobacterial species make up a large portion of the marine biosphere[6], they are relatively understudied as producers of marine natural products. The erythrazoles demonstrate unusual structural characteristics, which intrigued us to studying the mechanism of their biosynthesis. The discovery of novel biosynthetic enzymes can enable new chemical syntheses and the discovery of similar kinds of metabolites, which is what we hope to accomplish by studying the erythrazoles. This chapter outlines the significance of biocatalysis and biosynthetic gene clusters, what is known about bacterial species similar to SNB-035, and the intriguing characteristics of erythrazole A and B.

1.2 Biocatalysis in natural products, harnessing enzymes for new chemical transformations

1.2.1 Biocatalysis

The enzymes which produce secondary metabolites have been harnessed to perform useful and interesting reactions *in vitro*, a field known as biocatalysis. These enzymes have undergone many rounds of evolutionary selection, optimizing their metabolites for their biological target[1]. Biocatalysis can be a preferable option to standard synthesis as it is often more safe, environmentally friendly, and capable of chemical transformations that are not accessible with total synthesis routes[3,7,8]. In addition to these benefits, a major limitation of producing natural products for pharmaceutical and industrial relevance is that they often have low production levels in their natural producers. Therefore, harnessing biosynthetic enzymes for chemical synthesis can be a more efficient way to produce natural products in large scale.

Biocatalysis has been used for many kinds of chemical reactions including redox, substitution, addition, and elimination reactions[3]. These various enzymes are also capable of forming carbon-oxygen (C-O), carbon-nitrogen (C-N), carbon-carbon (C-C), and carbon hydrogen (C-H) bonds[9]. One of the first examples of biocatalysis was using nitrile hydratases (NHase) for hydration in chemical synthesis. This reaction has been applied in large scale

for the production of polyacrylamide, which is used to produce hundreds of thousands of tons per year of acrylamide[9]. Nitrile hydratase-catalyzed hydration is also used for the industrial production of vitamin B_3[10]. Some other applications of hydrolases includes aspartame[11], β-lactam antibiotics[12], and glucose production[13]. Another early example of the success of biocatalysis is the discovery of an enzyme from *Rhizopus nigricans* to oxidize preogesterone in order to make cortisone, which replaced a 31-step chemical synthesis[14].

Since its discovery, biocatalysis has advanced to more complex reactions. C-H functionalization has been one of the most important types of biocatalytic reactions as they are efficient, less accessible through purely synthetic methods, and can produce diverse and complex molecular scaffolds. These reactions can also be late-stage, site- and stereoselective. C-H functionalization has been used in antibiotic biosynthesis, production of steroid-based drugs, and diversification of shellfish toxins[15,16]. These reactions can also be accomplished on important natural product scaffolds containing amino acids and terpenes[17].

New C-C bond forming reactions have also been discovered and advanced throughout the years. One example of this is the discovery of the enzyme, CylK, which performs an intermolecular Friedel-Crafts alkylation. CylK has been found to be highly promiscuous, allowing it to have a widespread application for other molecules' biocatalysis[18]. Several C-C bond forming enzymes have been discovered from natural sources, including with carbene

transfers, a common synthetic reaction that was previously not known to occur in nature[19].

The number of known biocatalysts and reactions capable by biocatalysts is always growing. The field in its current state has progressed to directed evolution and engineering of these enzymes to optimize their catalytic abilities[4]. A hallmark example of the successful application of biocatalysis is that for islatavir, an HIV drug whose synthesis has been accomplished by five engineered enzymes with significantly higher yield than previous synthetic methods[20]. The capabilities and significance of biocatalytic enzymes on both an academic and industrial scale has been demonstrated for many years, exhibiting the importance of studying naturally occurring molecules and the enzymes which produce them.

1.2.2 Canonical biosynthetic gene clusters enable genome mining for natural product discovery

Biocatalysis and biosynthesis research was greatly accelerated by the advances and increase in availability of genome sequencing. The sequencing of known natural product producers, *Streptomyces coelicolor* and *Stretomyces avermitilis* identified 20-25 biosynthetic gene clusters (BGCs), when only 3-4 natural products were currently known to be produced[21–23]. However, genetic screening for biocatalysts is often limited by the identification of typically seen biosynthetic genes. The best studied examples of natural products are non-

ribosomal peptides (NRPs) and polyketides (PKs). These molecules are biosynthesized by the megasynthases, non-ribosomal polypeptide synthetases (NRPS) and polyketide syntases (PKS). Both NRPS and PKS megasynthases are composed of a series of functional modules, and act in a sequential manner to lengthen the molecule at each module. These modules are comprised of domains, one of which is a carrier protein domain bears a thiol-terminated phosphopantetheine moiety to covalently bind the polyketide or polypeptide before passing it to the next module. The other domains are specific to NRPSs or PKSs, each of which performs a specified function, ultimately causing chain elongation of the molecule and determining the structure of the metabolite which is being produced[24,25]. Figure 1.1 depicts the similar modular fashion of a PKS and NRPS with their differing domains. Similar interactions occur in both PKSs and NRPSs where the megasynthase proteins interact to pass the molecular scaffold from one protein to the next which are called docking domains in a PKS and communication (COM) domains in an NRPS. This is also depicted in figure 1.1. Figure 1.2 shows an example this assembly-line biosynthesis occurs in the case of the PKS, erythromycin. In this example, malonyl-coA units are modified by the domains shown and passed on from module to module in order to yield the final structure of erythromycin.

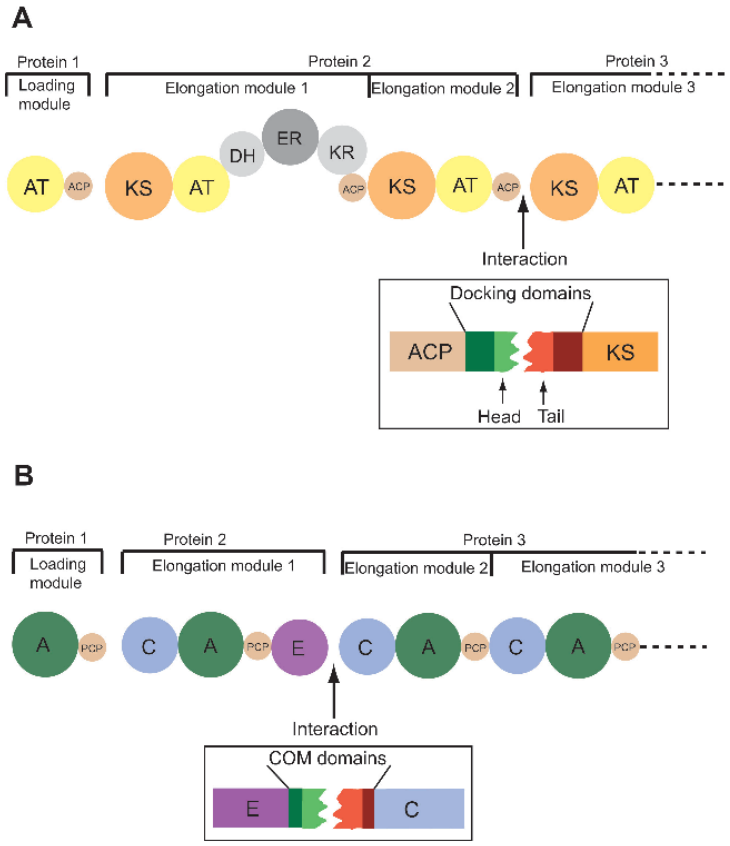


Figure 1.1 PKS and NRPS modules with similar structure but comprised of different domains. A. An example of PKS domains in a typical PKS megasynthase. B. An example of NRPS domains within a typical NRPS megasynthase. Abbreviations: KS, ketosynthase; AT: acyltransferase; ACP: acyl carrier protein; DH: dehydratase; ER: enoyl reductase; KR: ketoreductase; C: condensation; A, adenylation domain; PCP: peptidyl carrier protein; E: epimerization domain. Figure from Jenke-Kodama, H. and Dittman, E. Bioinformatic perspectives on NRPS/PKS megasynthases: Advances and challenges. *Natural Product Reports* **2009**, 26 (7), 874-883. https://doi.org/10.1039/B810283J

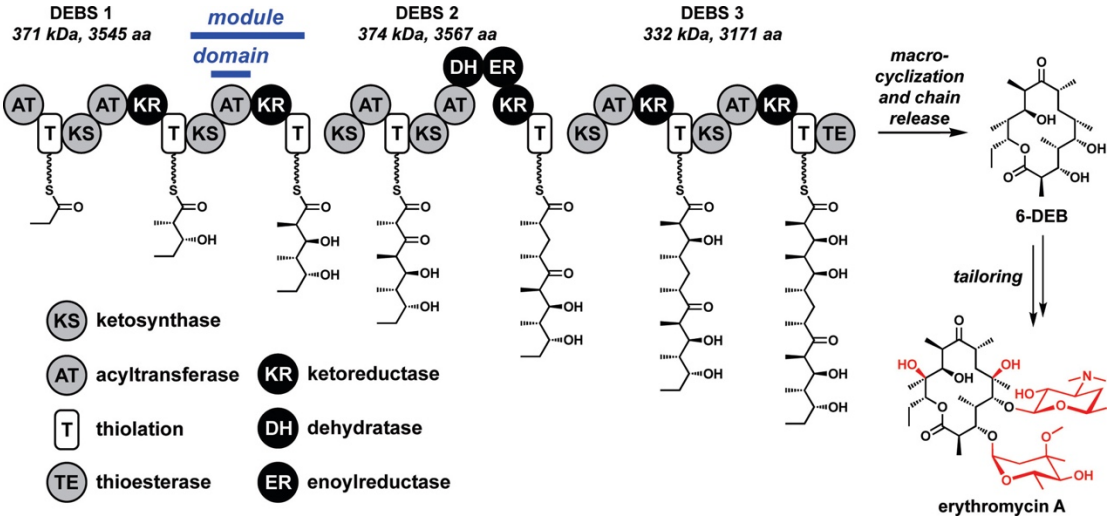


Figure 1.2 Example of assembly-line biosynthesis of secondary metabolites in the PKS erythromycin. This PKS was the first discovery of assembly-line biosynthesis and changed the way scientists understand assembly line biosynthesis. Reprinted with permission from Walsh, C. and Fischbach, M. Natural Products Version 2.0: Connecting Genes to Molecules. *Journal of the American Chemical Society* **2010**, 132(8), 2469-2493. https://doi.org/10.1021/ja909118a. Copyright {2010} American Chemical Society.

PKS and NRPS domains are highly conserved, which enables large scale screening for the discovery of new biosynthetic gene clusters (BGCs)[26]. To demonstrate the widespread presence of these megasynthases, Wang et al. searched the sequences of 2,699 genomes representing 4 different phyla of bacteria and 1 from fungi. This analysis identified 3,399 gene clusters that were either NRPS, PKS, or NRPS/PKS hybrid[27]. Many of the products of these BGCs were unknown, highlighting the value of genome mining techniques for natural product discovery. PKs, NRPs and PK-NRP hybrids produce widely used pharmaceutics including ikarugamycin (antibiotic, anti-leukemia, and anti-inflammatory), lovastatin (anti-cholesterolemic), erythromycin (antibiotic), and cyclosporin (antibiotic)[3,25,28]. While many tools have been developed to identify

BGCs *in silico*, they primarily identify NRPS, PKS, ribosomally synthesized and post-translationally modified peptides (RiPPs), and terpenes[26]. However, there is still a gap in knowledge regarding genome mining for BGCs outside of these groups.

1.2.3 Noncanonical biosynthesis in natural products

While gene clustering is generally accepted as the standard for bacterial secondary metabolite biosynthetic genes, several metabolites have been discovered for which this is not the case. For these noncanonical biosynthetic gene clusters, standard genome mining strategies cannot be used. However, due to their increase in discovery, studying these kinds of biosynthesis are likely to enable the discovery of more noncanonical BGCs, which are able to produce metabolites with novel structure and bioactivity. This will also enable the expansion of current chemical and enzymatic understanding of biosynthetic machineries[29,30].

A prime example of novel biosynthesis discovery leading to further discoveries is that of *trans*-acyl transferase polyketide synthases (*trans*-AT PKSs). *Trans*-AT-PKSs differ from a typical PKS because they lack acyltransferase (AT) domains within the megasynthase. The AT activity occurs instead at each elongation step *in trans* by one or a few free-standing proteins elsewhere in the BGC or sometimes outside of the BGC altogether. *Trans*-AT-

PKSs also have greater architectural diversity, which includes modules with unique or unusually ordered domains, non-elongating modules, action by domains across different modules, modules split into more than one protein, and can have other *in trans* activity in addition to the *in trans* ATs[31,32]. *Trans*-AT-PKSs are often not found in the well-studied, actinomycetes, which is partially why they were overlooked for so long. They have been seen more commonly in less well-studied bacteria from unusual taxa, bacteria that are challenging to cultivate in a laboratory setting, mutualists and pathogens[30].

Due to their prevalence in less studied bacteria and often non-colinear nature, these megasynthases were not discovered until 2008 whereas the NRPS/PKS machinery was discovered in the 1990's[31]. Though genome mining tools were able to detect *trans*-AT PKSs early on, their biosynthetic abilities were poorly understood. Further research has allowed a deeper understanding of *trans*-AT PKSs, and genome mining tools have been developed specifically for them, such as transATor, which predicts the *trans*-AT PKS-derived polyketide core structure[32]. This tool can be used for dereplication, isolation, and structure elucidation. This example demonstrates how studying noncanonical BGCs can enable future discoveries of new classes of compounds and biocatalysts.

Other examples of biosynthetic pathway discovery enabling new genome mining strategies includes the discovery of the BGC for the acetylenic meroterpenoid biscognienyne B, which enabled a new handle to identify other

natural products with acetylenic prenyl chains[33]. Similarly, a domain which incorporates sulfur in leinamycin biosynthesis has been used to discover new analogs to the anti-cancer metabolite in various producers[34]. The discovery of Diels-Alderases, enzymes which catalyze a naturally occurring Diels-Alder reaction, led to the discovery that these enzymes were more broadly found in biosynthetic pathways and enabled the discovery of previously unknown biosynthetic gene clusters and natural products[35].

Noncanonical natural product biosynthesis has also been seen in secondary metabolites which have been derived from primary metabolites, which differ from the typical amino acid or malonyl-coA building blocks of NRPSs or PKSs. These building blocks are often thought of as enzyme cofactors, but there is an increasing amount of known example where they are actually used as precursors of secondary metabolite biosynthesis[29]. One example of this is seen in the biosynthesis of Altemicidin, SB-203207 and SB-203208 from nicotinamide adenine dinucleotide (NAD). NAD is an important metabolite used by all living organisms as an electron acceptor and carrier. Meanwhile, Altemicidin has been identified as secondary metabolite which has anti-cancer activity and SB-203207 and SB-203208 are secondary metabolites with Ile-tRNA synthetase inhibitor activity. Determining the biosynthesis of these compounds was challenging, as they do not have any NRPS, PKS, RiPP or terpene moieties. While clustered biosynthetic genes were ultimately identified through isotope labeling experiments and resistance gene-guided

genome mining, the biosynthesis of these compounds demonstrated a novel kind of BGC which was not identifiable through standard genome mining techniques[36]. This discovery also set the precedent for NAD involvement in secondary metabolism.

Some other examples of primary metabolites being precursors of secondary metabolites are roseoflavin and tolyphorin. Roseoflavin is an antibiotic which is derived from riboflavin, or vitamin B_2. Riboflavin is converted to flavin mononucleotide (FMN) by riboflavin kinase. *Streptomyces davawensis* and *Streptomyces cinnabarinus* have then been shown to use two non-clustered enzymes to facilitate the transformation of FMN to roseoflavin[37,38]. Similarly, tolyphorin A, a tetrapyrrole macrocycle produced by several cyanobacterial species is derived from the heme biosynthetic pathway. A BGC has been identified for these molecules which includes many of the heme biosynthetic genes. It has been hypothesized that the heme intermediate, protoporphyrinogen IX is converted by several additional enzymes in the BGC to produce tolyphorin A[39].

Similarly, S-adenosyl methionine (SAM), is a commonly used enzyme cofactor which has also been seen to act as a substrate for natural product biosynthesis[40]. Some examples of this are shown in Figure 1.3, with varying degrees to which they incorporate SAM. While these metabolites have a variety of different functions (hormone precursor, quorum sensing, cell density and

biofilm regulation, antibiotics), they mostly do not involve a typical biosynthetic pathway that is easily detectable. Many of the enzymes discovered in these biosyntheses do not have homology to previously known enzymes, highlighting their novelty and the importance of these discoveries[29].

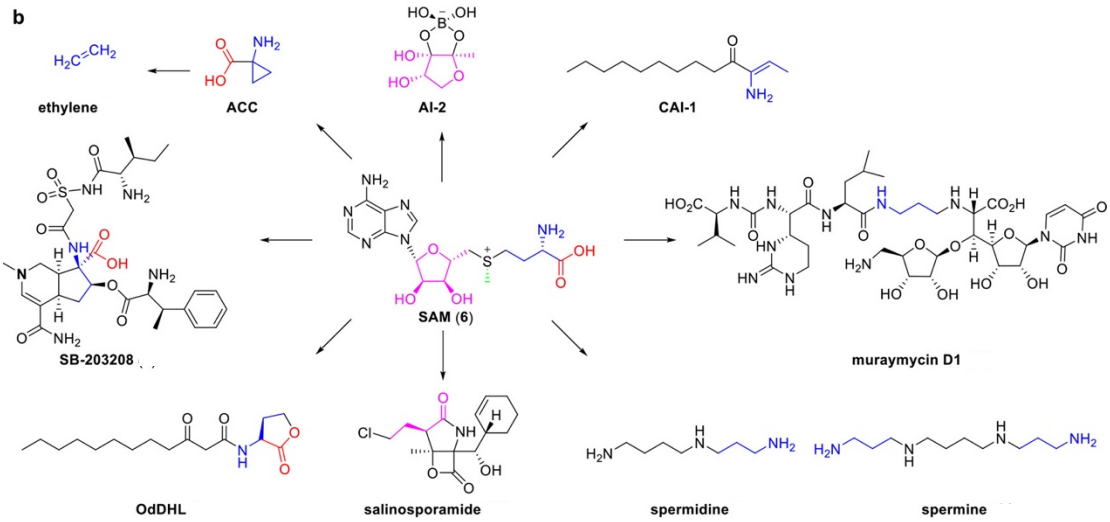


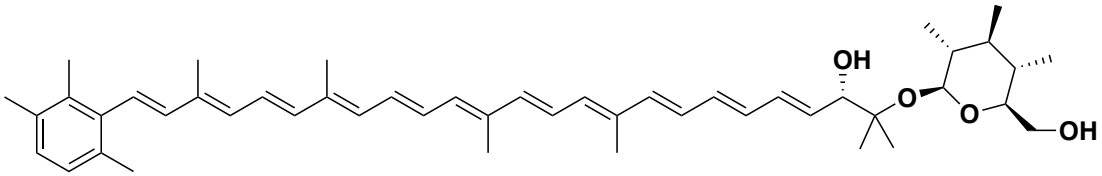
Figure 1.3 Natural products which use SAM as a precursor. Figure from Barra, Awakawa, and Abe. Noncanonical Functions of Enzyme Cofactors as Building Blocks in Natural Product Biosynthesis. *JACS Au* **2022**, 2(9), 1950-1963. https://doi.org/10.1021/jacsau.2c00391.

These instances of secondary metabolites being derived from primary metabolism showcase a novel method of natural product biosynthesis which is still poorly understood. Some of them closely resemble their primary metabolite precursor, while others are more divergent. Discovering these biosynthetic mechanisms is challenging due to their lack of precedent and sometimes lack of clustering of the biosynthetic genes. Due to these challenges, there is likely more prevalence of this type of biosynthesis than is currently known.

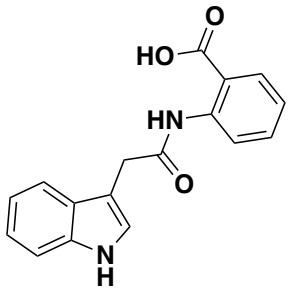
In addition to these examples, there have been other instances where biosynthetic genes were found to be un-clustered with no precedent in known natural product biosynthesis. One example of this is sioxanthin, a glycosylated carotenoid from the actinomycete, *Salinispora tropica* CNB-440. While carotenoid biosynthetic genes are typically clustered in bacterial genomes, sioxanthin biosynthesis occurs by two separate gene clusters and two independent genes[41]. Similar non-clustering of carotenoid biosynthetic genes has been seen in the Gram negative, *Gemmatimonas aurantiaca* strain T-27[42] and the extremophiles of the phylum Deinococcus-Thermus[43]. The co-occurrence of non-clustering carotenoid biosynthetic genes across these phyla indicate that this may be more prevalent than previously thought.

Other examples of non-clustering in biosynthetic genes are in the production of lindolin A and B and naringenin. Lindolin A and B have anti-plant pathogen activity and are produced by the fungi of the order, Kickxellales. These fungi produce lindolin A and B from indole-3-acetic acid with two, non-clustered enzymes[44]. Naringenin is a flavanone with anti-inflammatory, chemoprotective and antitumor activity. While it has been previously identified as a plant metabolite, it recently was discovered from *Streptomyces clavuligerus*. The biosynthesis identified in *S. clavuligerus* involves three biosynthetic genes, two of which are clustered together and one which is located 47.3kb away from those genes in the genome. It was proven that all

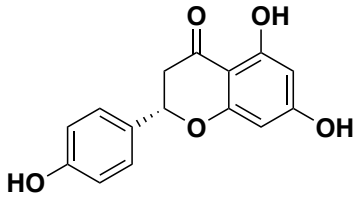
three genes are necessary for biosynthesis of naringenin through knockout studies[45].



sioxanthin



lindolin A



naringenin

Figure 1.4 Secondary metabolites produced by non-clustered biosynthetic genes.

1.3 Marine bacteria, SNB-035, and its natural products

The marine bacteria, SNB-035 was isolated from a sediment sample collected from Trinity Bay, Galveston, TX and isolated on seawater-based acidic gauze media. Initial 16S rRNA analysis classified SNB-035 as an *Erythrobacter* species. However, current 16S analysis shows the highest similarity to *Qipengyuania xiamensis* (99.45%). In 2020, the family Erythrobacter was re-classified and several species that were previously in the *Erythrobacter* genus were assigned to the *Qipengyuania* genus[46]. As this reclassification is a relatively new development, there has been less published about the species

in the *Qipengyuania* genus. Therefore, precedent for metabolite production was investigated in the both the *Erythrobacter* and *Qipengyuania* genus.

Erythrobacter species and *Qipengyuania* species are Gram-negative bacteria and belong to the family, *Erythrobacteraceae* of the order, *Sphingomonadales* of the class, *alphaproteobacterial*. They are marine aerobic anoxygenic phototrophs and while they require an organic carbon substrate for growth, they are also able to supplement their metabolic requirements with photosynthetically derived energy[6,47,48]. The Erythrobacter genus was first described in 1982 with the discovery of the species *Erythrobacter longus*[49] and several different species have been discovered since. *Erythrobacter* species have mainly been studied due to their phototrophic ability and their prolific production of carotenoids. While *Erythobacter* species and *Qipengyuania* species have physiological differences and are evolutionarily distinct, they share many of these same characteristics[50].

Most species in the *Erythobacter* and *Qipengyuania* genus' produce bacteriochlorophyll-α, a chlorophyll which enables photosynthesis by these species. However, the overall energetic contribution of photosynthesis to the organisms is relatively low[51]. The discovery of their phototrophic abilities changed the conception of the prominence of this trait in marine bacteria, as *Erythrobacter* species make up 1-15% of the total bacteria in the upper ocean[6]. *Erythrobacter litoralis* has been shown to use visible light as a general stress

response, which has been suggested to be the purpose of their photo-sensing abilities rather than for energy metabolism[52].

1.3.1 Previously studied metabolites from *Erythrobacter* and *Qipengyuania* species

The metabolites which have been mainly studied from *Erythrobacter* spcies are carotenoids. Upon the discovery of *E. longus*, 20 different carotenoids were identified as being produced. These include β-carotene, β-cryptoxanthin, zeaxanthin, caloxanthin, and nostoxanthin in addition to derivatives of these[53]. Several erythrobacter species have also been shown to produce sulfur-containing carotenoids. Carotenoid sulfates were first identified from *E. longus*, including erythroxanthin sulfate and caloxanthin sulfate[54]. Additionally, the sulfated carotenoids, bacteriorubioxanthinal and erythroxanthin sulfate have been identified in *E. litoralis*[51] and *E. flavus* strain has been shown to produce erythroxanthin sulphate and caloxanthin sulphate in addition to zeaxanthin sulfate[55].

Prior to our lab's publication of the discovery of erythrazoles A-B[56] and erythrolic acids A-E[57], no other natural products had been reported from an *Erythrobacter* species. The erythrolic acids are meroterpenoids, which contain a terpene chain in addition to a 4-hydroxybenzoic acid. While neither the erythrazoles nor erythrolic acids showed remarkable bioactivity, both structures show an interesting, modified terpene chain. Typically, terpenes contain monoterpenes (C5), diterpenes (C10), sesquiterpenes (C15), or

sesterterpenes (C25). However, the erythrolic acids contain a C22, C17 or C12 terpene side chain[57]. The erythrazoles are comprised of a C20 and C22 terpene chain. The fact that these chains are unusual lengths and end in a carboxylic acid, suggests that the isoprene chain is modified by either an acetate extension to add carbons or an oxidative cleavage, cleaving a longer terpene chain.

In the past two years, there has been several novel discoveries of natural products and bioactivity from *Qipengyuania* species. *Qipengyuania pacifica* has demonstrated antimicrobial potential, though the metabolite responsible for this activity has not been identified[58]. *Qipengyuania* sp. 3-20A1M has been shown to produce pyrrole-2-carboxlic acid, which showed algicidal activity[59]. It was also discovered that *Erythrobacter aquimaris*, which was renamed *Qipengyuania aquimaris*, can produce the polymer, poly-beta-hydroxybutyrate[60]. The recent discovery of natural products from bacteria in the *Qipengyuania* genus suggests that this genus may have novel, chemically interesting capabilities that remain undiscovered.

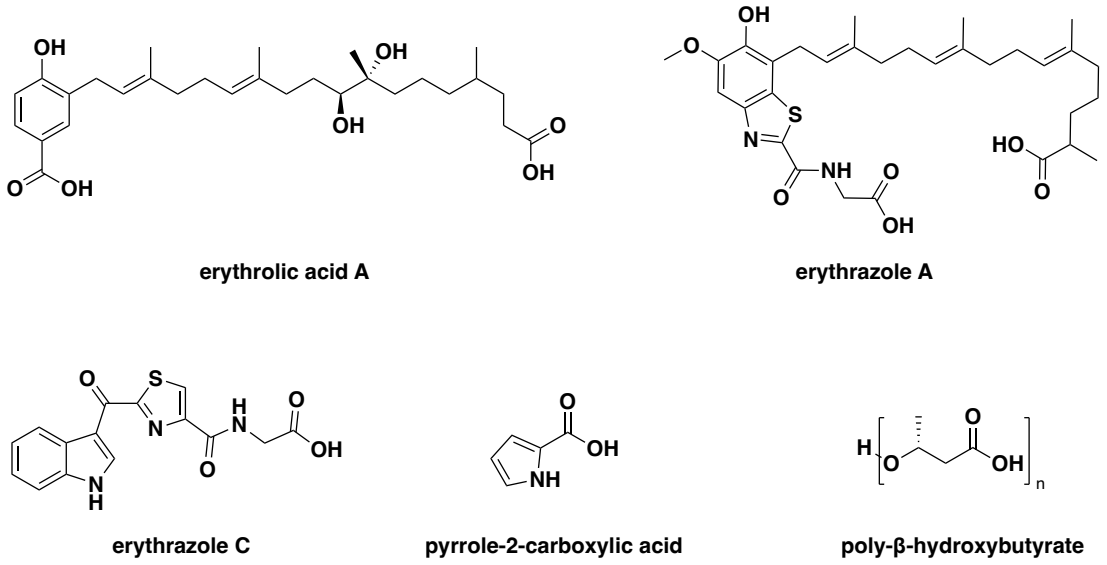


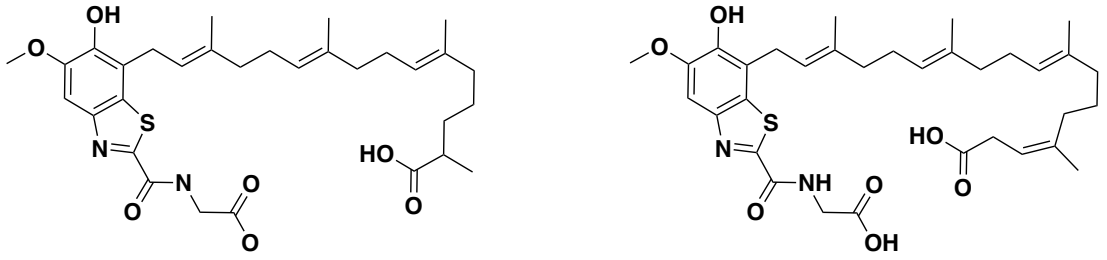
Figure 1.5 Natural products discovered from *Erythrobacter* or *Qipengyuania* species.

While alphaproteobacteria are the most abundant bacteria found in nature, there have been relatively few natural products identified to be produced by them, partially due to culturing challenges[61]. Of all of the known microbial secondary metabolites, approximately 70% are derived from actinomycetes[62,63]. Contrastingly, 50-80% of the microbes in marine environments are in the Proteobacteria phylum while only 5-10% are Actinobacteria[64]. This highlights the unexplored potential for marine natural products from Proteobacteria.

1.4 Significance of erythrazoles A-B and their structure

Erythrazole A and B (EA and EB) were identified via bioassay guided isolation[56], however they drew our attention due to their unusual structure

consisting of amino acids, a benzothiazole, and a C20 or C22 terpene chain ending in a carboxylic acid. Besides the unusual hybrid nature of these molecules, we were particularly intrigued by the terminal oxidation of the isoprene chain and the benzothiazole formation. Though these molecules do not show significant bioactivity, we were interested in studying the unusual composition of their structure to understand novel mechanisms in natural product biosynthesis and to seek out the enzymes responsible for their construction.



Eryrthazole A (EA) **Erythrazole B (EB)**

Figure 1.6 Structures of erythrazole A (EA) and erythrazole B (EB).

1.4.1 Benzothiazoles in natural products

Benzothiazoles have a wide range of biological activities including anticancer, anti-inflammatory and antibacterial properties, making them important and interesting to study[65,66]. However, these moieties are often found in synthetic molecules but are rarely found in natural products. The first benzothiazole marine natural products were isolated from a *Micrococcus*

species in 1991[67]. These were simple benzothiazoles including 2-mercaptobenzothiazole and 2-hydroxybenzothiazole. Several years later, the β_2-adrenoreceptor agonist, S1319 was isolated from a marine sponge[68]. Several naturally occurring benzothiazoles were discovered since, though these remain relatively rare. They have been found from a variety of sources including luciferin from beetles[69,70], violatinctamine from a tunicate[71], and dercitin which was isolated from a sponge (Figure 1.7).

The only known microbially produced benzothiazoles other than the erythrazoles include the thiazinotrienomycines F and G from a *Streptomyces* sp. MJ672-m3[72], rifamycins P and Q, which were identified from *Nocardia mediterranea*[73], and mevashuntin from *Streptomyces prunicolo*r[74] (Figure 1.7). There are two prominent proposals as to how benzothiazoles are naturally formed.

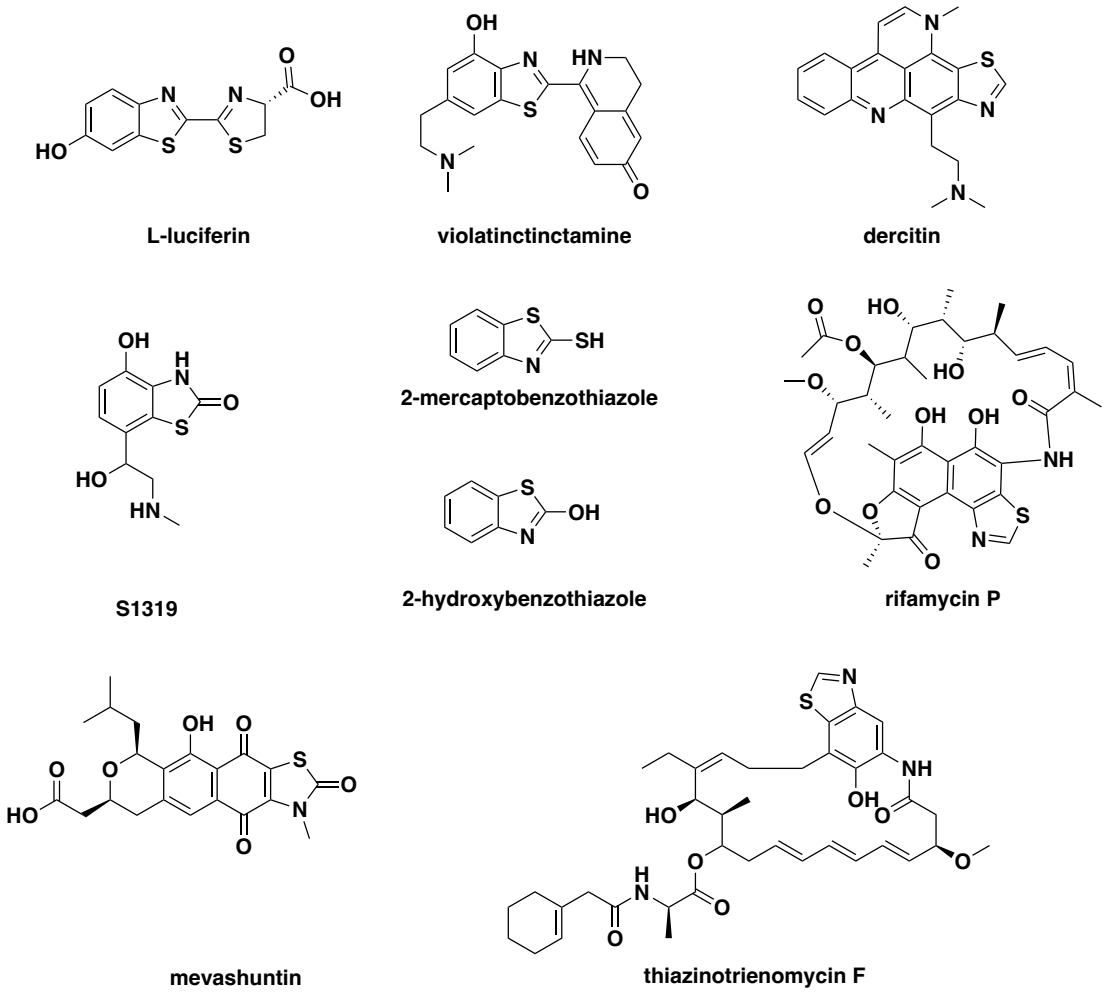


Figure 1.7 Naturally occurring benzothiazoles.

While luciferin, the molecule responsible for beetle bioluminescence, was discovered in the 1970's, its biosynthesis still remains largely unclear. However, the formation of the benzothiazole moiety has been shown to occur through the decarboxylation of cysteine. In this proposal, L-cysteine is incorporated onto p-benzoquinone, which then undergoes elimination of the carboxyl group of the cysteine to form the benzothiazole moiety. Then, a second L-cysteine is incorporated to form the thiazoline moiety (Figure 1.8).

This was demonstrated through stable isotope labeling of the L-cysteine and detected with mass spectrometry[69].

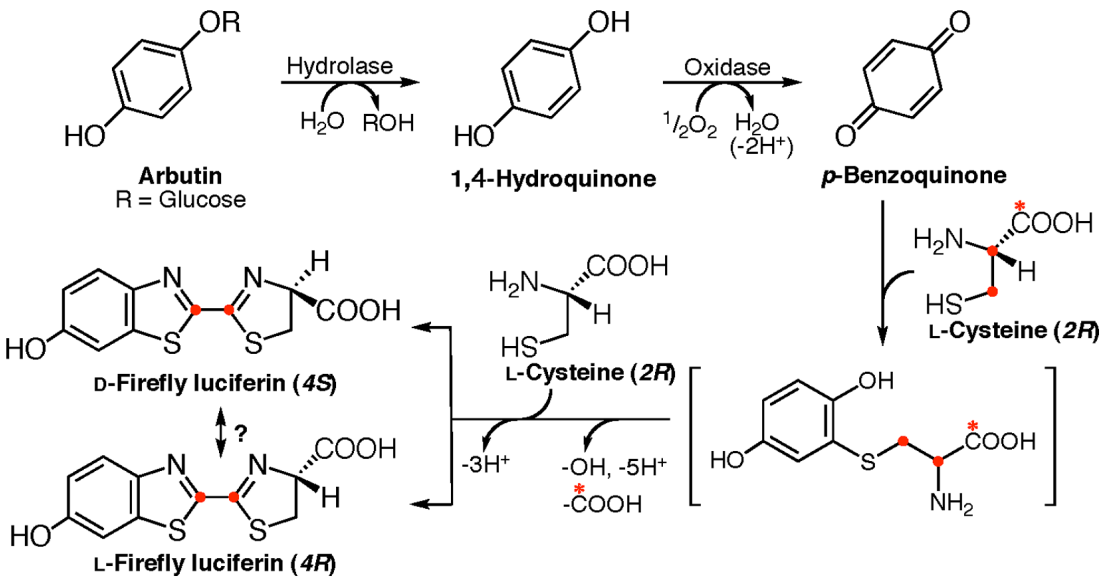


Figure 1.8 Benzothiazole formation mechanism in firefly luciferin demonstrated through stable isotope labeling of cysteine. Figure from Oba et al. Biosynthesis of Firefly Luciferin in Adult Lantern: Decarboxylation of L-cysteine is a Key Step for Benzothiazole Ring Formation in Firefly Luciferin Synthesis. *PLoS ONE* **2013**, 8(12), 1-15. https://doi.org/10.1371/journal.pone.0084023.

Another proposed mechanism of benzothiazole biosynthesis is from the amino acid 5-*S*-cysteinyl-DOPA, which was first proposed in the biosynthesis of the red hair pigment, pheomelanin[75] and later in violatinctamine[71]. It has been proposed that in these molecules, the benzothiazole moiety forms as a product of oxidation in the presence of iron or copper ions at physiological pHs. This proposal suggests that iron ions chelate at the N and O atoms of the benzothiazine, which causes a shift in equilibrium of the benzothiazine to the 4H tautomer. This would allow the generation of an electrophilic species

susceptible of attack at the 2 position by water, which would promote the ring contraction. Further oxidation would then occur followed by an iron-assisted rearrangement to produce the benzothiazole[71,75] (Figure 1.8).

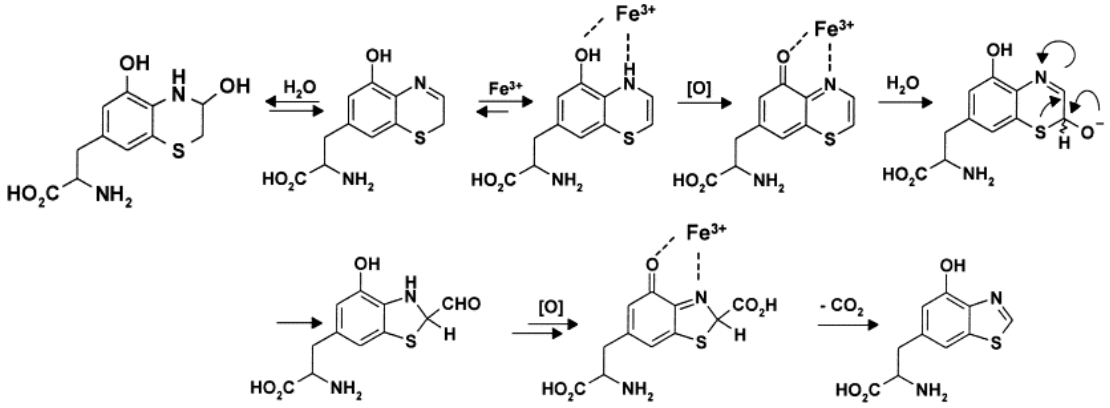


Figure 1.9 Proposed biosynthesis of benzothiazoles through iron assisted oxidation. This has also been proposed to apply with copper ions. Reprinted from *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1571 (2), Di Donato, P.; Napolitano, A.; Prota, G., Metal ions as potential regulatory factors in the biosynthesis of red hair pigments: a new benzothiazole intermediate in the iron or copper assisted oxidation of 5-S-cysteinyldopa. 157-166, **2002**, with permission from Elsevier.

To our knowledge no enzyme has been discovered which can form the benzothiazole moiety of these molecules. We hypothesize that there may be an enzyme capable of these transformations, particularly due to the mechanism we have shown to form the benzothiazole of the erythrazoles (Chapter 3). The discovery of an enzyme that plays a role in benzothiazole biosynthesis would not only be useful in synthetic chemistry but would enable the discovery of other benzothiazoles by genome mining.

1.4.2 Resemblance to ubiquinone and primary metabolism

The substituted aromatic ring of the erythrazoles closely resembles the structure of ubiquinone. Ubiquinone, or coenzyme Q, is a well-studied primary metabolite which transfers electrons in the respiratory electron transport chain and other energy generating processes[76–78]. Its biosynthesis begins with chorismic acid, which is then converted to 4-hydroxybenzoic acid and then is decarboxylated to form 3-octaprenyl-phenol (number of prenyl units varies depending on the species). Then, a series of oxidations and methylations occurs to decorate the aromatic ring of the molecule (Figure 1.10). These steps closely resemble what we predict erythrazole biosynthesis should look like, deviating at the C1-demethyl-C6-demethoxy-Q intermediate (DDMQ). Due to these similarities, we considered the possibility that there is a relationship between ubiquinone and the erythrazoles. As there is precedent of unusual natural products using primary metabolites like NAD, SAM, FMN and heme as substrates (see 1.2.2), we suggest that perhaps the erythrazoles occur from a similar mechanism but with ubiquinone or its intermediates as a precursor. The experiments described in this work aim to characterize the biosynthesis of the erythrazoles and to explore this idea.

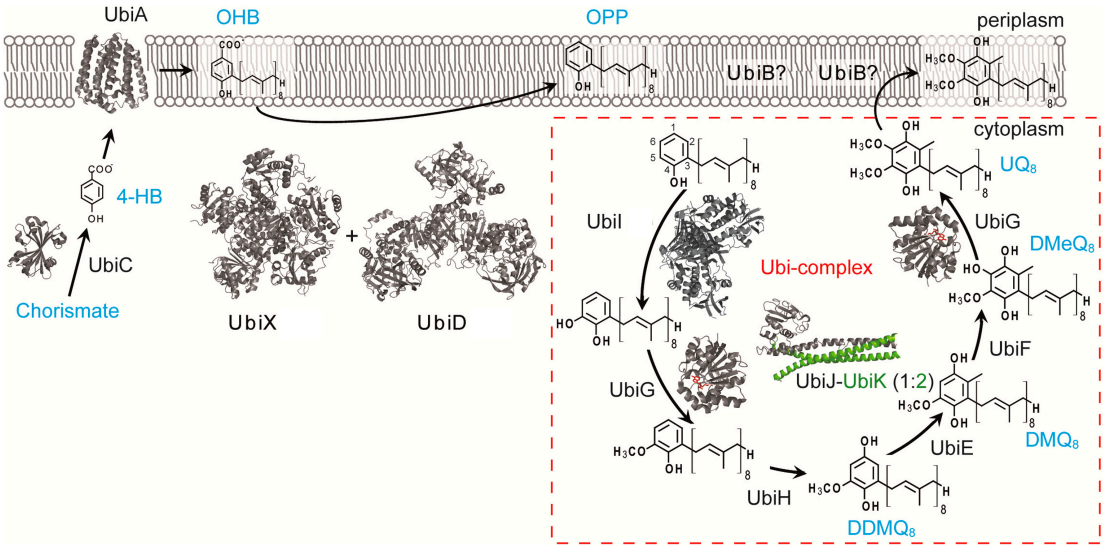


Figure 1.10 Coenzyme-Q_8 biosynthesis in *E. coli*. Reprinted from Figure from *Biochim. et Biophys. Acta (BBA) – Bioenergetics*, 1861 (11), Sophie Saphia Abby, Katayoun Kazemzadeh, Charles Vragniau, Ludovic Pelosi, Fabien Pierrel, Advances in bacterial pathways for the biosynthesis of ubiquinone, page 148259, Copyright (**2020)**, with permission from Elsevier. https://doi.org/10.1016/j.bbabio.2020.14.

1.4.3 Other biosynthetic proposals

In April, 2022, Xie et al. identified meroterpenthiazole A (MerA), a compound with several structural similarities to erythrazoles A-B (Figure 1.11). The only structural differences between these compounds are that meroterpenthiazole A has a cyclized, C15 terpene moiety whereas the erythrazoles have a 20-22 carbon chain and that these molecules have slightly different benzothiazole connectivity. Interestingly, this molecule was isolated from the fungus, *Pencillium allii-sativi*. The authors also found that meroterpenthiazole A inhibits retinoid X receptor (RXR)-α, a target for antitumor

drug design[79]. Due to their similar structures, it would be interesting to determine if the erythrazoles had a similar bioactivity.

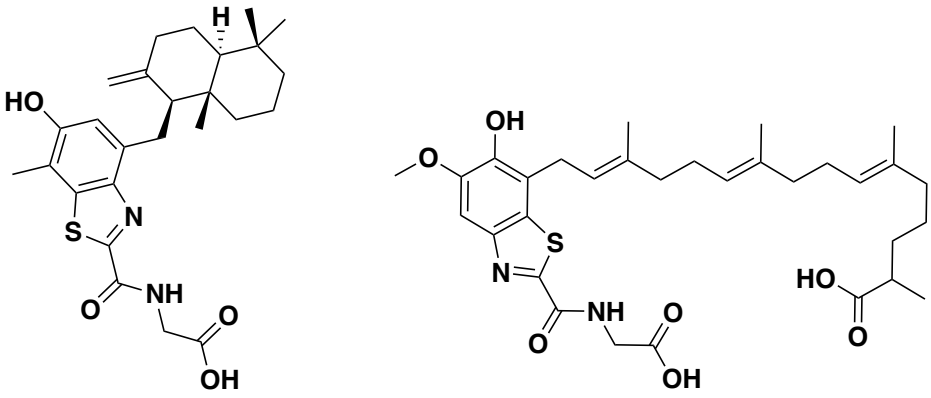


Figure 1.11 Meroterpenthiazole A and erythrazole A.

The authors proposed that 5-farnesyltoluquinol is an intermediate of MerA, which would arise from a biosynthetic gene cluster similar to that of yanuthone D. They proposed that the terpene cyclase *macJ*, identified in macrophorin biosynthesis, would cyclize the sesquiterpene of the molecule[79]. They were able to find homology in the *P. allii-sativi* genome to both the yanuthone D and macrophorin biosynthetic gene clusters to support this hypothesis (Figure 1.12). However, this BGC does not account for the incorporation of cysteine, the ring contraction to form the benzothiazole, and the glycine incorporation of the molecule. They suggested that the ring contraction could arise by the mechanism used in firefly luciferin. However, this mechanism involves a decarboxylation, which would not explain the presence of the cysteine-derived carbonyl in MerA or in the erythrazoles. At least in the

case of the erythrazoles, decarboxylation of cysteine has been shown to not occur, which we demonstrate by stable isotope labeling in chapter 3.

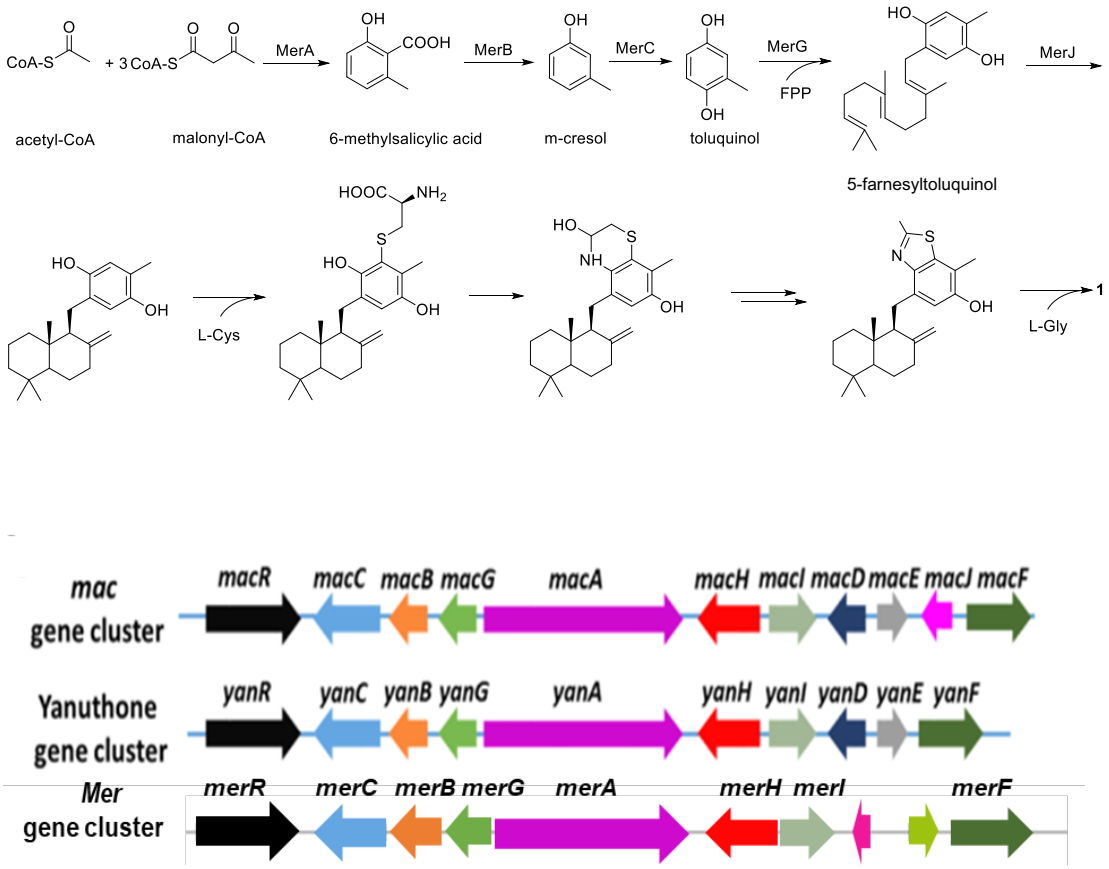


Figure 1.12 Proposed biosynthesis of meroterpenthiazole A and its biosynthetic gene cluster in comparison to the known yanuthone gene cluster and macrocyclin gene cluster. Reprinted from *Chinese Chemical Letters*, 33(4), Xie et al. Meroterpenthiazole A, a unique meroterpenoid from the deep-sea-derived *Penicillium allii-sativi*, significantly inhibited retinoid X receptor (RXR)-*α* transcriptional effect. Pages 2057-2059, Copyright (**2022**). With permission from Elsevier. https://doi.org/10.1016/j.cclet.2021.09.073

Upon querying the SNB-035 genome for this gene cluster, we did not detect any homology. The authors said that antiSMASH detected the BGC, which would suggest that we would be able to detect a similar BGC using this

method. One major discrepancy between an erythrazole BGC and that which they predicted for MerA is that their hypothesis is partially based on the presence of the terpene cyclase, *macJ*. Since the erythrazoles do not have a cyclized terpene, it is unlikely that their BGC would contain a similar enzyme. That being said, meroterpenthiazole A is produced by a fungal species while erythrobacters are produced by an alphaproteobacterial species. The evolutionary gap between these producers, especially since these molecules have not been found from other producers, suggests that they may arise from convergent evolution and therefore may use different mechanisms for biosynthesis.

1.5 Discussion

The evolution of the field of biocatalysis which started from simple reactions using full bacterial cells to entire syntheses by megasynthases or bioengineered enzymes[4,9,17] highlights the potential for growth and new discovery in the field. Regardless of the complexity of the reaction, biocatalysts provide a safer, greener, and often more efficient method for synthesizing molecules[7,80]. Several examples have shown how biosynthetic discoveries can lead to changes in understanding of biosynthetic logic, setting a precedent for future discoveries, and providing access to novel biocatalysts[30,81]. Access to full genome sequencing has greatly increased the ability to detect biosynthetic genes. However, there are still limitations to detecting biosynthetic genes which do not contain well studied core structures like polyketides or polypeptides.

The erythrazoles show interesting potential for biocatalytic discovery due to their unusual hybrid-like structure, their benzothiazole moiety, and because they are produced by a less-studied species in the field of natural products. There has been precedence for non-clustered biosynthetic gene clusters and secondary metabolites being derived from primary metabolites. We predict that the erythrazoles A-B may fall under both of these categories due to their resemblance of ubiquinone and lack of a clear BGC through standard genome mining strategies. Discovering noncanonical biosynthetic genes and novel biosynthetic mechanisms is challenging but rewarding. These discoveries may set the precedent for future discoveries of novel metabolites and their biosyntheses in addition to potentially providing a new biocatalyst for chemical syntheses.

We believe that studying the erythrazoles may lead to uncovering novel mechanisms and biocatalysts for benzothiazole formation and terpene chain modification. Currently accepted mechanisms for benzothiazole formation include the carboxylic acid removal and cyclization of cysteine in firefly luciferin (Figure 1.8) or the iron mediated oxidation that occurs in the pheomelanins (Figure 1.9). Our work shows a novel mechanism to form a benzothiazole, which differs from both of these mechanisms and suggests the potential for the discovery of an enzyme which is capable of forming benzothiazoles. This enzyme could have a wide range of applications as benzothiazoles are prevalent in many bioactive and industrially relevant compounds[65,66]. Terpenes

are also found in a wide range of natural products and used in various industries, so detecting an enzyme which can manipulate them could have biocatalytic potential in addition to setting precedent for other secondary metabolites and their biosynthesis to be discovered. The following chapters aim to pursue erythrazole biosynthesis as a means to understanding the biosynthetic logic behind these unusual molecules in addition to identifying potential biocatalysts.

1.6 Chapter one references

(1) Scott, T. A.; Piel, J. The Hidden Enzymology of Bacterial Natural Product Biosynthesis. *Nat. Rev. Chem.* **2019**, *3* (7), 404–425. https://doi.org/10.1038/s41570-019-0107-1.

(2) Atanasov, A. G. et al; the International Natural Product Sciences Taskforce. Natural Products in Drug Discovery: Advances and Opportunities. *Nat. Rev. Drug Discov.* **2021**, *20* (3), 200–216. https://doi.org/10.1038/s41573-020-00114-z.

(3) Yi, D.; Bayer, T.; Badenhorst, C. P.; Wu, S.; Doerr, M.; Höhne, M.; Bornscheuer, U. T. Recent Trends in Biocatalysis. *Chem. Soc. Rev.* **2021**, *50* (14), 8003–8049.

(4) Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K. Engineering the Third Wave of Biocatalysis. *Nature* **2012**, *485* (7397), 185–194. https://doi.org/10.1038/nature11117.

(5) Butler, M. S. The Role of Natural Product Chemistry in Drug Discovery. *J. Nat. Prod.* **2004**, *67* (12), 2141–2153. https://doi.org/10.1021/np040106y.

(6) Zheng, Q.; Lin, W.; Liu, Y.; Chen, C.; Jiao, N. A Comparison of 14 Erythrobacter Genomes Provides Insights into the Genomic Divergence and Scattered Distribution of Phototrophs. *Front. Microbiol.* **2016**, *7*. https://doi.org/10.3389/fmicb.2016.00984.

(7) Alcántara, A. R.; Domínguez de María, P.; Littlechild, J. A.; Schürmann, M.; Sheldon, R. A.; Wohlgemuth, R. Biocatalysis as Key to Sustainable Industrial Chemistry. *ChemSusChem* **2022**, *15* (9), e202102709. https://doi.org/10.1002/cssc.202102709.

(8) Brustad, E. M. New Recipes for Biocatalysis. *Nat. Chem. Biol.* **2014**, *10* (3), 170–171. https://doi.org/10.1038/nchembio.1457.

(9) Sheldon, R. A.; Brady, D.; Bode, M. L. The Hitchhiker's Guide to Biocatalysis: Recent Advances in the Use of Enzymes in Organic Synthesis. *Chem. Sci.* **2020**, *11* (10), 2587–2605.

(10) Li, B.; Su, J.; Tao, J. Enzyme and Process Development for Production of Nicotinamide. *Org. Process Res. Dev.* **2011**, *15* (1), 291–293. https://doi.org/10.1021/op100309g.

(11) Isowa, Y.; Ohmori, M.; Ichikawa, T.; Mori, K.; Nonaka, Y.; Kihara, K.; Oyama, K.; Satoh, H.; Nishimura, S. The Thermolysin-Catalyzed Condensation Reactions of n-Substituted Aspartic and Glutamic Acids with Phenylalanine Alkyl Esters. *Tetrahedron Lett.* **1979**, *20* (28), 2611–2612. https://doi.org/10.1016/S0040-4039(01)86363-2.

(12) Bruggink, A.; Roos, E. C.; de Vroom, E. Penicillin Acylase in the Industrial Production of β-Lactam Antibiotics. *Org. Process Res. Dev.* **1998**, *2* (2), 128–133. https://doi.org/10.1021/op9700643.

(13) Bertoldo, C.; Antranikian, G. Starch-Hydrolyzing Enzymes from Thermophilic Archaea and Bacteria. *Curr. Opin. Chem. Biol.* **2002**, *6* (2), 151–160. https://doi.org/10.1016/S1367-5931(02)00311-3.

(14) Sheldon, R. A.; Brady, D. The Limits to Biocatalysis: Pushing the Envelope. *Chem. Commun.* **2018**, *54* (48), 6088–6104.

(15) Lukowski, A. L.; Mallik, L.; Hinze, M. E.; Carlson, B. M.; Ellinwood, D. C.; Pyser, J. B.; Koutmos, M.; Narayan, A. R. H. Substrate Promiscuity of a Paralytic Shellfish Toxin Amidinotransferase. *ACS Chem. Biol.* **2020**, *15* (3), 626–631. https://doi.org/10.1021/acschembio.9b00964.

(16) Lukowski, A. L.; Ellinwood, D. C.; Hinze, M. E.; DeLuca, R. J.; Du Bois, J.; Hall, S.; Narayan, A. R. H. C–H Hydroxylation in Paralytic Shellfish Toxin Biosynthesis. *J. Am. Chem. Soc.* **2018**, *140* (37), 11863–11869. https://doi.org/10.1021/jacs.8b08901.

(17) Pyser, J. B.; Chakrabarty, S.; Romero, E. O.; Narayan, A. R. H. State-of-the-Art Biocatalysis. *ACS Cent. Sci.* **2021**, *7* (7), 1105–1116. https://doi.org/10.1021/acscentsci.1c00273.

(18) Nakamura, H.; Schultz, E. E.; Balskus, E. P. A New Strategy for Aromatic Ring Alkylation in Cylindrocyclophane Biosynthesis. *Nat. Chem. Biol.* **2017**, *13* (8), 916–921. https://doi.org/10.1038/nchembio.2421.

(19) Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H. Olefin Cyclopropanation via Carbene Transfer Catalyzed by Engineered Cytochrome P450 Enzymes. *Science* **2013**, *339* (6117), 307–310. https://doi.org/10.1126/science.1231434.

(20) Huffman, M. A.; Fryszkowska, A.; Alvizo, O.; Borra-Garske, M.; Campos, K. R.; Canada, K. A.; Devine, P. N.; Duan, D.; Forstater, J. H.; Grosser, S. T.; Halsey, H. M.; Hughes, G. J.; Jo, J.; Joyce, L. A.; Kolev, J. N.; Liang, J.; Maloney, K. M.; Mann, B. F.; Marshall, N. M.; McLaughlin, M.; Moore, J. C.; Murphy, G. S.; Nawrat, C. C.; Nazor, J.; Novick, S.; Patel, N. R.; Rodriguez-Granillo, A.; Robaire, S. A.; Sherer, E. C.; Truppo, M. D.; Whittaker, A. M.; Verma, D.; Xiao, L.; Xu, Y.; Yang, H. Design of an in Vitro Biocatalytic Cascade for the Manufacture of Islatravir. *Science* **2019**, *366* (6470), 1255–1259. https://doi.org/10.1126/science.aay8484.

(21) Ikeda, H.; Ishikawa, J.; Hanamoto, A.; Shinose, M.; Kikuchi, H.; Shiba, T.; Sakaki, Y.; Hattori, M.; Ōmura, S. Complete Genome Sequence and Comparative Analysis of the Industrial Microorganism Streptomyces Avermitilis. *Nat. Biotechnol.* **2003**, *21* (5), 526–531. https://doi.org/10.1038/nbt820.

(22) Bentley, S. D.; Chater, K. F.; Cerdeño-Tárraga, A.-M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; Bateman, A.; Brown, S.; Chandra, G.; Chen, C. W.; Collins, M.; Cronin, A.; Fraser, A.; Goble, A.; Hidalgo, J.; Hornsby, T.; Howarth, S.; Huang, C.-H.; Kieser, T.; Larke, L.; Murphy, L.; Oliver, K.; O'Neil, S.; Rabbinowitsch, E.; Rajandream, M.-A.; Rutherford, K.; Rutter, S.; Seeger, K.; Saunders, D.; Sharp, S.; Squares, R.; Squares, S.; Taylor, K.; Warren, T.; Wietzorrek, A.; Woodward, J.; Barrell, B. G.; Parkhill, J.; Hopwood, D. A. Complete Genome Sequence of the Model Actinomycete Streptomyces Coelicolor A3(2). *Nature* **2002**, *417* (6885), 141–147. https://doi.org/10.1038/417141a.

(23) Bode, H. B.; Müller, R. The Impact of Bacterial Genomics on Natural Product Research. *Angew. Chem. Int. Ed.* **2005**, *44* (42), 6828–6846. https://doi.org/10.1002/anie.200501080.

(24) Walsh, C. T.; Fischbach, M. A. Natural Products Version 2.0: Connecting Genes to Molecules. *J. Am. Chem. Soc.* **2010**, *132* (8), 2469–2493. https://doi.org/10.1021/ja909118a.

(25) Cane, D. E.; Walsh, C. T.; Khosla, C. Harnessing the Biosynthetic Code: Combinations, Permutations, and Mutations. *Science* **1998**, *282* (5386), 63–68. https://doi.org/10.1126/science.282.5386.63.

(26) Ziemert, N.; Alanjary, M.; Weber, T. The Evolution of Genome Mining in Microbes–a Review. *Nat. Prod. Rep.* **2016**, *33* (8), 988–1005.

(27) Wang, H.; Fewer, D. P.; Holm, L.; Rouhiainen, L.; Sivonen, K. Atlas of Nonribosomal Peptide and Polyketide Biosynthetic Pathways Reveals Common Occurrence of Nonmodular Enzymes. *Proc. Natl. Acad. Sci.* **2014**, *111* (25), 9259–9264. https://doi.org/10.1073/pnas.1401734111.

(28) Fischbach, M. A.; Walsh, C. T. Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. *Chem. Rev.* **2006**, *106* (8), 3468–3496. https://doi.org/10.1021/cr0503097.

(29) Barra, L.; Awakawa, T.; Abe, I. Noncanonical Functions of Enzyme Cofactors as Building Blocks in Natural Product Biosynthesis. *JACS Au* **2022**, *2* (9), 1950–1963. https://doi.org/10.1021/jacsau.2c00391.

(30) Biermann, F.; Helfrich, E. J. N. Hidden Treasures: Microbial Natural Product Biosynthesis off the Beaten Path. *mSystems* **2021**, *6* (4), 10.1128/msystems.00846-21. https://doi.org/10.1128/msystems.00846-21.

(31) Nguyen, T.; Ishida, K.; Jenke-Kodama, H.; Dittmann, E.; Gurgui, C.; Hochmuth, T.; Taudien, S.; Platzer, M.; Hertweck, C.; Piel, J. Exploiting the Mosaic Structure of Trans-Acyltransferase Polyketide Synthases for Natural Product Discovery and Pathway Dissection. *Nat. Biotechnol.* **2008**, *26* (2), 225–233. https://doi.org/10.1038/nbt1379.

(32) Helfrich, E. J. N.; Ueoka, R.; Dolev, A.; Rust, M.; Meoded, R. A.; Bhushan, A.; Califano, G.; Costa, R.; Gugger, M.; Steinbeck, C.; Moreno, P.; Piel, J. Automated Structure Prediction of Trans-Acyltransferase Polyketide Synthase Products. *Nat. Chem. Biol.* **2019**, *15* (8), 813–821. https://doi.org/10.1038/s41589-019-0313-7.

(33) Lv, J.-M.; Gao, Y.-H.; Zhao, H.; Awakawa, T.; Liu, L.; Chen, G.-D.; Yao, X.-S.; Hu, D.; Abe, I.; Gao, H. Biosynthesis of Biscognienyne B Involving a Cytochrome P450-Dependent Alkynylation. *Angew. Chem. Int. Ed.* **2020**, *59* (32), 13531–13536. https://doi.org/10.1002/anie.202004364.

(34) Pan, G.; Xu, Z.; Guo, Z.; null; Ma, M.; Yang, D.; Zhou, H.; Gansemans, Y.; Zhu, X.; Huang, Y.; Zhao, L.-X.; Jiang, Y.; Cheng, J.; Nieuwerburgh, F. V.; Suh, J.-W.; Duan, Y.; Shen, B. Discovery of the Leinamycin Family of Natural Products by Mining Actinobacterial Genomes. *Proc. Natl. Acad. Sci.* **2017**, *114* (52), E11131–E11140. https://doi.org/10.1073/pnas.1716245115.

(35) Tan, D.; Jamieson, C. S.; Ohashi, M.; Tang, M.-C.; Houk, K. N.; Tang, Y. Genome-Mined Diels–Alderase Catalyzes Formation of the Cis-Octahydrodecalins of Varicidin A and B. *J. Am. Chem. Soc.* **2019**, *141* (2), 769–773. https://doi.org/10.1021/jacs.8b12010.

(36) Barra, L.; Awakawa, T.; Shirai, K.; Hu, Z.; Bashiri, G.; Abe, I. β-NAD as a Building Block in Natural Product Biosynthesis. *Nature* **2021**, *600* (7890), 754–758. https://doi.org/10.1038/s41586-021-04214-7.

(37) Jhulki, I.; Chanani, P. K.; Abdelwahed, S. H.; Begley, T. P. A Remarkable Oxidative Cascade That Replaces the Riboflavin C8 Methyl with an Amino Group during Roseoflavin Biosynthesis. *J. Am. Chem. Soc.* **2016**, *138* (27), 8324–8327. https://doi.org/10.1021/jacs.6b02469.

(38) Tongsook, C.; Uhl, M. K.; Jankowitsch, F.; Mack, M.; Gruber, K.; Macheroux, P. Structural and Kinetic Studies on RosA, the Enzyme Catalysing the Methylation of 8-Demethyl-8-Amino-d-Riboflavin to the Antibiotic Roseoflavin. *FEBS J.* **2016**, *283* (8), 1531–1549. https://doi.org/10.1111/febs.13690.

(39) Jin, X.; Zhang, Y.; Zhang, R.; Nguyen, K.-U.; Lindsey, J. S.; Miller, E. S. Identification of Putative Biosynthetic Gene Clusters for Tolyporphins in Multiple Filamentous Cyanobacteria. *Life* **2021**, *11* (8). https://doi.org/10.3390/life11080758.

(40) Sufrin, J. R.; Finckbeiner, S.; Oliver, C. M. Marine-Derived Metabolites of S-Adenosylmethionine as Templates for New Anti-Infectives. *Mar. Drugs* **2009**, *7* (3), 401–434. https://doi.org/10.3390/md7030401.

(41) Richter, T. K. S.; Hughes, C. C.; Moore, B. S. Sioxanthin, a Novel Glycosylated Carotenoid, Reveals an Unusual Subclustered Biosynthetic Pathway. *Environ. Microbiol.* **2015**, *17* (6), 2158–2171. https://doi.org/10.1111/1462-2920.12669.

(42) Takaichi, S.; Maoka, T.; Takasaki, K.; Hanada, S. Carotenoids of Gemmatimonas Aurantiaca (Gemmatimonadetes): Identification of a

Novel Carotenoid, Deoxyoscillol 2-Rhamnoside, and Proposed Biosynthetic Pathway of Oscillol 2,2′-Dirhamnoside. *Microbiology*, 2010, *156*, 757–763. https://doi.org/10.1099/mic.0.034249-0.

(43) Tian, B.; Hua, Y. Carotenoid Biosynthesis in Extremophilic Deinococcus–Thermus Bacteria. *Trends Microbiol.* **2010**, *18* (11), 512–520. https://doi.org/10.1016/j.tim.2010.07.007.

(44) Rassbach, J.; Hilsberg, N.; Haensch, V. G.; Dörner, S.; Gressler, J.; Sonnabend, R.; Semm, C.; Voigt, K.; Hertweck, C.; Gressler, M. Non-Canonical Two-Step Biosynthesis of Anti-Oomycete Indole Alkaloids in Kickxellales. *Fungal Biol. Biotechnol.* **2023**, *10* (1), 19. https://doi.org/10.1186/s40694-023-00166-x.

(45) Álvarez-Álvarez, R.; Botas, A.; Albillos, S. M.; Rumbero, A.; Martín, J. F.; Liras, P. Molecular Genetics of Naringeninbiosynthesis, a Typical Plant Secondary Metabolite Produced by Streptomyces Clavuligerus. *Microb. Cell Factories* **2015**, *14* (1), 178. https://doi.org/10.1186/s12934-015-0373-7.

(46) Liu, Y.; Pei, T.; Du, J.; Yao, Q.; Deng, M.-R.; Zhu, H. Comparative Genomics Reveals Genetic Diversity and Metabolic Potentials of the Genus Qipengyuania and Suggests Fifteen Novel Species. *Microbiol. Spectr.* **2022**, *10* (3), e0126421. https://doi.org/10.1128/spectrum.01264-21.

(47) Denner, E. B.; Vybiral, D.; Koblízek, M.; Kämpfer, P.; Busse, H.-J.; Velimirov, B. Erythrobacter Citreus Sp. Nov., a Yellow-Pigmented Bacterium That Lacks Bacteriochlorophyll a, Isolated from the Western Mediterranean Sea. *Int. J. Syst. Evol. Microbiol.* **2002**, *52* (5), 1655–1661.

(48) Koblížek, M.; Béjà, O.; Bidigare, R. R.; Christensen, S.; Benitez-Nelson, B.; Vetriani, C.; Kolber, M. K.; Falkowski, P. G.; Kolber, Z. S. Isolation and Characterization of Erythrobacter Sp. Strains from the Upper Ocean. *Arch. Microbiol.* **2003**, *180*, 327–338.

(49) SHIBA, T.; SIMIDU, U. Erythrobacter Longus Gen. Nov., Sp. Nov., an Aerobic Bacterium Which Contains Bacteriochlorophyll a. *International Journal of Systematic and Evolutionary Microbiology*, 1982, *32*, 211–217. https://doi.org/10.1099/00207713-32-2-211.

(50) Feng, X.-M.; Mo, Y.-X.; Han, L.; Nogi, Y.; Zhu, Y.-H.; Lv, J. Qipengyuania Sediminis Gen. Nov., Sp. Nov., a Member of the Family Erythrobacteraceae Isolated from Subterrestrial Sediment. *International Journal of Systematic and Evolutionary Microbiology*, 2015, *65*, 3658–3665. https://doi.org/10.1099/ijsem.0.000472.

(51) Yurkov, V.; Stackebrandt, E.; Holmes, A.; Fuerst, J. A.; Hugenholtz, P.; Golecki, J.; Gad'on, N.; Gorlenko, V. M.; Kompantseva, E. I.; Drews, G. Phylogenetic Positions of Novel Aerobic, Bacteriochlorophyll a-Containing Bacteria and Description of Roseococcus Thiosulfatophilus Gen. Nov., Sp. Nov., Erythromicrobium Ramosum Gen. Nov., Sp. Nov., and Erythrobacter Litoralis Sp. Nov. *Int. J. Syst. Evol. Microbiol.* **1994**, *44* (3), 427–434.

(52) Fiebig, A.; Varesio, L. M.; Alejandro Navarreto, X.; Crosson, S. Regulation of the Erythrobacter Litoralis DSM 8509 General Stress Response by Visible Light. *Mol. Microbiol.* **2019**, *112* (2), 442–460. https://doi.org/10.1111/mmi.14310.

(53) Takaichi, S.; Shimada, K.; Ishidsu, J. Carotenoids from the Aerobic Photosynthetic Bacterium, Erythrobacter Longus: β-Carotene and Its Hydroxyl Derivatives. *Arch. Microbiol.* **1990**, *153*, 118–122.

(54) Takaichi, S.; Furihata, K.; Ishidsu, J.; Shimada, K. Carotenoid Sulphates from the Aerobic Photosynthetic Bacterium, Erythrobacter Longus. *Phytochemistry* **1991**, *30* (10), 3411–3415. https://doi.org/10.1016/0031-9422(91)83219-B.

(55) Setiyono, E.; Heriyanto; Pringgenies, D.; Shioi, Y.; Kanesaki, Y.; Awai, K.; Brotosudarmo, T. H. P. Sulfur-Containing Carotenoids from a Marine Coral Symbiont Erythrobacter Flavus Strain KJ5. *Mar. Drugs* **2019**, *17* (6), 349.

(56) Hu, Y.; MacMillan, J. B. Erythrazoles A–B, Cytotoxic Benzothiazoles from a Marine-Derived *Erythrobacter* Sp. *Org. Lett.* **2011**, *13* (24), 6580–6583. https://doi.org/10.1021/ol202944g.

(57) Hu, Y.; Legako, A. G.; Espindola, A. P. D.; MacMillan, J. B. Erythrolic Acids A–E, Meroterpenoids from a Marine-Derived Erythrobacter Sp. *J. Org. Chem.* **2012**, *77* (7), 3401–3407.

(58) Tareen, S.; Risdian, C.; Müsken, M.; Wink, J. Qipengyuania Pacifica Sp. Nov., a Novel Carotenoid-Producing Marine Bacterium of the Family Erythrobacteraceae, Isolated from Sponge (Demospongiae), and Antimicrobial Potential of Its Crude Extract. *Diversity* **2022**, *14* (4), 295.

(59) Ko, S.-R.; Le, V. V.; Srivastava, A.; Kang, M.; Oh, H.-M.; Ahn, C.-Y. Algicidal Activity of a Novel Bacterium, Qipengyuania Sp. 3-20A1M, against Harmful Margalefidinium Polykrikoides: Effects of Its Active Compound. *Mar. Pollut. Bull.* **2023**, *186*, 114397. https://doi.org/10.1016/j.marpolbul.2022.114397.

(60) Mostafa, Y. S.; Alrumman, S. A.; Otaif, K. A.; Alamri, S. A.; Mostafa, M. S.; Sahlabji, T. Production and Characterization of Bioplastic by Polyhydroxybutyrate Accumulating Erythrobacter Aquimaris Isolated from Mangrove Rhizosphere. *Molecules* **2020**, *25* (1), 179.

(61) Joint, I.; Mühling, M.; Querellou, J. Culturing Marine Bacteria – an Essential Prerequisite for Biodiscovery. *Microb. Biotechnol.* **2010**, *3* (5), 564–575. https://doi.org/10.1111/j.1751-7915.2010.00188.x.

(62) Khalifa, S. A. M.; Elias, N.; Farag, M. A.; Chen, L.; Saeed, A.; Hegazy, M.-E. F.; Moustafa, M. S.; Abd El-Wahed, A.; Al-Mousawi, S. M.; Musharraf, S. G.; Chang, F.-R.; Iwasaki, A.; Suenaga, K.; Alajlani, M.; Göransson, U.; El-Seedi, H. R. Marine Natural Products: A Source of Novel Anticancer Drugs. *Mar. Drugs* **2019**, *17* (9). https://doi.org/10.3390/md17090491.

(63) Carroll, A. R.; Copp, B. R.; Davis, R. A.; Keyzers, R. A.; Prinsep, M. R. Marine Natural Products. *Nat. Prod. Rep.* **2020**, *37* (2), 175–223.

(64) Desriac, F.; Jégou, C.; Balnois, E.; Brillet, B.; Chevalier, P. L.; Fleury, Y. Antimicrobial Peptides from Marine Proteobacteria. *Mar. Drugs* **2013**, *11* (10), 3632–3660. https://doi.org/10.3390/md11103632.

(65) Keri, Rangappa S.; Patil, M.; Patil, S.; Budagumpi, S. A Comprehensive Review in Current Developments of Benzothiazole- Based Molecules in Medicinal Chemistry. *Eur. J. Med. Chem.* **2015**, *89*, 207–251.

(66) Irfan, A.; Batool, F.; Zahra Naqvi, S. A.; Islam, A.; Osman, S. M.; Nocentini, A.; Alissa, S. A.; Supuran, C. T. Benzothiazole Derivatives as Anticancer Agents. *J. Enzyme Inhib. Med. Chem.* **2020**, *35* (1), 265–279. https://doi.org/10.1080/14756366.2019.1698036.

(67) Stierle, A. A.; Cardellina II, J. H.; Singleton, F. L. Benzothiazoles from a Putatitve Bacterial Symbiont of the Marine Sponge Tedania Ignis. *Tetrahedron Lett.* **1991**, *32* (37), 4847–4848.

(68) Suzuki, H.; Shindo, K.; Ueno, A.; Miura, T.; Takei, M.; Sakakibara, M.; Fukamachi, H.; Tanaka, J.; Higa, T. S1319: A Novel B2-Adrenoceptor Agonist from a Marine Sponge Dysidea Sp. *Bioorg. Med. Chem. Lett.* **1999**, *9* (10), 1361–1364. https://doi.org/10.1016/S0960-894X(99)00205-X.

(69) Oba, Y.; Yoshida, N.; Kanie, S.; Ojika, M.; Inouye, S. Biosynthesis of Firefly Luciferin in Adult Lantern: Decarboxylation of L-Cysteine Is a Key Step for Benzothiazole Ring Formation in Firefly Luciferin Synthesis. *PloS One* **2013**, *8* (12), e84023. https://doi.org/10.1371/journal.pone.0084023.

(70) Zhang, R.; He, J.; Dong, Z.; Liu, G.; Yin, Y.; Zhang, X.; Li, Q.; Ren, Y.; Yang, Y.; Liu, W.; Chen, X.; Xia, W.; Duan, K.; Hao, F.; Lin, Z.; Yang, J.; Chang, Z.; Zhao, R.; Wan, W.; Lu, S.; Peng, Y.; Ge, S.; Wang, W.; Li, X. Genomic and Experimental Data Provide New Insights into Luciferin

Biosynthesis and Bioluminescence Evolution in Fireflies. *Sci. Rep.* **2020**, *10* (1), 15882. https://doi.org/10.1038/s41598-020-72900-z.

(71) Chill, L.; Rudi, A.; Benayahu, Y.; Kashman, Y. Violatinctamine, a New Heterocyclic Compound from the Marine Tunicate Cystodytes Cf. Violatinctus. *Tetrahedron Lett.* **2004**, *45* (42), 7925–7928. https://doi.org/10.1016/j.tetlet.2004.08.137.

(72) Hosokawa, N.; Naganawa, H.; Hamada, M.; Iinuma, H.; Takeuchi, T.; Tsuchiya, K. S.; Hori, M. New Triene-Ansamycins, Thiazinotrienomycins F and G and a Diene-Ansamycin, Benzoxazomycin. *J. Antibiot. (Tokyo)* **2000**, *53* (9), 886–894. https://doi.org/10.7164/antibiotics.53.886.

(73) Cricchio, R.; Antonini, P.; Sartori, G. THIAZORIFAMYCINS. III BIOSYNTHESIS OF RIFAMYCINS P, Q AND VERDE, NOVEL METABOLITES FROM A MUTANT OF NOCARDIA MEDITERRANEA. *J. Antibiot. (Tokyo)* **1980**, *33* (8), 842–846.

(74) Shin-ya, K.; Umeda, Y.; Chijiwa, S.; Furihata, K.; Furihata, K.; Hayakawa, Y.; Seto, H. Mevashuntin, a Novel Metabolite Produced by Inhibition of the Mevalonate Pathway in Streptomyces Prunicolor. *Tetrahedron Lett.* **2005**, *46* (8), 1273–1276. https://doi.org/10.1016/j.tetlet.2004.12.134.

(75) Di Donato, P.; Napolitano, A.; Prota, G. Metal Ions as Potential Regulatory Factors in the Biosynthesis of Red Hair Pigments: A New Benzothiazole Intermediate in the Iron or Copper Assisted Oxidation of 5-S-Cysteinyldopa. *Biochim. Biophys. Acta BBA-Gen. Subj.* **2002**, *1571* (2), 157–166.

(76) Olson, R. E.; Rudney, H. Biosynthesis of Ubiquinone; Aurbach, G. D., McCormick, D. B., Eds.; Vitamins & Hormones; Academic Press, 1983; Vol. 40, pp 1–43. https://doi.org/10.1016/S0083-6729(08)60431-8.

(77) Meganathan, R. Ubiquinone Biosynthesis in Microorganisms. *FEMS Microbiol. Lett.* **2001**, *203* (2), 131–139. https://doi.org/10.1111/j.1574-6968.2001.tb10831.x.

(78) Abby, S. S.; Kazemzadeh, K.; Vragniau, C.; Pelosi, L.; Pierrel, F. Advances in Bacterial Pathways for the Biosynthesis of Ubiquinone. *Biochim. Biophys. Acta BBA - Bioenerg.* **2020**, *1861* (11), 148259. https://doi.org/10.1016/j.bbabio.2020.148259.

(79) Xie, C.-L.; Zhang, D.; Guo, K.-Q.; Yan, Q.-X.; Zou, Z.-B.; He, Z.-H.; Wu, Z.; Zhang, X.-K.; Chen, H.-F.; Yang, X.-W. Meroterpenthiazole A, a Unique Meroterpenoid from the Deep-Sea-Derived Penicillium Allii-Sativi, Significantly Inhibited Retinoid X Receptor (RXR)-α Transcriptional Effect. *Chin. Chem. Lett.* **2022**, *33* (4), 2057–2059. https://doi.org/10.1016/j.cclet.2021.09.073.

(80) Bell, E. L.; Finnigan, W.; France, S. P.; Green, A. P.; Hayes, M. A.; Hepworth, L. J.; Lovelock, S. L.; Niikura, H.; Osuna, S.; Romero, E.; Ryan, K. S.; Turner, N. J.; Flitsch, S. L. Biocatalysis. *Nat. Rev. Methods Primer* **2021**, *1* (1), 46. https://doi.org/10.1038/s43586-021-00044-z.

(81) Scherlach, K.; Hertweck, C. Mining and Unearthing Hidden Biosynthetic Potential. *Nat. Commun.* **2021**, *12* (1), 3864. https://doi.org/10.1038/s41467-021-24133-5.

CHAPTER TWO

EXPLORATION OF ERYTHRAZOLE A AND B BIOSYNTHESIS THROUGH STABLE ISOTOPE LABELING

2.1 Identification and characterization of erythrazole A-B

Erythrazole A and B were first identified by the MacMillan lab where the authors were able to obtain 4.1g of extract from 30 L of bacterial culture which after chromatography methods yielded erythrazole A (1.0mg) and erythrazole B (0.6mg). The following methods were based on those previously described. Detection of these molecules was performed using HR-LC-MS.

2.1.1 Cultivation and extraction

Bacterium SNB-035 was cultured in 5 L in the seawater based medium, A1BFe+C (10g starch, 4g yeast extract, 2g peptone, 1g $CaCO_3$, 40mg $Fe_2(SO_4)3*4H_2O$, 100 mg KBr). After 5 days of shaking at 200 rpm at 30°C, sterilized XAD-7-HP resin was added to the cultures and left shaking for 2 hours. The resin was then filtered through cheesecloth and submerged in acetone shaking overnight at 30°C, 200rpm. This resulted in 506.9mg of crude extract, which was later purified for further analysis. For smaller scale extractions (50mL-1L growths), liquid:liquid partitioning was performed. 1:1 ethyl acetate was added to bacterial cultures, shaken and allowed to separate. The ethyl acetate soluble layer was collected in a round bottom flask and this was repeated 3 times. This resulted in ~300mg of crude per 1 liter of culture.

2.1.2 Isolation

Isolation method used was adapted from the method described in Hu and MacMillan[1]. The crude extract was fractionated by flash column chromatography, (Redisep column: C18 30g Gold, 35 ml/minute, UV 200nm - 360 nm) using a gradient solvent system of H_2O and MeOH, with increasing MeOH from 5% to 100% over 20 minutes. 25 fractions were collected. Fractions were screened for presence of EA and EB using LC-MS/MS methods as described below. Those which contained one or both erythrazoles were pooled for further isolation (fractions 16-19). The pooled fractions were purified by reverse phase HPLC (Phemonenex Luna, Phenyl-Hexyl, 250 x 10.0 mm, 2.5 mL/minute, 5 um, UV = 330 nm) using a gradient solvent system from 50% to 90% CH_3CN (0.1% formic acid) over 26 minutes to afford erythrazole A (t_R = 18.4 minutes) and erythrazole B (t_R = 20.0 minutes) (Figure 2.1).

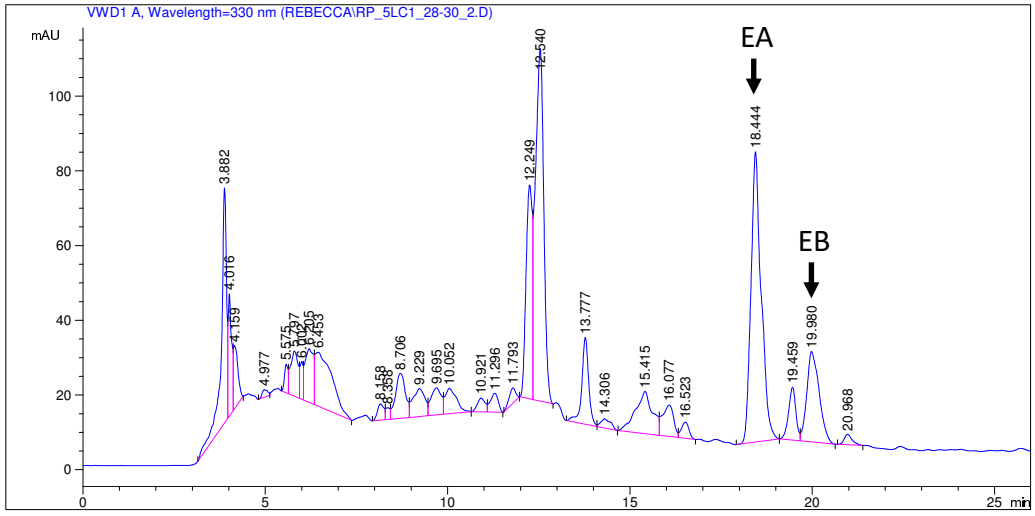


Figure 2.1 HPLC Chromatogram of SNB-035 extract to isolate erythrazole A and erythrazole B. Gradient of 50-90% ACN over 25 minutes yields EA

eluting at 18.44 minutes, EB at 19.98 minutes. Phenyl-hexyl column, wavelength = 330nm.

2.1.3 Identification of erythrazoles using LC-HRMS/MS

A Dionex UltiMate 3000 UHPLC paired with a Thermo Orbitrap Velos Pro LC-MS was used to detect the erythrazoles. Based on a method involving a mobile phase of 10% to 100% acetonitrile (ACN) in water containing 0.1% formic acid, retention times of 6.75 minutes for EA and 6.9 minutes for EB were identified and used for further detection of the molecules. A characteristic tandem mass spectrometry (MS/MS) spectrum was also identified (Figure 2.2), which showed fragments that include a loss of carboxylic acid, glycine, and acetylglycine. Maximum erythrazole detection was observed at day 8 of growth (Figure 2.3). We also used this method to determine peak erythrazole production by taking samples from cultures for extraction and analysis every 2 days from day 2-15.

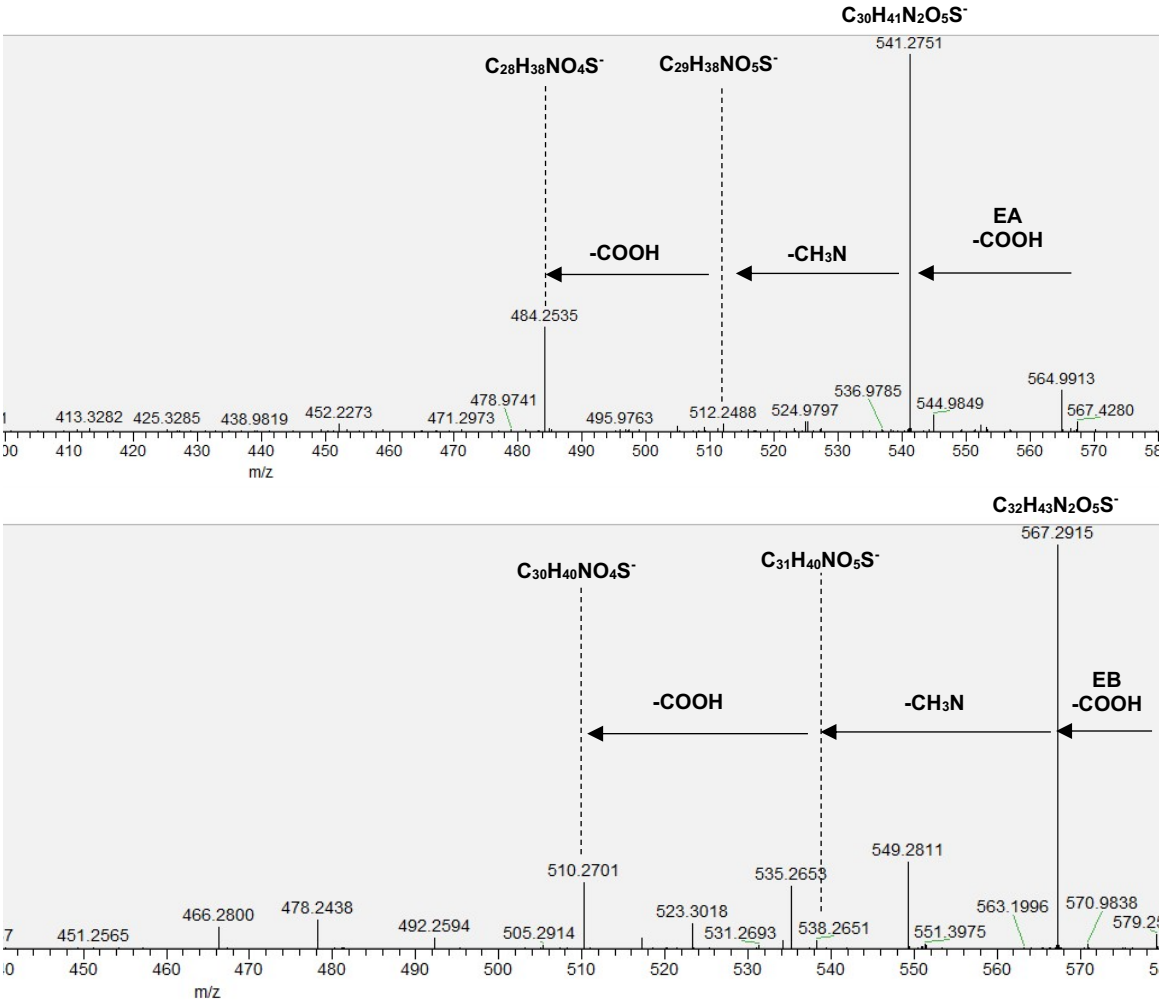


Figure 2.2 MS^2 fragmentation of erythrazoles A and B. Characteristic fragments are observed for both EA and EB including a loss of carboxylic acid (-COOH), glycine (-$C_2H_5NO_2$), and acetylglycine (-$C_3H_4NO_3$). Collision energy = 35.0V.

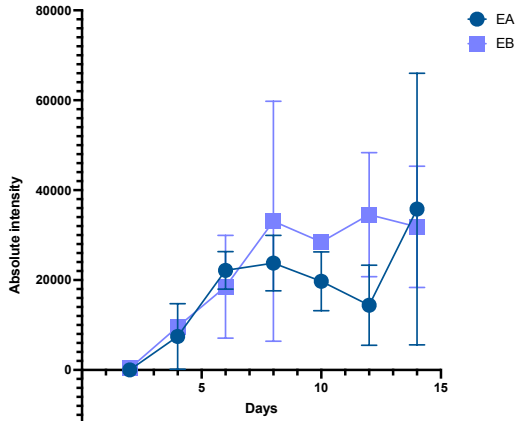


Figure 2.3 Time course study to determine peak erythrazole production through LC-MS analysis over 2 weeks of growth. 50mL cultures were inoculated at day 0. Every 2 days, 5mL was taken from the culture and extracted using liquid:liquid methods. These extracts were then dried and analyzed at 1mg/mL concentrations (n=2).

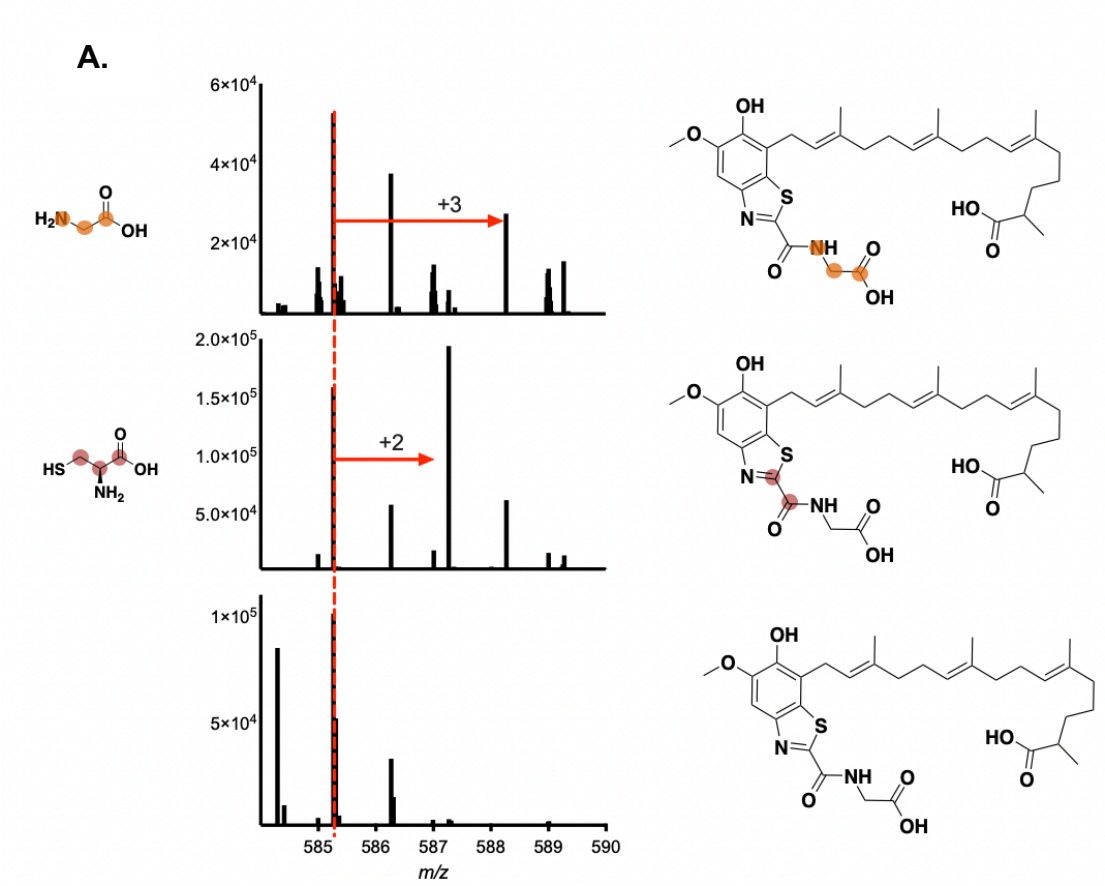
2.2 Stable isotope labeling of the amino acid component of the erythrazoles

In order to learn more about these unusual molecules, we employed a more traditional method of natural product biosynthesis probing; stable isotope labeling (SIL).

2.2.1 Cysteine, glycine and glutathione labeling of erythrazole A and B

On the C-terminal end of the erythrazoles are glycine and a moiety which resembles cysteine with a missing carbon. This led us to believe that these amino acids were both incorporated in the biosynthesis of the erythrazoles, and that a carbon cleavage occurs of cysteine to form the benzothiazole moiety. To

explore the incorporation of these amino acids, we first grew SNB-035 in liquid culture media with supplemented **L**-cysteine-^{13}C3 (1mM) or glycine-^{13}C2, ^{15}N (1mM) to confirm the amino acid incorporation of cysteine and glycine. Cultures were grown for 5 days and extracted with liquid:liquid partitioning (described above). Crude extracts were then analyzed by LC-HRMS (Figure 2.4). Interestingly, while both SIL precursors had 3 heavy labeled atoms, there was an observed shift in *m/z* of +3 in the glycine-^{13}C2, ^{15}N supplemented culture, while there was only a *m/z* shift of +2 in the culture with added **L**-cysteine-^{13}C3. This was observed in both EA and EB, and while the shifted *m/z* features had low relative abundance in some cases, the incorporation of labeled atoms was confirmed via MS/MS analysis. The MS/MS showed successful incorporation as the fragment features shared the same *m/z* as the unlabeled erythrazole when the enriched atoms were lost in fragmentation. Meanwhile, if the atoms were incorporated in a position that was not lost by fragmentation (i.e. the carbon in the thiazole), then we did see a shift in the fragment feature (Figure 2.5). This observation confirmed our hypotheses that cysteine and glycine are building blocks of the erythrazoles and that the benzothiazole formation involves the incorporation of cysteine, followed by a 1-carbon cleavage of cysteine.


A.
6×10^4
4×10^4
2×10^4
+3
2.0×10^5
1.5×10^5
1.0×10^5
5.0×10^4
+2
1×10^5
5×10^4
585
586
587
588
589
590
m/z
OH
O
S
N
NH
O
O
OH
HO
O

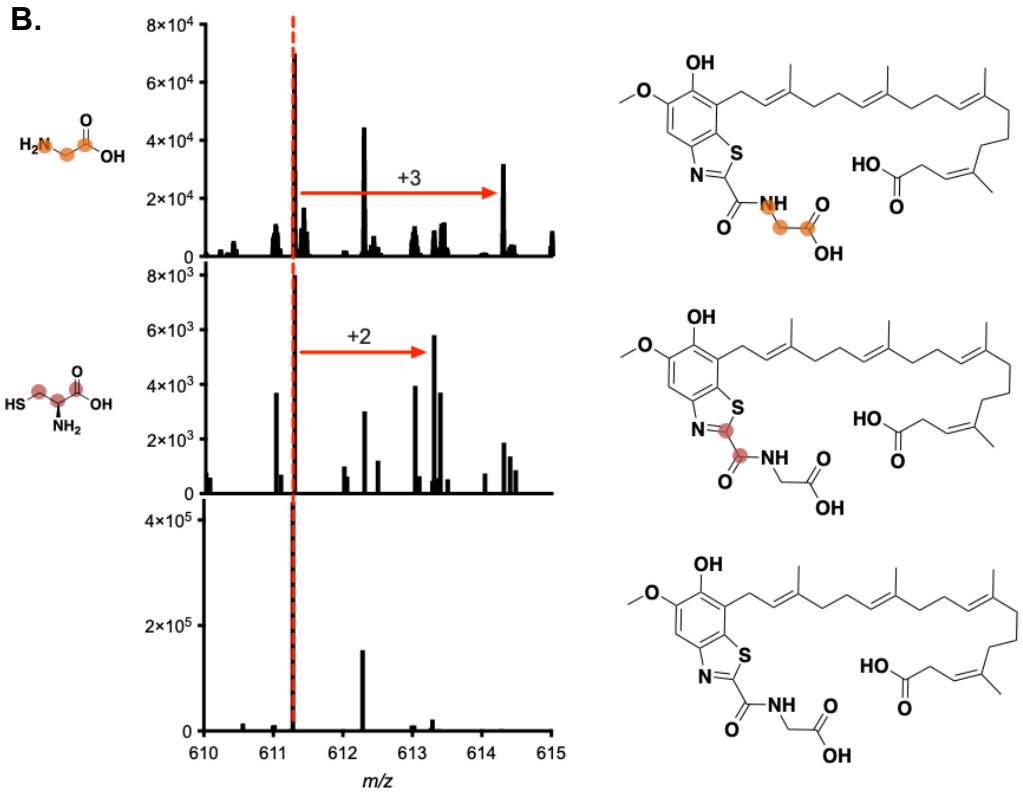
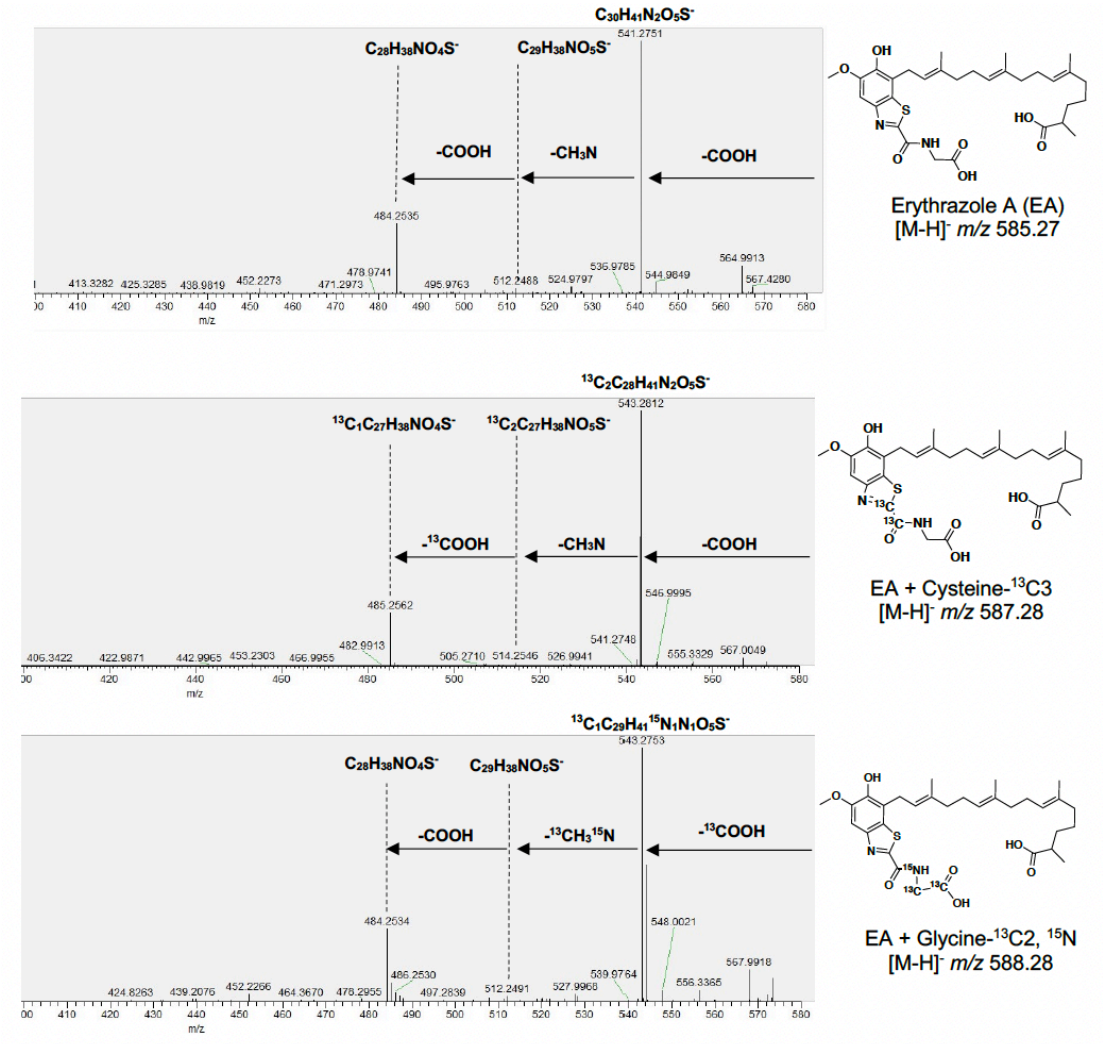


Figure 2.3 Stable isotope labeling using cysteine and glycine. A. Incorporation of L-cysteine $^{13}C3$ and glycine $^{13}C2$, ^{15}N into erythrazole A. **B.** Incorporation of L-cysteine $^{13}C3$ and glycine $^{13}C2$, ^{15}N into erythrazole B.

A.


C30H41N2O5S-
541.2751
C28H38NO4S-
C29H38NO5S-
-COOH
-CH3N
-COOH
484.2535
564.9913
536.9785
478.9741
544.9849
567.4280
413.3282 425.3285 438.9819 452.2273 471.2973 495.9763 512.2488 524.9797
m/z
Erythrazole A (EA)
[M-H]- m/z 585.27
13C2C28H41N2O5S-
543.2812
13C1C27H38NO4S-
13C2C27H38NO5S-
-13COOH
-CH3N
-COOH
485.2562
546.9995
482.9913
541.2748
555.3329
567.0049
406.3422 422.9871 442.9965 453.2303 466.9955 505.2710 514.2546 526.9941
m/z
EA + Cysteine-13C3
[M-H]- m/z 587.28
13C1C29H41 15N1N1O5S-
543.2753
C28H38NO4S-
C29H38NO5S-
-COOH
-13CH3 15N
-13COOH
484.2534
548.0021
567.9918
486.2530
539.9764
556.3365
424.8263 439.2076 452.2266 464.3670 478.2955 497.2839 512.2491 527.9968
m/z
EA + Glycine-13C2, 15N
[M-H]- m/z 588.28

B.

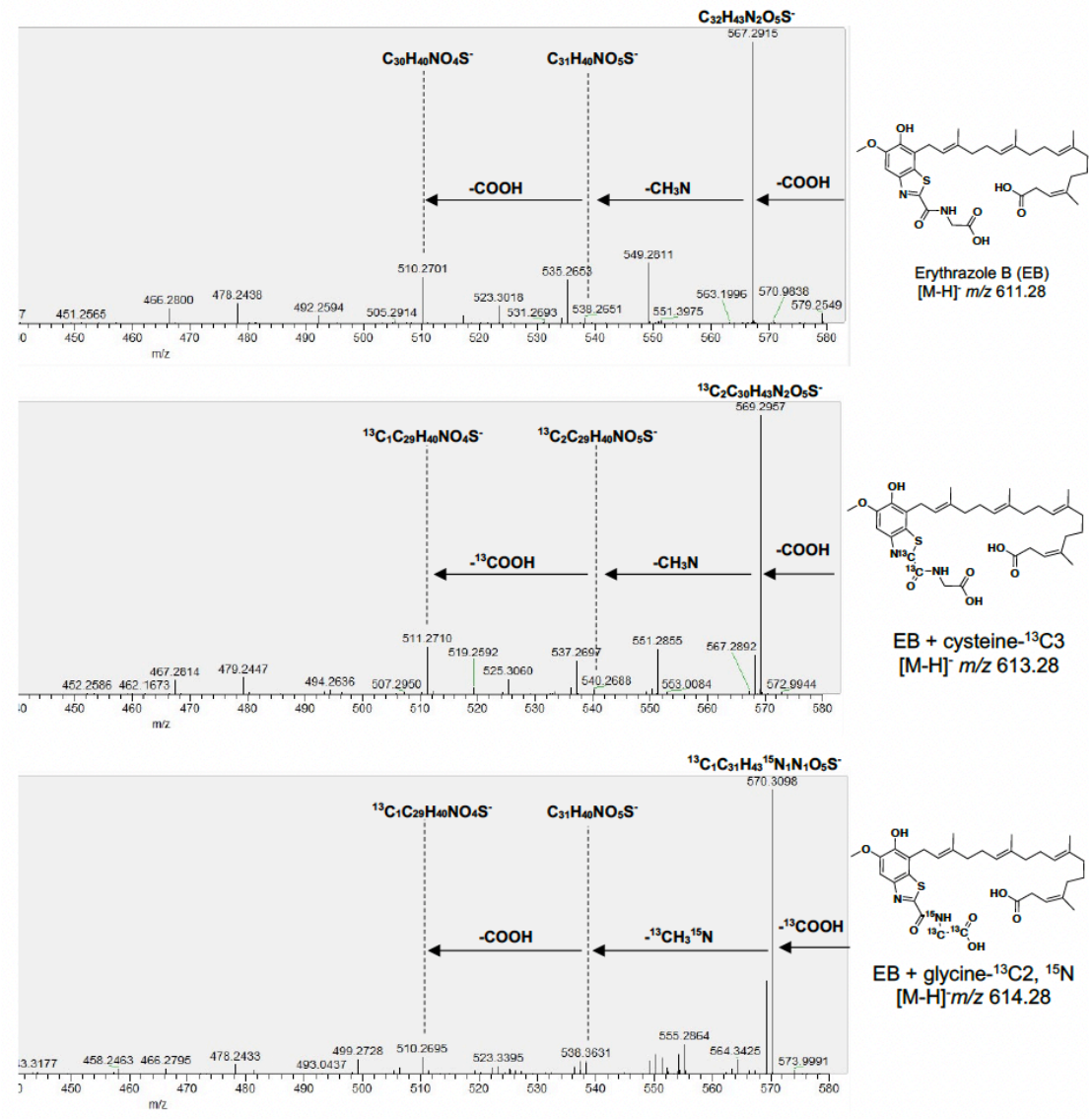


Figure 2.4 MS/MS fragmentation of the erythrazole A (A) and erythrazole B (B) with and without stable isotope labeling. Fragments showing a loss of carboxylic acid, glycine and acetylglycine were consistently observed with the described methods for LC-MS analysis. Collision energy = 35.0V.

Due to the hypothesized involvement of primary metabolism in the biosynthesis of EA/EB and the resemblance to glutathione, we suggested that perhaps the amino acid side chain of EA/EB could be derived from glutathione incorporation. Therefore, we cultured SNB-035 with either glycine-$^{13}C2$, ^{15}N or glutathione (glycine-$^{13}C2$, ^{15}N). The isotopically labeled glycine was fully incorporated into the C-terminus of the molecule when either glycine-$^{13}C2$ alone or ^{15}N or glutathione (glycine-$^{13}C2$, ^{15}N) was fed into the culture (Figure 2.6). For both EA and EB, glycine was incorporated with greater efficiency than glutathione (50% and 20% in EA, 50% and 7% in EB, respectively). It is unclear whether this indicates that glycine is better incorporated because it is in fact being incorporated as a single amino or because glutathione is bulkier and is therefore more difficult to incorporate. Similarly, we cannot rule out that the isotopically labeled glutathione may break down in the culture media, which may account for the labeling in the glutathione (glycine-$^{13}C2$, ^{15}N) fed culture. It would be useful to repeat this with fully isotopically labeled glutathione. However, this is not currently available to us.

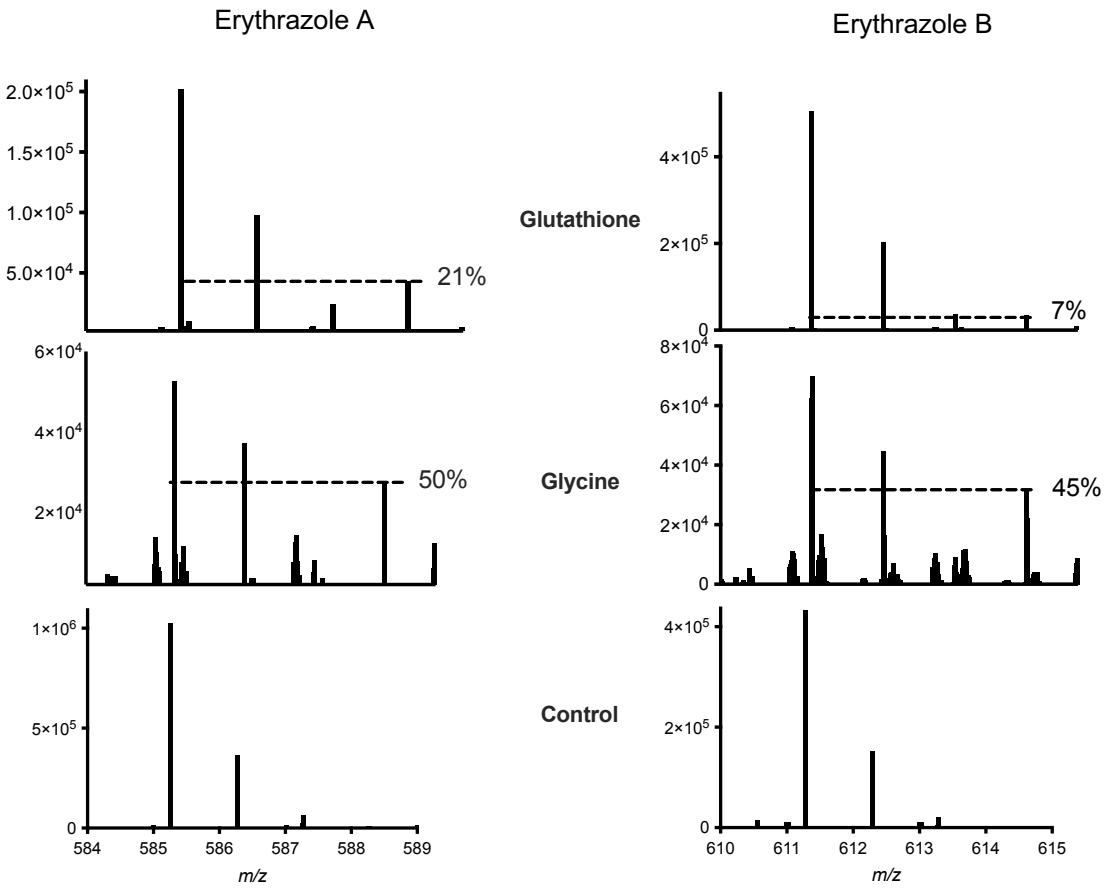
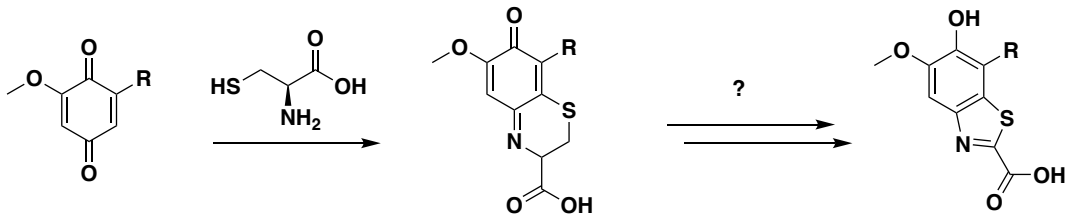


Figure 2.5 Glycine vs. glutathione labeling of erythrazole A and B. Percentages show the relative abundance of the labeled features compared to the unlabeled feature.

To further investigate the benzothiazole formation mechanism, we wanted to determine which carbon was being cleaved in the ring formation. A similar carbon cleavage has been observed in RiPP biosynthesis of 3-thiaglutamate, where a β-carbon cleavage of cysteine occurs by the enzymes, TglHI[2]. Therefore, we predicted that similarly, the observed carbon cleavage of cysteine occurs on the β-carbon. To probe this, we inoculated another culture of SNB-035 with only the β-carbon of cysteine labeled (L-cysteine (3-^{13}C)). As predicted, no incorporation of the ^{13}C was observed, confirming that the

benzothiazole is formed by a β-carbon cleavage (Figure 2.7). While this is not the first observation of this mechanism, the only other known occurrence is in a ribosomally synthesized and post-translationally modified (RiPP) molecule. Since EA/EB are not peptides and do not resemble RiPPs, it is likely that a different enzyme is at play here, albeit likely using a similar mechanism. The SNB-035 genome was also searched for enzymes with homology to TglHI but none were identified (Chapter 3.2).

A.


O
O
R
O
HS
OH
NH2
?
OH
S
N
Benzothiazine
Benzothiazole

B.

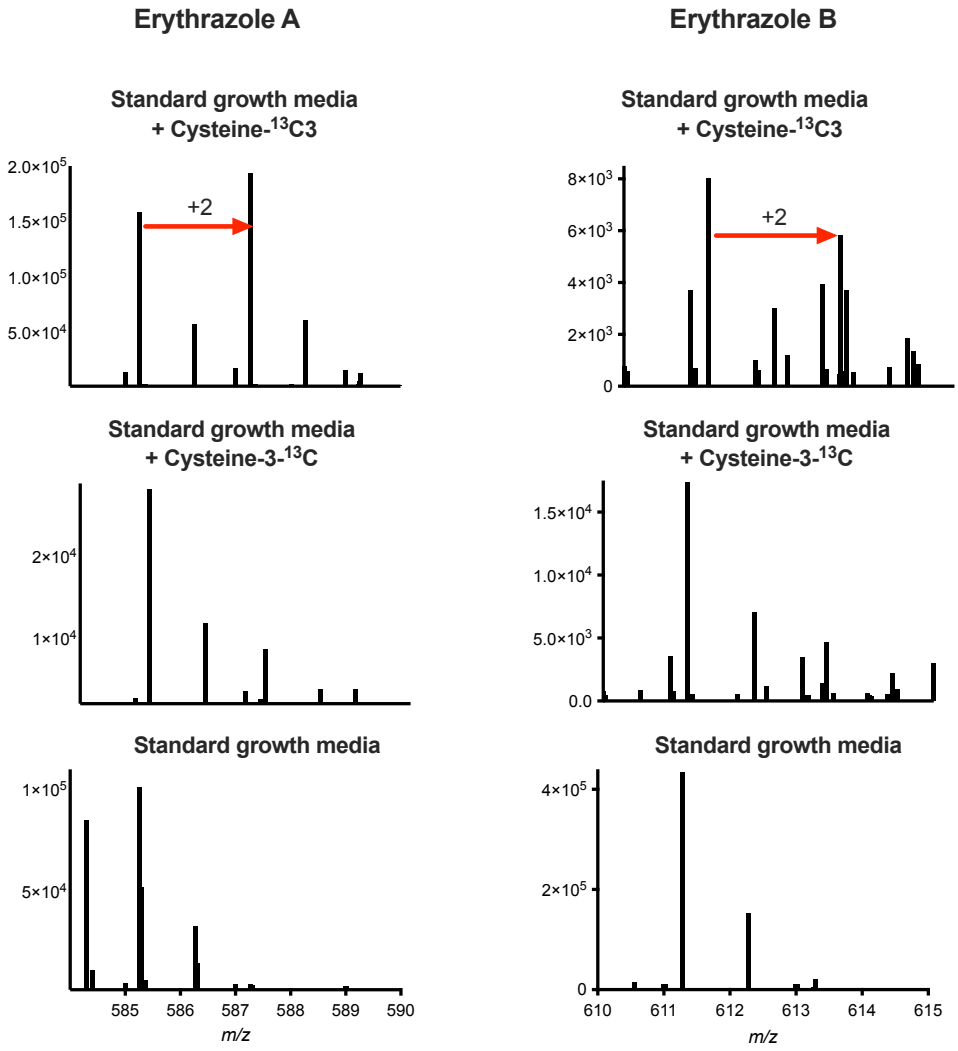

Erythrazole A
Erythrazole B
Standard growth media
+ Cysteine-13C3
+2
Standard growth media
+ Cysteine-3-13C
Standard growth media
m/z

Figure 2.6 ^{13}C-cysteine labeling of erythrazoles A-B reveals β-carbon cleavage in the formation of the benzothiazole moiety. a) Proposed mechanism of benzothiazole formation. b) LC-MS analysis of erythrazole A and erythrazole B from culture grown in standard growth media or with the addition of L-cysteine ^{13}C3 or L-cysteine 3-^{13}C.

2.3.2 Conversion between EA and EB

Another question which we wished to answer was if the terminal carboxylic acid on the terpene was derived by oxidative cleavage or by acetate extension. We reasoned that if oxidative cleavage was responsible, EB would be a precursor of EA. Contrastingly, if acetate extension occurs then EA would be a precursor of EB. To test this, we used the labeled erythrazoles as probes to look for interconversion. EA-glycine-^{13}C2, ^{15}N and EB-glycine-^{13}C2, ^{15}N were enriched from the crude extract of a 1 liter culture of SNB-035 supplemented with glycine-^{13}C2, ^{15}N. Isolation methods for these molecules have been previously described[1]. Separate cultures were started of SNB-035, which were then left to grow for 2 days at 30°C shaking at 300rpm. After 2 days of growth, they were supplemented with either EA-glycine-^{13}C2, ^{15}N or EB-glycine-^{13}C2, ^{15}N. Samples were taken from the culture and crude extract was obtained at days 3, 5, 7 and 11. These crudes were then analyzed with LC-MS/MS single ion monitoring (SIM) for the *m/z* of EA (585.26), EA-glycine-^{13}C2, ^{15}N (588.26), EB (611.28) and EB-glycine-^{13}C2, ^{15}N (614.28).

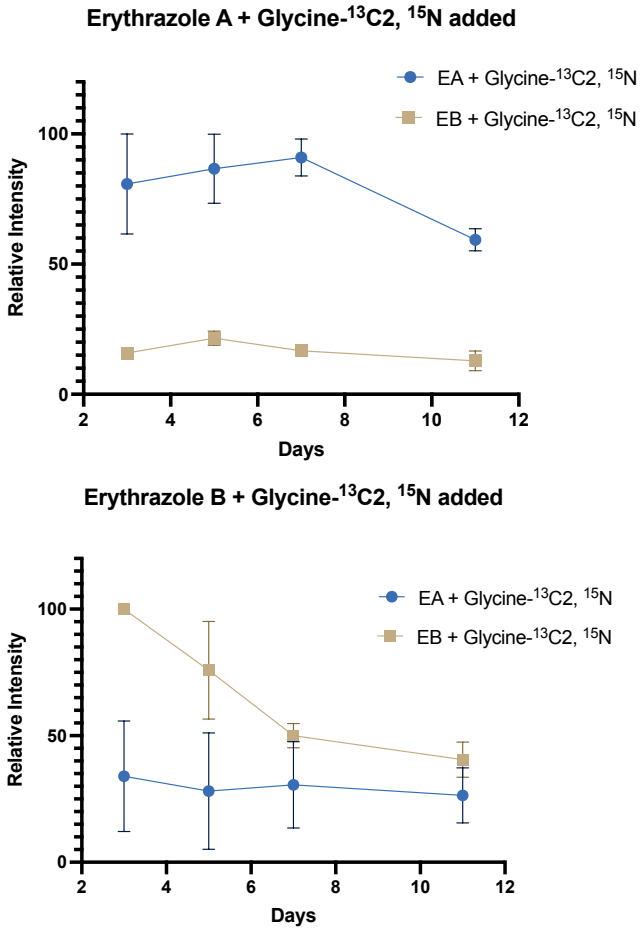


Figure 2.7 Results of conversion test to examine if interconversion occurs between EA and EB. Single ion monitoring (SIM) of the *m/z* corresponding with EA+3 and EB+3 in culture of SNB-035 supplemented EB-glycine-^{13}C2, ^{15}N on days 3, 5, 7 and 11 of growth (n=2). Labeled erythrazole was added on day 2. Results are normalized to relative intensity with the initial intensity of labeled erythrazole detected normalized to 100 and subsequent measurements as percentages of the initial intensity detected.

Surprisingly, the results indicated that in both cultures, the amount of supplemented erythrazole was decreasing over time but that the amount of the other erythrazole either stayed the same or slightly decreased (Figure 2.8). While it was expected to see the signal of the supplemented erythrazole

decrease in intensity over time, we expected to see a contrasting increase in the other labeled erythrazole's signal. This implies that conversion between the erythrazoles does not occur and that they are broken down or metabolized further.

We had hypothesized that one of the erythrazoles preceded the other, assuming that their biosynthesis was canonical. However, this data suggests that while oxidative cleavage likely occurs to form the terminal carboxylic acid, there is not an initial cleavage forming EB with a subsequent cleavage forming EA. Rather, they arise independently from the cleavage occurring at different positions. In order to determine the selectivity of the cleavage and whether it occurs before or after the amino acids are incorporated, we used the Global Natural Product Social Molecular Networking (GNPS)[3] to search for molecules similar to EA and EB but with varying isoprene chain lengths. MS/MS data of the erythrazoles share a characteristic fragmentation pattern (Figure 2.2), which we expected would identify similar molecules through networking if they were produced. While EA and EB clustered in a network, no molecules with a greater mass clustered with them or had similar MS/MS patterns. While there are larger features in the cluster (637.395 and 635.411), neither of these features shared the characteristic fragmentation pattern of the erythrazoles. Specifically, we looked for the acetylglycine fragment which we see consistently in the erythrazoles in relativey high abundance and would indicate the presence of the amino acid tail. The lack of similar molecules with a longer terpene chain

tells us that the oxidative cleavage occurs selectively and that the cleavage step of the biosynthesis likely precedes the amino acid incorporation.

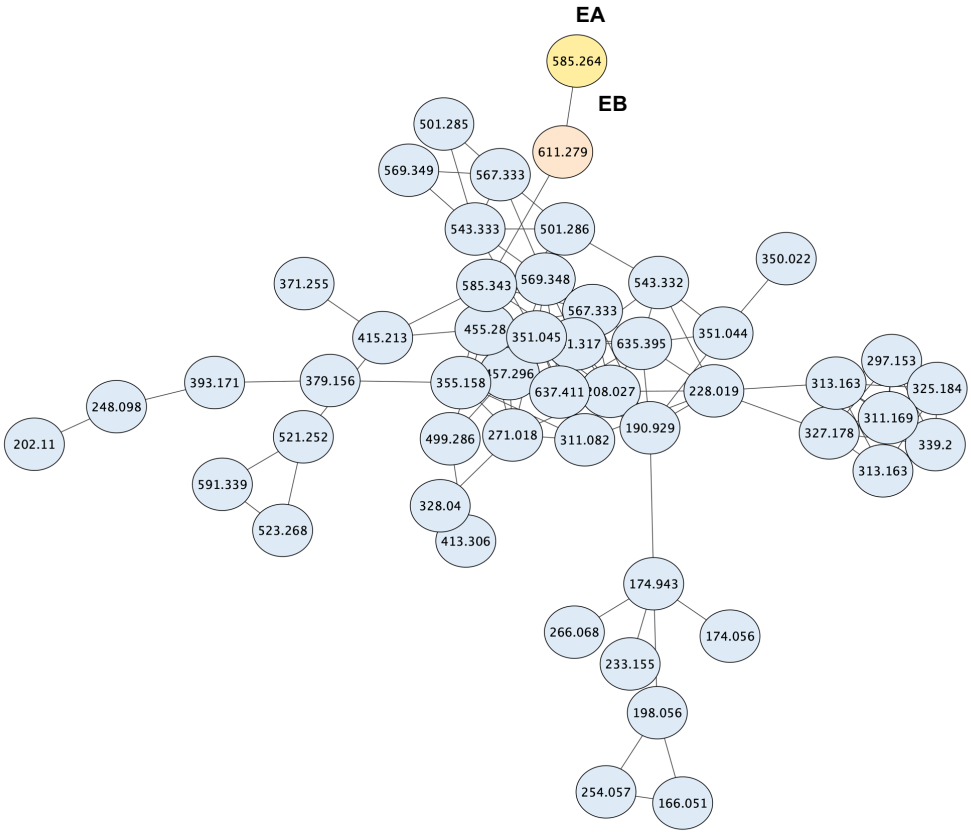


Figure 2.8 GNPS molecular networking of metabolites from crude extracts of SNB-025. EA and EB are highlighted. This network was constructed with the following conditions: Precursor Ion Mass Tolerance: 1 Da, Fragment Ion Mass Tolerance: 0.5 Da, Min Pairs Cos: 0.6, Minimum Matched Fragment Ions: 2, Minimum cluster size: 2.

We used this information to update our hypothetical biosynthesis (Figure 2.10). If our hypothesis is correct that the erythrazoles derive from an intermediate in the ubiquinone pathway, we suggest that the intermediate $DDMQ_{10}$ is used. Ubiquinone biosynthesis has been extensively characterized[4–6]. Following, $DDMQ_{10}$ in the ubiquinone biosynthesis is where

the structure begins to deviate from that of the erythrazoles. The first transformation to $DDMQ_{10}$ would then be the oxidative cleavage, resulting in an isoprene chain comprised of 20 (EA) or 22 (EB) carbons. Next, cysteine is incorporated followed by the β-carbon cleavage. Finally, glycine is attached. We also suppose that it is possible that the glycine and cysteine are added together, due to the inconclusiveness of the glutathione results (Figure 2.6). The experiments described here provide evidence for all of these steps, with the exception of the discrepancy between the amino acids being incorporated together or separately.

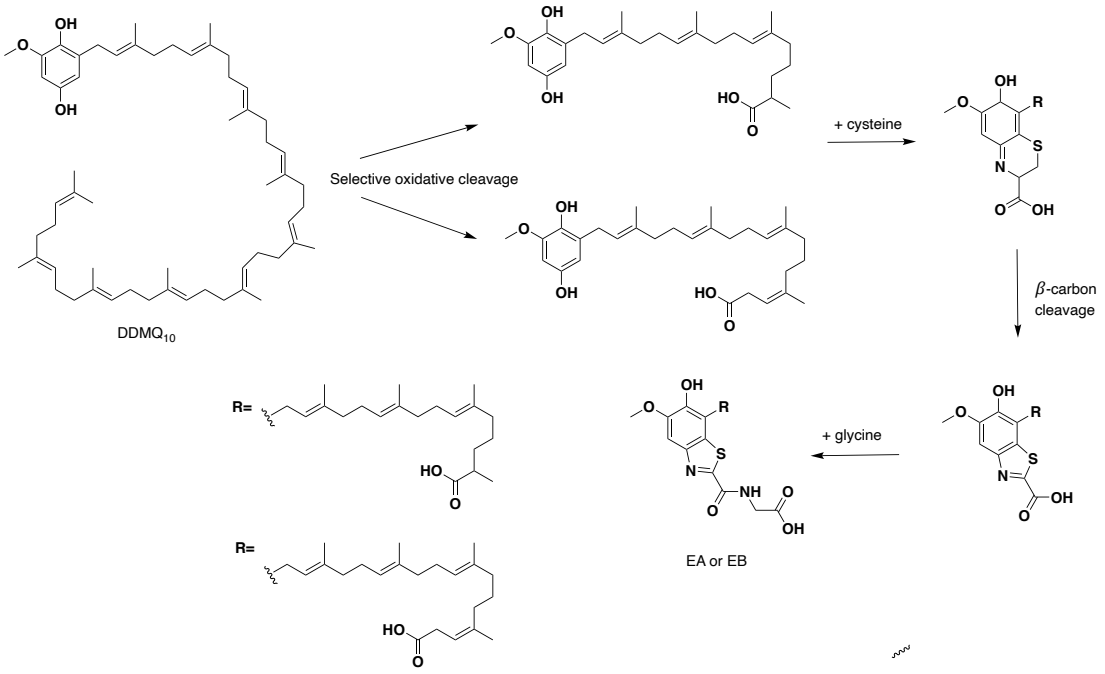


Figure 2.9 Proposed biosynthesis of erythrazoles A and B with feeding study results. $DDMQ^{10}$ is an intermediate from the ubiquinone pathway which was previously characterized[4,7,8]. Selective oxidative cleavage results in an isoprene chain comprised of 20 (EA) or 22 (EB) carbons. Next, SIL results show

that a cysteine is incorporated onto the quinol, which subsequently undergoes a β-carbon cleavage. Finally, glycine is incorporated.

2.3 Attempts of isoprenoid labeling

There are two potential pathways for the biosynthesis of terpenes in bacteria, the mevalonate pathway (MVA) and the methylerythritol phosphate pathway (MEP). These pathways have shown distinct labeling patterns by the incorporation of acetate[9–11] and glucose into isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP), the building blocks of terpenes. While we hypothesized that the MEP pathway is involved in erythrazole biosynthesis due to the presence of MEP genes in the SNB-035 genome, we wanted to confirm this by labeling the isoprene units.

2.3.1 Unsuccessful attempts of isoprenoid labeling using labeled Na-acetate

Figure 2.11 shows the anticipated labeling of DMAPP and the expected labeling patterns for the erythrazoles with sodium acetate. Overall, if the MVA pathway was utilized in erythrazole biosynthesis, Na-acetate (1-^{13}C) incorporation would should a *m/z* shift of +8 for erythrazole A and +9 for erythrazole B. Na-acetate (2-^{13}C) would show a *m/z* shift of 4 for eryrthazole A and 5 for erythrazole B. Na-acetate (1,2-^{13}C2) would show a *m/z* shift of +12 for erythrazole A and +14 for erythrazole B. Contrastingly, the MEP pathway would show much less incorporation of acetate. Na-acetate (1-^{13}C)

incorporation would should a *m/z* shift of +4 for erythrazole A and +4 for erythrazole B. Na-acetate (2-^{13}C) would not label the erythrazoles if the MEP pathway was used, therefore the Na-acetate (1,2-^{13}C2) would show the same labeling as the Na-acetate (1-^{13}C). Of note, these predictions are based on the conclusion that oxidative cleavage is forming the terminal carboxylic acid rather than an acetate extension. We also are assuming that the shikimate pathway is forming the substituted benzene. However, if the benzene was derived from a polyketide synthase (PKS) rather than the shikimate pathway as we predicted, this could account for 3 more labeled ^{13}C's to be incorporated. Regardless, the labeling patterns differ significantly enough for the pathway to be easily identified through acetate labeling.

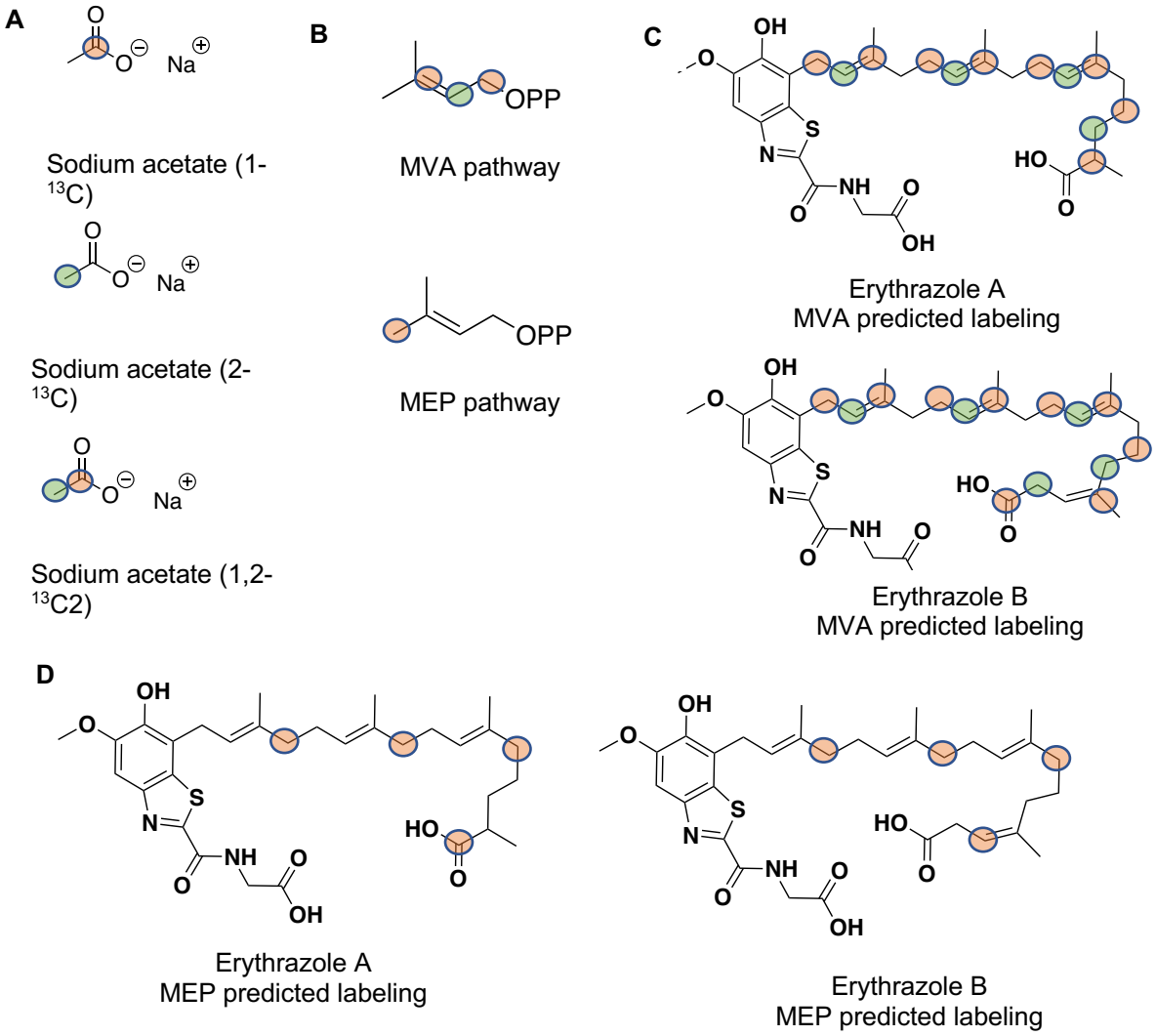


Figure 2.10 Anticipated labeling patterns of the erythrazoles with differently labeled sodium-acetate.

To carry out this experiment, I grew 1L of SNB-035 with 500mg/L of either Na-acetate (1-^{13}C), Na-acetate (2-^{13}C), or Na-acetate (1,2-^{13}C2). Labeled compounds were dissolved in water and sterile filtered before being added to the cultures. Cultures were allowed to grow for 5 days before liquid:liquid extraction with ethyl acetate. These crudes were then analyzed

using previously described methods with LC-HRMS. The results of the *m/z* shifts in these extracts are found in the table below:

	EA	**EB**
Control	22, 28, 30	16, 18, 28, 30
Unlabeled acetate	12, 28, 30	12, 14, 22, 28, 30
^{13}C1	14	11, 20
^{13}C2	4, 5, 6, 7	4, 5, 6
^{13}C1^{13}C2	16	4, 29

Table 2.1 Results from feeding studies with differently labeled Na-acetate. *M/z* shifts that were seen in the controls or unlabeled acetate were excluded from those listed in the C1, C2, and C1C2 results.

Since the extracts being analyzed were crudes, purification was necessary to be sure that these *m/z* shifts were labeled erythrazoles rather than separate compounds. Yield of erythrazoles from these cultures was very low (<1mg). However, we expected to be able to detect ^{13}C incorporation using NMR and LC-HRMS. Figure 2.12 shows the HR-LC-MS results from purified erythrazoles. After purification, we saw shifts of 18 and 22 for erythrazole B and 22 for erythrazole A. These shifts were much higher than we anticipated and we considered the possibility that they may be adducts. To check this, we used ^{13}C NMR (800 MHz, MeOD, 10,000 scans). While we did not expect to see unlabeled signals due to the small amount of material and low sensitivity of the

^{13}C NMR, the ^{13}C enriched signals should be detectable even with a small sample size. However, no signals were observed confirming that the labeling was unsuccessful. This was a surprising result as both isoprenoid biosynthetic pathways should incorporate acetate and therefore show labeling. One hypothesis for why this occurred is that there could be issues with competition from other carbon sources in the media, causing the bacteria to use these rather than the labeled Na-acetate.

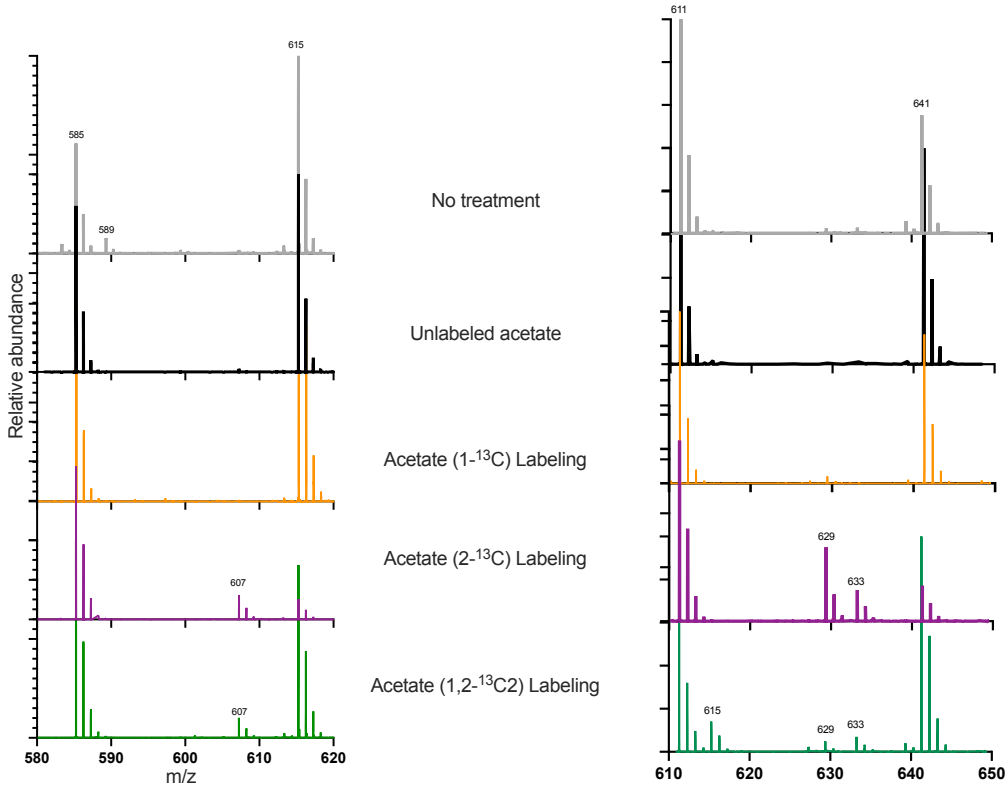
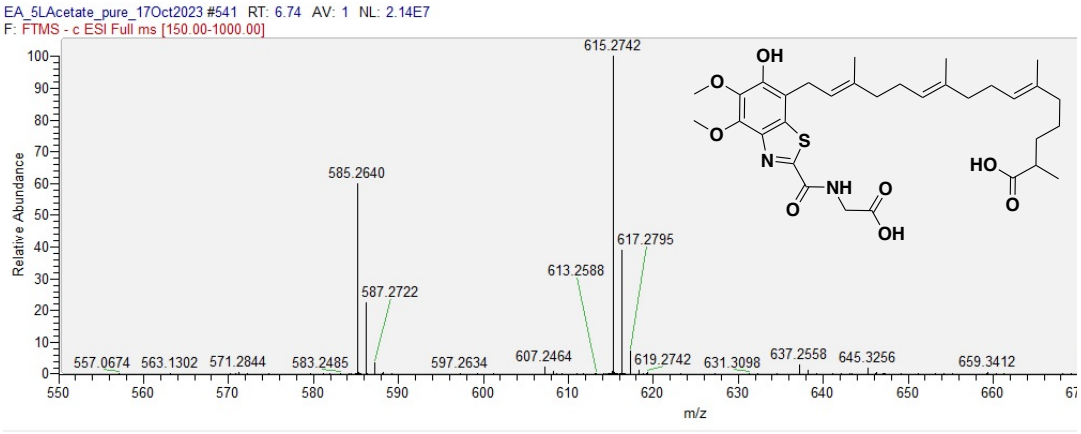
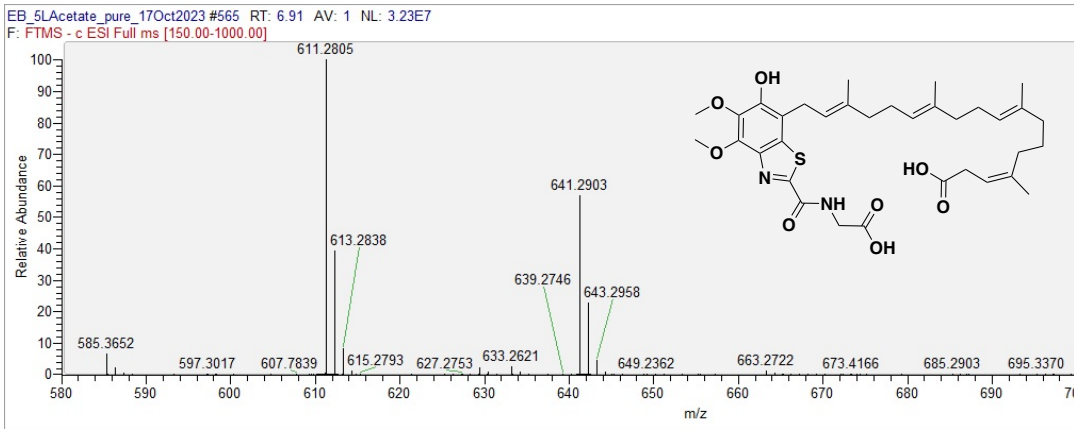


Figure 2.11 SIL of sodium acetate feeding study results.

While the labeling was unsuccessful, we did learn something interesting from these isolations which is that we consistently saw a *m/z* shift of +30 Da. This change in mass is consistent with that of a methoxy group, which is interesting because ubiquinone has an additional methoxy group on the benzene ring. The consistency of this mass appearing in the semi-isolated compounds supports that it must have very similar composure to the erythrazoles. Additionally, the MS/MS fragmentation patterns are consistent with that of the erythrazoles as is shown below in EA. This finding suggests that the erythrazoles may be even more closely tied to ubiquinone than we originally thought. The additional methoxy group suggests that erythrazole biosynthesis may deviate at a later point in the ubiquinone pathway. This also suggests the promiscuity of the enzymes which alter ubiquinone as they are able to add amino acids and perform the carbon cleavage of cysteine on both the singly o-methylated starter and doubly o-methylated started.

A.





B.

Figure 2.12 LC-HRMS of isolated erythrazoles indicates presence of erythrazoles with additional methoxy group. A. Both molecules show a feature 30 greater than the corresponding erythrazole *m/z*. This *m/z* aligns with erythrazoles with an additional methoxy group at the 5 position. **B.** MS^2 fragmentation of erythrazole A and the MS feature with a 30 Da higher *m/z*.

2.3.2 Unsuccessful attempts of isoprenoid labeling using D-glucose (1-^{13}C)

Literature has also shown successful labeling of terpenes using D-glucose[12–15]. Therefore, we attempted to label the isoprene units by feeding the culture of SNB-035 with D-glucose (1-^{13}C). For these experiments, 0.05M glucose was added at the time which the culture was inoculated and left to grow for 7 days. We had concerns about sugars in the media interfering with incorporation of the labeled glucose, which is why it was added immediately in this case. Expected labeling of isoprene units is seen in Figure 2.14. Expected erythrazole labeling would therefore have *m/z* shifts of +12 for both EA and EB for the MVA pathway and +8 or +9 respectively for the MEP pathway.

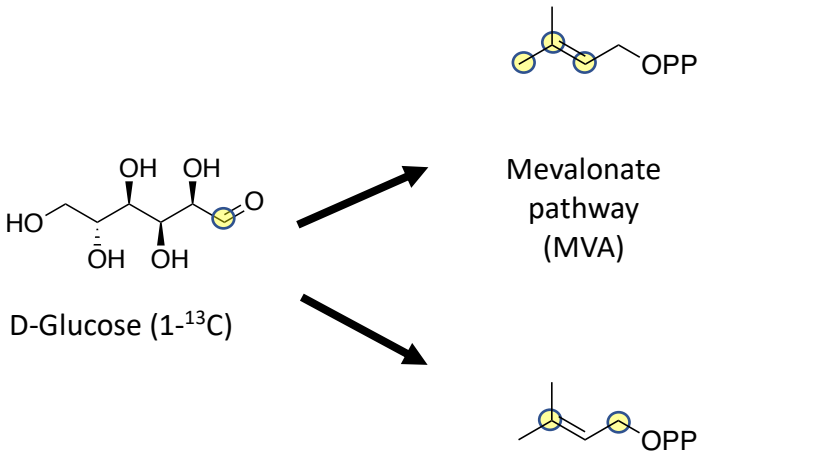


Figure 2.13 Expected labeling patterns of DMAPP using D-glucose (1-^{13}C).

Initial LC-MS analysis of the labeled crudes showed shifts of +14 for EA and +12, +16 for EB. After partial purification with flash column chromatography, shifts of +4, 14 were observed for EA and +5, +8, +18 for EB. However, these masses were also seen in controls where unlabeled glucose was added to the cultures. We concluded that no labeling was occurring in our regular seawater based media, A1BFe+C. We then supposed that growing the cultures in a minimal media may be more successful as there would be less interference from background sugars in the media. First, I kept the components of the media the same as our regular recipe but with 10% of the usual starch. These results were similar to what we observed in the regular media. Next, I tried M9 minimal medium[16] which I did not see any growth of SNB-035 in. Next, I tried using a TCG medium (3g tryptone, 5g casitone, 4g glucose, per liter of seawater). The microbe was able to grow in this media but erythrazole production was nearly depleted (Figure 2.15).

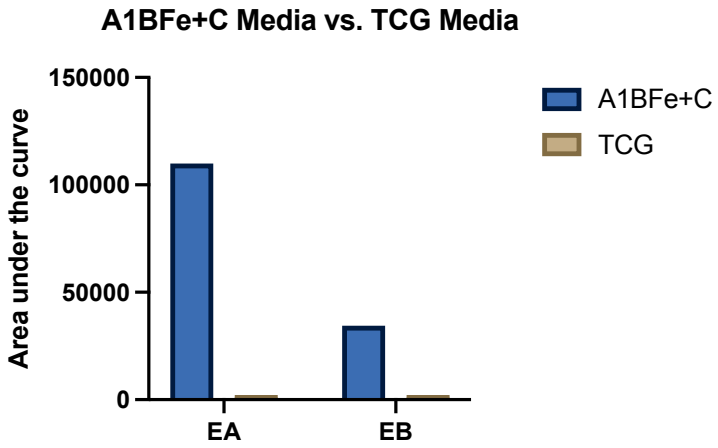


Figure 2.14 Area under the curve (AUC) LC-HRMS analysis of erythrazole A and B when cultured in either A1BFe+C or TCG media.

2.3.3 Media optimization for stable isotope labeling of erythrazoles

With these unsuccessful attempts, I resorted to trying to grow SNB-035 with lowered amounts of our A1BFe+C premix, which is comprised of 10g starch, 4g yeast extract, 2g peptone, 1g $CaCO_3$. Typically, we add 0.85g of this mix to a 50mL culture so I tried adding 0.4g and 0.2g to lower the amount of background components in the media. First, I tested the difference between lowering the media mixture with and without added unlabeled sodium acetate. I reasoned that the nutrient rich background in the media may also have been the cause for the lack of acetate labeling. Therefore, I chose to use sodium acetate rather than glucose because the labeled material is less expensive. Lowering the media nutrients to 0.2g/50mL culture (~23% of usual nutrients) and then supplementing with acetate caused an 8.1x or 29.9x increase in production of erythrazole A and B, respectively. When we lowered the media nutrients to 0.4g/50 mL (~47% of usual nutrients), we saw a 3.0x and 6.0x increase of erythrazole production with supplemented Na-acetate (Figure 2.16).

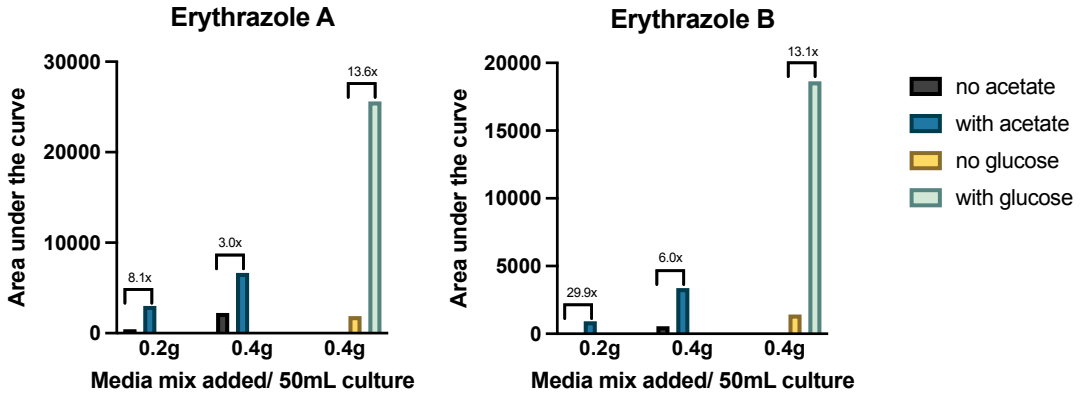


Figure 2.15 Low nutrient media with and without supplemented acetate or glucose.

This data supports that supplementing with labeled acetate should be incorporated into the erythrazoles. Though the 0.4g/50mL media showed less of a difference with or without acetate supplementation, the bacteria seemed to grow better overall in this media and overall erythrazole production was ~2.5x higher. Therefore, I chose to go forward with these experiments using the 0.4g of A1BFe+C premix/50mL media. Additionally, I repeated this experiment with and without glucose supplementation. Production of erythrazole A showed a 13.6x increase with supplemented glucose. Erythrazole B showed a 13.1x increase with glucose supplementation (Figure 2.16). This indicates that labeling should also occur with these conditions with glucose labeling, perhaps with even greater efficiency than with acetate. Next, I grew 2 cultures of SNB-035 in 0.4g/50mL media. One culture was supplemented with unlabeled Na-acetate and the other with Na-acetate (1-^{13}C). Na-acetate (1-^{13}C) was selected as it should label both MAV and MEP

pathways (Figure 2.14). LC-HRMS from these extracts indicated a *m/z* shift of +14 for both EA and EB. While this shift seemed higher than anticipated, we went forward with purification but again saw no signals with the ^{13}C NMR. Ultimately, we decided that these efforts were a lower priority and to focus on the amino acid installation and manipulation aspect of the erythrazoles.

2.4 Discussion

Together, the results from stable isotope labeling of the amino acids were able to reveal major insights into erythrazole biosynthesis. First, that the benzothiazole formation occurs from a β-carbon cleavage of cysteine, and second that a selective oxidative cleavage occurs to form the terminal carboxylic acid of the molecules. To our knowledge, there has only been one other example of a β-carbon cleavage of cysteine in a natural product biosynthesis[2]. Additionally, this is the first time where this has been seen in a benzothiazole formation. Benzothiazoles are important moieties which have a variety of bioactivity[17,18]. While several syntheses have been reported[19–24], the only known naturally occurring mechanisms of benzothiazole formation in nature are a decarboxylation of cysteine in firefly luciferin[25] and iron-assisted oxidation[26]. Our discovery of a benzothiazole being formed through a β-carbon cleavage of cysteine shows a novel mechanism that differs than those previously described. The discovery of the enzyme responsible for this transformation would open a new understanding of benzothiazole biosynthesis,

could be useful for new methods of biocatalysis to form benzothiazoles in chemical synthesis, and could identify new, naturally occurring benzothiazoles through genome mining. While other methods of benzothiazole synthesis are currently available[19–24], enzymatic reactions for chemical syntheses can be preferable as they are often site-specific, stereospecific, and often more environmentally friendly[27,28]. Future directions should work to elucidate the identity of the enzyme which is forming the benzothiazole of the erythrazoles.

The selective oxidative cleavage of the terpene chain to form the terminal carboxylic acid of the erythrazoles was supported by several experiments described here. First, the fact that the erythrazoles do not interconvert suggests that they arise separately rather than erythrazole A undergoing an acetate extension. This was also supported by the lack of labeling by sodium acetate. The site selectivity of the cleavage was supported by GNPS molecular networking, which only showed one, smaller molecule with the characteristic acetylglycine fragmentation of the erythrazoles. However, the *m/z* of this molecule is 328.04, which is not large enough for even a single isoprene unit to be attached to the rest of the scaffold. Because the only molecules which were larger than erythrazole B did not share the characteristic MS/MS fragmentation pattern as the erythrazoles, we can imply that the oxidative cleavage of the isoprene chain only occurs at the positions which produce EA and EB. From this data we also can imply that the oxidative cleavage of the isoprene chain occurs before the amino acids are incorporated,

as we do not see the characteristic glycine and acetylglycine fragmentation on any of the related molecules. Therefore, while these experiments indicated the presence of yet another interesting enzyme which is capable of selective oxidative cleavage of a terpene chain, they also gave us new information about the order of erythrazole biosynthesis.

While the labeling of the terpene chain was unsuccessful, the success of the amino acid labeling was still able to provide several new insights regarding erythrazole biosynthesis. Because the genome of SNB-035 contains genes from the MEP pathway rather than the MVA pathway, we can assume that the MEP pathway is responsible for biosynthesizing the isoprene units which are found in the erythrazoles. Future work could investigate this further with more media optimization or perhaps leaving the cultures growing for longer. Regardless of which pathway synthesizes the isoprene units, their incorporation into the erythrazoles would require an isoprenyl diphosphate synthase (IDS) to condense the isoprene units and a prenyltransferase to transfer them onto the rest of the molecule. The work described in chapter 4 seeks to identify these enzymes.

2.6 Chapter two references

(1) Hu, Y.; MacMillan, J. B. Erythrazoles A–B, Cytotoxic Benzothiazoles from a Marine-Derived *Erythrobacter* Sp. *Org. Lett.* **2011**, *13* (24), 6580–6583. https://doi.org/10.1021/ol202944g.

(2) Ting, C.P; Funk, M. A.; Halaby, S. L.; Zhang, Z.; Gonen, T.; van der Donk, W. A. Use of a Scaffold Peptide in Thebiosynthesis of Amino Acid–Derivednatural Products. *Science* **2019**, *365* (6450), 280–284.

(3) Wang, M. et al.; Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34* (8), 828–837. https://doi.org/10.1038/nbt.3597.

(4) Meganathan, R. Ubiquinone Biosynthesis in Microorganisms. *FEMS Microbiol. Lett.* **2001**, *203* (2), 131–139. https://doi.org/10.1111/j.1574-6968.2001.tb10831.x.

(5) Abby, S. S.; Kazemzadeh, K.; Vragniau, C.; Pelosi, L.; Pierrel, F. Advances in Bacterial Pathways for the Biosynthesis of Ubiquinone. *Biochim. Biophys. Acta BBA - Bioenerg.* **2020**, *1861* (11), 148259. https://doi.org/10.1016/j.bbabio.2020.148259.

(6) Aussel, L.; Pierrel, F.; Loiseau, L.; Lombard, M.; Fontecave, M.; Barras, F. Biosynthesis and Physiology of Coenzyme Q in Bacteria. *Biochim. Biophys. Acta BBA - Bioenerg.* **2014**, *1837* (7), 1004–1011. https://doi.org/10.1016/j.bbabio.2014.01.015.

(7) Abby, S. S.; Kazemzadeh, K.; Vragniau, C.; Pelosi, L.; Pierrel, F. Advances in Bacterial Pathways for the Biosynthesis of Ubiquinone. *Biochima Biophys. Acta BBA - Bioenerg.* **2020**, *1861* (11).

(8) Aussel, L.; Pierrel, F.; Loiseau, L.; Lombard, M.; Fontecave, M.; Barras, F. Biosynthesis and Physiology of Coenzyme Q in Bacteria. *18th Eur. Bioenerg. Conf. 2014 Lisbon Port.* **2014**, *1837* (7), 1004–1011. https://doi.org/10.1016/j.bbabio.2014.01.015.

(9) Rohmer, M.; Knani, M.; Simonin, P.; Sutter, B.; Sahm, H. Isoprenoid Biosynthesis in Bacteria: A Novel Pathway for the Early Steps Leading to Isopentenyl Diphosphate. *Biochem. J.* **1993**, *295* (2), 517–524. https://doi.org/10.1042/bj2950517.

(10) Zhou, D.; White, R. H. Early Steps of Isoprenoid Biosynthesis in Escherichia Coli. *Biochem. J.* **1991**, *273* (3), 627–634. https://doi.org/10.1042/bj2730627.

(11) Rohmer, M. Diversity in Isoprene Unit Biosynthesis: The Methylerythritol Phosphate Pathway in Bacteria and Plastids. *Pure Appl. Chem.* **2007**, *79* (4), 739–751. https://doi.org/doi:10.1351/pac200779040739.

(12) Frank, A.; Groll, M. The Methylerythritol Phosphate Pathway to Isoprenoids. *Chem. Rev.* **2017**, *117* (8), 5675–5703. https://doi.org/10.1021/acs.chemrev.6b00537.

(13) Orsi, E.; Beekwilder, J.; Peek, S.; Eggink, G.; Kengen, S. W. M.; Weusthuis, R. A. Metabolic Flux Ratio Analysis by Parallel 13C Labeling of Isoprenoid Biosynthesis in Rhodobacter Sphaeroides. *Metab. Eng.* **2020**, *57*, 228–238. https://doi.org/10.1016/j.ymben.2019.12.004.

(14) Rohmer, M.; Sutter, B.; Sahm, H. Bacterial Sterol Surrogates. Biosynthesis of the Side-Chain of Bacteriohopanetetrol and of a Carbocyclic Pseudopentose from 13C-Labelled Glucose in Zymomonas Mobilis. *J Chem Soc Chem Commun* **1989**, No. 19, 1471–1472. https://doi.org/10.1039/C39890001471.

(15) Rohmer, M. Mevalonate-Independent Methylerythritol Phosphate Pathway for Isoprenoid Biosynthesis. Elucidation and Distribution. *Pure Appl. Chem.* **2003**, *75* (2–3), 375–388. https://doi.org/doi:10.1351/pac200375020375.

(16) Miller, J. H. M9 Minimal Medium (Standard). *Cold Spring Harb. Protoc.* **2010**.

(17) Irfan, A.; Batool, F.; Zahra Naqvi, S. A.; Islam, A.; Osman, S. M.; Nocentini, A.; Alissa, S. A.; Supuran, C. T. Benzothiazole Derivatives as Anticancer Agents. *J. Enzyme Inhib. Med. Chem.* **2020**, *35* (1), 265–279. https://doi.org/10.1080/14756366.2019.1698036.

(18) Keri, Rangappa S.; Patil, M.; Patil, S.; Budagumpi, S. A Comprehensive Review in Current Developments of Benzothiazole-Based Molecules in Medicinal Chemistry. *Eur. J. Med. Chem.* **2015**, *89*, 207–251.

(19) Bahrami, K.; Khodaei, M. M.; Naali, F. Mild and Highly Efficient Method for the Synthesis of 2-Arylbenzimidazoles and 2-Arylbenzothiazoles. *J. Org. Chem.* **2008**, *73* (17), 6835–6837. https://doi.org/10.1021/jo8010232.

(20) Fan, X.; He, Y.; Wang, Y.; Xue, Z.; Zhang, X.; Wang, J. A Novel and Practical Synthesis of 2-Benzoylbenzothiazoles and 2-Benzylbenzothiazoles. *Tetrahedron Lett.* **2011**, *52* (8), 899–902.

(21) Seijas, J. A.; Vázquez-Tato, M. P.; Carballido-Reboredo, M. R.; Crecente-Campo, J.; Romar-Lopez, L. Lawesson's Reagent and Microwaves: A New Efficient Access to Benzoxazoles and Benzothiazoles from Carboxylic Acids under Solvent-Free Conditions. *Synlett* **2007**, *2007* (02), 0313–0317.

(22) Sun, Y.; Jiang, H.; Wu, W.; Zeng, W.; Wu, X. Copper-Catalyzed Synthesis of Substituted Benzothiazoles via Condensation of 2-Aminobenzenethiols with Nitriles. *Org. Lett.* **2013**, *15* (7), 1598–1601.

(23) Xiao, R.; Hao, W.; Ai, J.; Cai, M.-Z. A Practical Synthesis of 2-Aminobenzothiazoles via the Tandem Reactions of 2-Haloanilines with Isothiocyanates Catalyzed by Immobilization of Copper in MCM-41. *J. Organomet. Chem.* **2012**, *705*, 44–50.

(24) Gao, X.; Liu, J.; Zuo, X.; Feng, X.; Gao, Y. Recent Advances in Synthesis of Benzothiazole Compounds Related to Green Chemistry. *Molecules* **2020**, *25* (7). https://doi.org/10.3390/molecules25071675.

(25) Oba, Y.; Yoshida, N.; Kanie, S.; Ojika, M.; Inouye, S. Biosynthesis of Firefly Luciferin in Adult Lantern: Decarboxylation of L-Cysteine Is a Key Step for Benzothiazole Ring Formation in Firefly Luciferin Synthesis. *PloS One* **2013**, *8* (12), e84023. https://doi.org/10.1371/journal.pone.0084023.

(26) Di Donato, P.; Napolitano, A.; Prota, G. Metal Ions as Potential Regulatory Factors in the Biosynthesis of Red Hair Pigments: A New Benzothiazole Intermediate in the Iron or Copper Assisted Oxidation of 5-S-Cysteinyldopa. *Biochim. Biophys. Acta BBA-Gen. Subj.* **2002**, *1571* (2), 157–166.

(27) Chakrabarty, S.; Romero, E. O.; Pyser, J. B.; Yazarians, J. A.; Narayan, A. R. H. Chemoenzymatic Total Synthesis of Natural Products. *Acc. Chem. Res.* **2021**, *54* (6), 1374–1384. https://doi.org/10.1021/acs.accounts.0c00810.

(28) Wu, S.; Snajdrova, R.; Moore, J. C.; Baldenius, K.; Bornscheuer, U. T. Biocatalysis: Enzymatic Synthesis for Industrial Applications. *Angew. Chem. Int. Ed.* **2021**, *60* (1), 88–119. https://doi.org/10.1002/anie.202006648.

CHAPTER THREE

GENOME MINING AND MANIPULATION TO INVESTIGATE ERYTHRAZOLE A AND B BIOSYNTHESIS

3.1 Genome mining strategies to search for biosynthetic gene clusters or genes involved in erythrazole biosynthesis

3.1.1 Initial genome mining with standard strategies

In order to determine the biosynthetic genes and enzymes responsible for the biosynthesis of EA/EB and their interesting chemical moieties, we had the SNB-035 genome sequenced. Illumina HiSeq sequencing was performed with 75.0x genome coverage by Jonathan Goodson at the University of Maryland. Annotations were added by the NCBI Prokaryotic Genome Annotation Pipeline. The sequence and annotations are deposited on GenBank accession: GCA_029760855.1. The sequenced SNB-035 genome was first evaluated using antiSMASH[1] which we then investigated further by applying the deep learning strategy, DeepBGC[2] (Figure 3.1). Based on the structure of the molecules, we deduced that the gene cluster responsible for these molecules should have genes responsible for terpene synthesis (isoprenyl diphosphate synthases) and incorporation (prenyltransferases), shikimate pathway genes, an *O*-methyltransferase and genes encoding enzymes for amino acid incorporation. However, no such gene clusters were identified which had all of these components in relative proximity to each other.

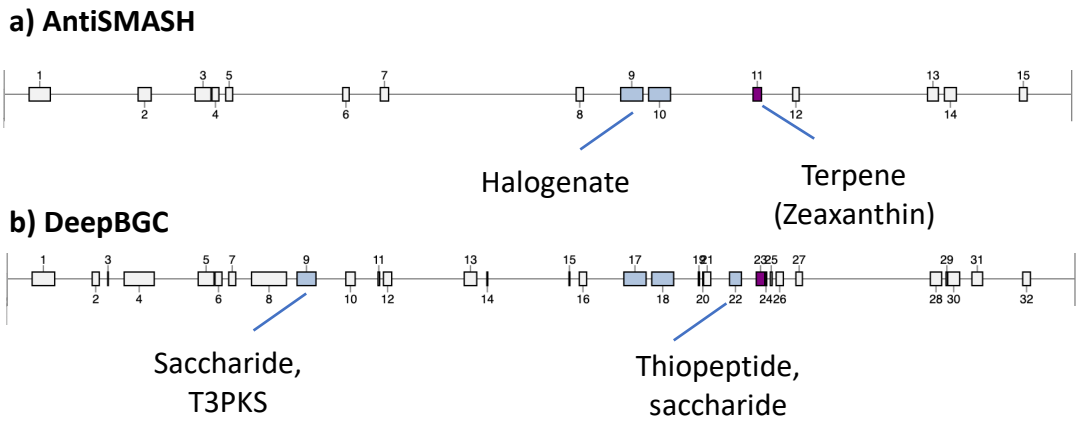


Figure 3.1 Antismash results and DeepBGC results. A.15 BGCs initially identified with antiSMASH[1]. Halogenated saccharide; most similar known cluster (MSKC): vazabitide A (NRP) 4%. Terpene; MSKC: zeaxanthin (66%). Unlabeled clusters are saccharides, fatty acids, and halogenated. **B.** 32 BGCs identified with DeepBGC[2]. Saccharide, T3PKS, MSKC: isoindolinomycin (PKS) 5%. Thiopeptide, saccharide; no similar cluster listed. Unlabeled clusters are saccharides, unknowns, fatty acids and halogenated.

We have proposed that the lack of a definitive biosynthetic gene cluster may be due to the erythrazoles being shunt products of the primary metabolism pathway which produces ubiquinone. The use of cofactors like NAD, FMN, and SAM as precursors for the biosynthesis of secondary metabolites[3] as described in chapter one shows precedence for secondary metabolism to use primary metabolites as building blocks. Since primary metabolism tends to be less clustered within bacterial genomes, this may explain the lack of observed clustering in the genes responsible for erythrazole biosynthesis and thereby gene cluster identification. We hypothesize that at the very least, tailoring enzymes, which convert ubiquinone or ubiquinone intermediates to the

erythrazoles, would be under similar regulation and therefore would be clustered. However, this may be a smaller, more discrete gene cluster.

The pathways which interested us in exploring due to their hypothesized role in erythrazole metabolism were the methylerythritol phosphate (MEP) pathway, ubiquinone pathway, and shikimate pathway. In order to identify genes from these pathways that may be related to erythrazole biosynthesis, we wanted to determine what was “normal” for these pathways in terms of distribution throughout the genome in alphaproteobacteria. To do this, these three pathways were identified in various alphaproteobacterial species and compared to those in the SNB-035 genome. Species were selected based on accessibility through Kyoto Encyclopedia of Genes and Genomes (KEGG)[4], which I used to identify the distribution of genes in these pathways.

3.1.2 Methylerythritol phosphate (MEP) pathway comparisons in alphaproteobacteria

Terpenes can be synthesized through the mevalonate pathway (MVA) or the methylerythritol phosphate (MEP) pathway, which differs based on species[5]. Most bacteria use the MEP pathway and annotations in the SNB-035 genome indicated that this organism uses the MEP pathway based on the presence of many MEP pathway genes. Therefore, we hypothesized that SNB-

035 uses the MEP pathway to biosynthesize the isoprenoid units of the erythrobacters. To assess the distribution of the MEP genes in SNB-035, we compared them to *Agrobacterium radiobacter* (KEGG entry ara00900) and *Sphingomonas henguiensis* (map00900). I identified that in *A. radiobacter*, *ispA* and *ispG* had some proximity to each other as do *ispB* and *ispE*. Meanwhile, in *S. hengshuiensis*, *ispE* is found near *ispDF* and DXR. These patterns are important to identify in order to distinguish between what is typical clustering of genes in these pathways vs. what may be unusual in SNB-035 and therefore suggesting a deviation from the typical primary metabolism pathway. In SNB-035 the only clustering of these genes was DXP and undecaprenyl diphosphate synthase. This clustering was noted as it differs from the other alphaproteobacterial species examined, there is no significant clustering or duplicates of any of the genes found in the other organisms.

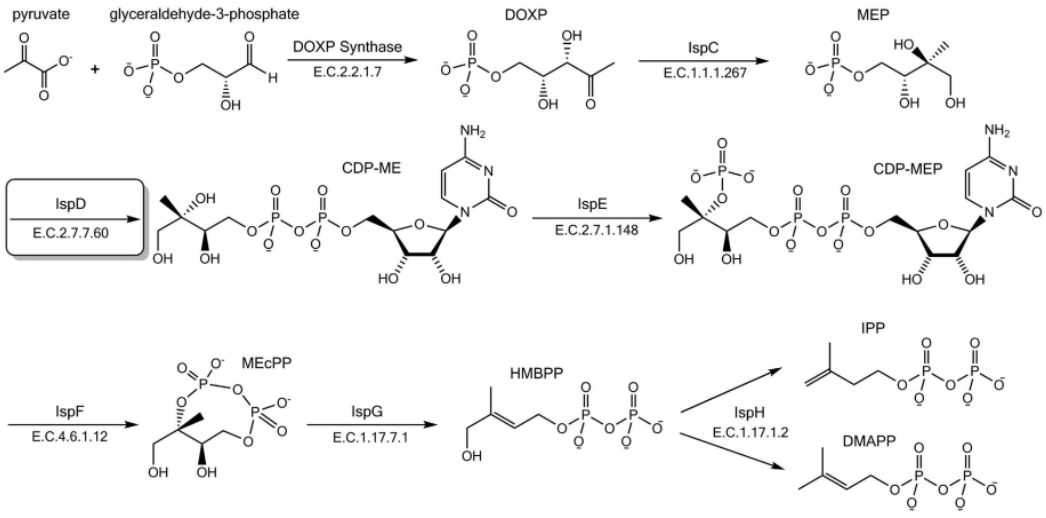


Figure 3.2 The methylerythritol phosphate pathway of isoprenoid biosynthesis. Figure from Price, K., Armstrong, C., Imlay, L. *et al.* Molecular

Mechanism of Action of Antimalarial Benzoisothiazolones: Species-Selective Inhibitors of the *Plasmodium* spp. MEP Pathway enzyme, IspD. *Sci Rep* **2006**, 6, 36777 (2016). https://doi.org/10.1038/srep36777

3.1.3 Shikimate pathway comparisons in alphaproteobacteria

Next, the shikimate pathway was examined in *Agrobacterium fabrum* and *Sphingomonas taxi*. The shikimate pathway is responsible for the biosynthesis of chorismate (also known as shikimic acid), which is the precursor of aromatic amino acids, specifically phenylalanine, tyrosine and tryptophan. It metabolizes phosphoenolpyruvate and erythraose 4-phosphate into chorismate, in 7 steps (Figure 3.3)[6].

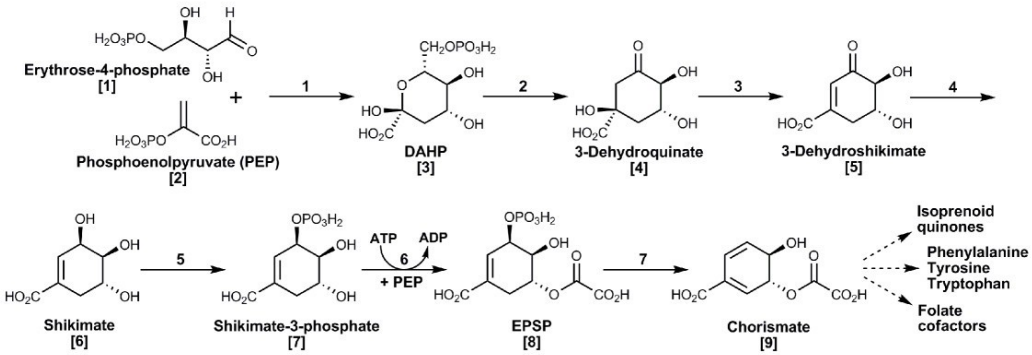


Figure 3.3 The shikimate pathway for aromatic amino acid biosynthesis. Figure from Zucko, J., Dunlap, W.C., Shick, J.M. et al. Global genome analysis of the shikimic acid pathway reveals greater gene loss in host-associated than in free-living bacteria. BMC Genomics **2010**, 11, 628. https://doi.org/10.1186/1471-2164-11-628.

In the comparison of the shikimate pathway in the *Agrobacterium* and *Sphingomonas* species, I found that 3-dehydroquinate synthase was found

clustered near shikimate kinase in both genomes. Additionally, in *S. taxi* 3-phosphoshikimate 1-carboxyvinyltransferase was located near chorismate synthase. Otherwise the genes were dispersed throughout the genome. In SNB-035, I observed the same pattern of clustering between 3-dehydroquinate synthase and shikimate kinase but no other clustering.

3.1.4 Ubiquinone biosynthesis pathway comparisons in alphaproteobacteria

The ubiquinone pathway comprises the shikimate pathway, MEP pathway, in addition to other steps. Figure 3.4b demonstrates metabolic engineering of *C. gluamicum* for CoQ10 production. This figure nicely depicts the relationship between these pathways in ubiquinone biosynthesis. Coenzyme Q10 is the primary form of ubiquinone found in erythrobacters[7–9]. The quinoid nucleus of ubiquinone is derived from the shikimate pathway, which is where the ubiquinone biosynthesis pathway begins (Figure 3.4a). For these comparisons I compared the pathways of *Agrobacterium rhizogenes* and *Zymomonas mobilis*. In the *A. rhizogenes* I saw no clustering of the ubiquinone biosynthesis genes. However, in the *Z. mobilis* genome *ubiG*, *ubiH* and *ubiF* were clustered. In SNB-035, *ubiG* is clustered with *ubiE*. While *ubiG* occurs twice in the ubiquinone pathway, there seems to only be one copy in the *A,*

rhizogenes and *Z. mobilis* genomes. This may indicate that one of these copies is involved in secondary metabolism.

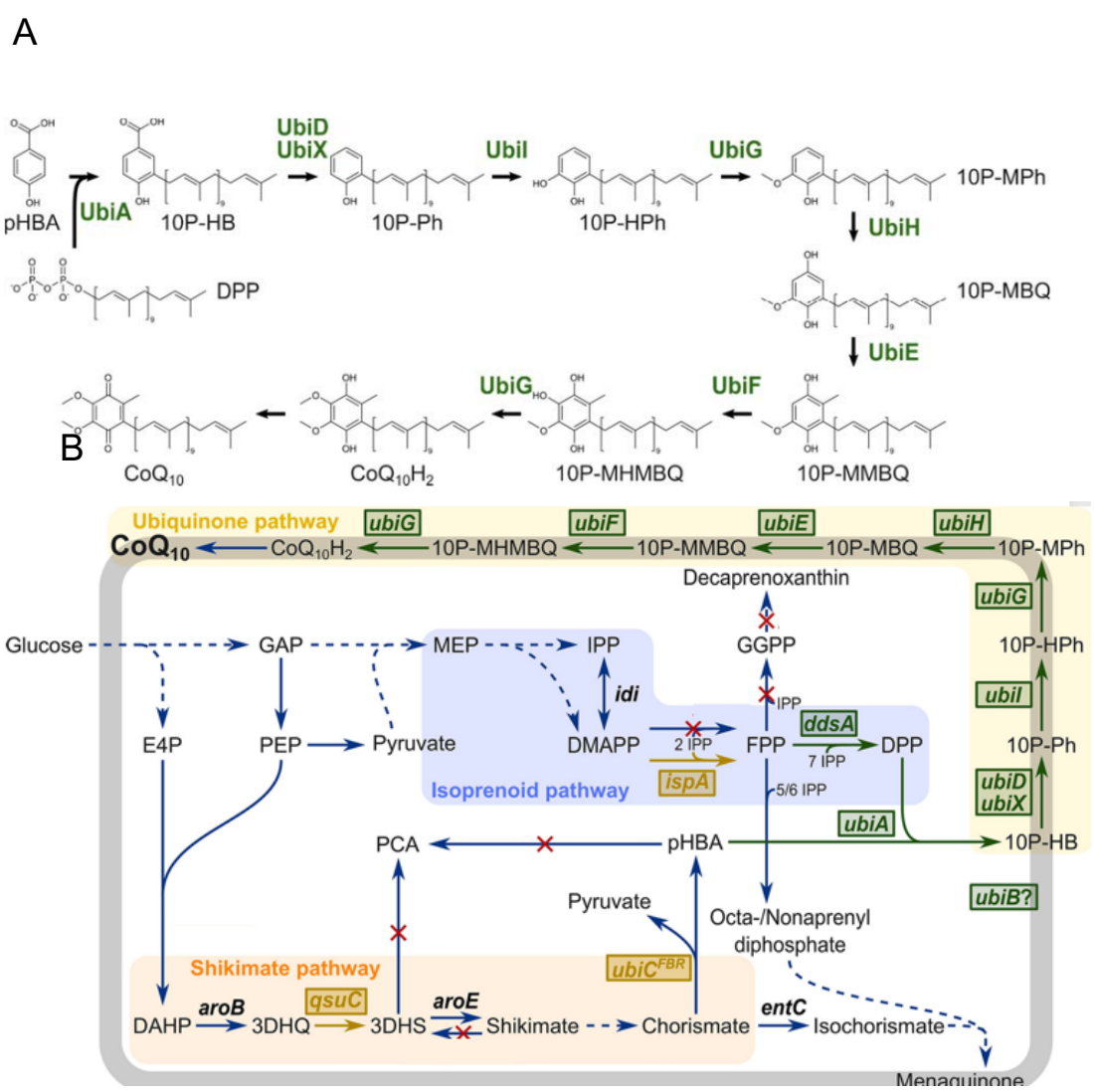


Figure 3.4 Biosynthesis of coenzyme Q_{10}. A. Ubiquinone biosynthesis **B.** Metabolic engineering of *C. glutamicum* for CoQ10 production. This figure demonstrates the connection between the shikimate pathway and MEP pathway in ubiquinone biosynthesis. Notes depicting gene manipulation can be disregarded for this work. From Burgadt, A. et al. Coenzyme Q10 Biosynthesis Established in the Non-Ubiquinone Containing *Corynebacterium glutamicum*

by Metabolic Engineering. *Front. Bioeng. Biotechnol*. **2021**, 9. https://doi.org/10.3389/fbioe.2021.650961.

3.2 Using custom profile Hidden Markov Models (HMMs) to identify discrete biosynthetic genes in the SNB-035 genome.

Hidden Markov Models (HMMs) are probability models which are generally applicable to time series of linear sequences[10]. Profile HMMs were established by Krogh et al.[11] to represent profiles of multiple sequence alignments. Profile HMMs have been used broadly by computational biologists and several software programs have been developed to implement them including anti-SMASH[1]. Others in the field have been using HMMs to identify novel biosynthetic genes which are out of the scope of typical programs that are designed to search for BGCs with well-established architecture like PKS, NRPS, RiPPs, and terpenes.[12–14] Profile HMMs can be tailored to the species of interest by building a multiple sequence alignment (MSA), which uses seeded guide trees and HMM profile-profile techniques to generate alignments between three or more sequences. HMMER then uses these alignments to establish homology between the sequences, which can then be used as a template to detect other genes with homology in the genome of interest. This method can also be useful as annotations of genes may not always be accurate. Using sequence homology to identify closely related proteins may

reveal a more accurate function of a gene than an annotation that may be outdated or overly broad.

We had predicted several families of enzymes which may play a role in erythrazole biosynthesis. These families include isoprenoid diphosphate synthases (IDS), prenyltransferases, cytochrome P450's (CYP450), and *O*-methyltransferases (OMT). I identified proteins in these families using Uniprot and National Center for Biotechnology Information (NCBI). These were then used to generate a multiple sequence alignment (MSA) through Clustal Omega[15], which was then used to build a profile HMM through HMMER 3.3.2[16]. HMMER 3.3.2 was then used to search the SNB-035 genome for the protein family of interest.

IDS	**GGPP**	**FPP**
O24743	P22873	Q9KWR7
P21684	P17060	ODT00158.1
WP_004941318.1	A0A0F7M0F7	MBA4008217.1
A0A6G9NBK1	C7C529	MBA4052658.1
A0A7D3XJB0	A0A0J7LXT2	MBA4043629.1
A0A345YE64	C7C5A8	GGA02957.1
	A0A0C5VAR7	MAQ66253.1
	Q6N9M1	OCJ62390.1
	A0A022PJ84	OCJ30308.1

	A0A0C5VQG7	PIW55970.1
	A9DA47	OJW76665.1
	A0A5E8GWJ1	
	Q08PP2	
	A0A7W4XP05	
	A0A7W9FRV5	
	A0A7W6RG93	
	A0A7W8MBK9	
	A0A7W9ZYZ0	
	A0A7Y9VW46	

Table 3.1 Accession numbers used to build multiple sequence alignments to build isopentyl diphosphate synthase (IDS) profile HMMs for HMMer analysis. Sequences are from either Uniprot or NCBI. Sequences annotating as an IDS were analyzed both together and separately than those annotating as a geranylgeranyl diphosphate (GGPP) synthase or farnesyl diphosphate (FPP) synthase.

OMT	**CYP450**	**Prenyltransferases**
WP_115415557.1	A0A2Z2PJ64_AGRFH	A0A7D4CCZ6
WP_173996927.1	WP_252258479.1	MBL44208.1
WP_103657808.1	KEO86203.1	NNC59157.1
WP_060977772.1	KZX88281.1	AXK43537.1
WP_209161946.1	Q5F4D9_SPHMC	BBI20526.1
A0A1S7RPR2	MBO6769075.1	UXS96154.1

WP_111991312.1	WP_011415403.1	269697892.1
MBD3756352.1	WP_173212924.1	MBF2715896.1
WP_146872800.1	NNC53505.1	OHC45330.1
Q2N843	GIX21007.1	161861282.1
WP_176274028.1	QIQ87614.1	WP_060722937.1
WP_279350884.1	MCL9983357.1	WP_159712261.1
NNC47323.1		WP_292930732.1
WP_107964481.1		WP_101796952.1
WP_115415805.1		WP_160613583.1
WP_160727546.1		BBI20419.1
MBB11474.1		KZX87712.1
		WP_263606722.1
		HBQ54415.1
		WP_292971945.1
		PHR02177.1
		WP_305958556.1
		NJN05327.1
		WP_179905136.1
		WP_119134007.1

Table 3.2 Accession numbers used to build multiple sequence alignments for profile HMMs for *O*-methyltransferases (OMT), cytochrome P450s (CYP450), and prenyltransferases for HMMER analysis. Sequences are from either Uniprot or NCBI.

Separate and combined searches were performed using genes annotating as isoprenoid diphosphate synthases (IDS), farnesyl diphosphate synthases (FPPS), and geranylgeranyl diphosphate synthase (GGPPS) (Table 3.1). IDSs are enzymes which catalyze the condensation of the C_5 substrates, DMAPP and IPP, which are biosynthesized through either the MEV or MVA pathway[17]. This class of enzymes can condense these units by increments of 1 (geranyl diphosphate synthase), 3 (FPPS), 4 (GGPPS), or sometimes 5 (Geranylfarnesyl diphosphate synthase). The length of the isoprene chain on the erythrazole A is comprised of 20 carbons and erythrazole B of 22 carbons. While EA would typically be produced by a GGPPS, we are considering the possibility that erythrazoles undergo oxidative cleavage to form the carboxylic acid from a previously longer chain. Therefore, I had to consider the possibility that other IDS's could be involved in the isoprene chain biosynthesis. The homology search identified a GGPPS and a gene annotated as polyprenyl synthetase. GGPPS was found near other putative biosynthetic genes but the polyprenyl synthase was not. All of the alignments created from similar IDS, FPPS, and GGPPS enzymes identified these same two genes in the SNB-035 genome, which makes sense because FPPS and GFPS are forms of IDSs which all share homology.

We hypothesized that CYP450's may play a role in the oxidative cleavage of the erythrazoles as they have been shown to catalyze a diverse range of oxidations and have been shown to act on terpenes[18] and even in aryl

terpenoids[19]. HMMER searches for cytochrome P450's identified three different genes spread throughout the genome. Two of these were located near other oxidative proteins, which could indicate the potential for the oxidative chemistry occurring in erythrazole biosynthesis. However, none of these were near anything that may incorporate the amino acids, which would be necessary for a gene cluster. HMMER searches for *O*-methyltransferases identified 4 proteins in the SNB-035 genome. The prenyltransferase HMMER search identified 5 proteins. These annotated as 4-hydroxybenzoate octaprenyl transferase (the enzyme which transfers prenyl groups in ubiquinone biosynthesis (*ubiA*)), undecaprenyl diphosphate synthase, a tRNA dimethylallyltransferase (DMAPP), and 2 hypothetical proteins. The hypothetical proteins were particularly interesting as the others all play known roles in primary metabolism. All of these identified proteins were noted but once again did not cluster near other relevant proteins.

Since we hypothesized that a carotenoid cleavage dioxygenase (CCD) may perform oxidative cleavage on the isoprene chain we also used this method to query the SNB-035 genome for enzymes with homology to CCDs and TglHI, the enzyme responsible for the β-carbon cleavage of cysteine in thiaglutamate biosynthesis.

TglH (DUF692)	Carotenoid cleavage dioxygenases
MBI3777992.1	YP_616744.1
MBA2351288.1	YP_616058.1
WP 086741107.1	ZP_01913312.1
WP_173212681.1	ZP_01912480.1
WP_067602679.1	NP_485149.1
WP_212449445.1	WP_194020870
WP_142788030.1	WP_061430495
WP_247270652.1	NP_189064.1
NNF94114.1	WP_011027026.1
KUO56976.1	
WP_122485290.1	
WP_024694451.1	
WP_003611576.1	
WP_065083569.1	
7TCR_A	
7TCW_A	
WP_024694451.1	

Table 3.3 Accession numbers used to build multiple sequence alignments to build profile HMMs for carotenoid cleavage dioxygenases and TglH (or DUF692) for HMMER analysis. Sequences are from either Uniprot or NCBI.

CCDs have been found to cleave terpene chains to form a terminal aldehyde, may be responsible for the formation of the carboxylic acid at the end of the erythrazoles. Bixin biosynthesis has been shown to use a CCD in conjunction with an aldehyde dehydrogenase to form a carboxylic acid from the cleavage of a longer isoprene chain[20]. Therefore, we predicted that a CCD may catalyze the oxidative cleavage of the erythrazoles. Hoffman et al. identified CCDs in two strains of alphaproteobacteria, *S. alaskensis RB226* and *P. pacifica SIR-1*. We searched for homology to these enzymes with precedent in alphaproteobacterial (first 4 accessions in table, Table 3.3), but no homology was detected in the SNB-035 genome. The multiple sequence alignment used was also then expanded to CCDs in fungi and cyanobacteria (Table 3.3). HMMER search for carotenoid cleavage dioxygenases yielded no results in the SNB-035 genome. This indicates that this class of enzymes is likely not responsible for the oxidative cleavage of the erythrazoles or that if they do, the one used here is significantly different than those seen before.

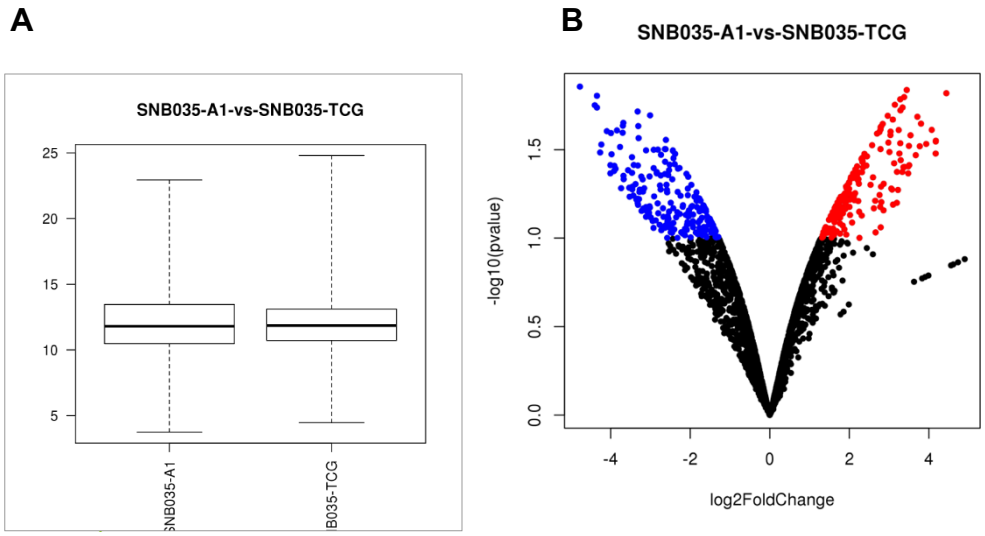
TglHI has been shown to catalyze a β-carbon cleavage of cysteine in thiaglutamate biosynthesis by *Pseudomonas syringae*[21]. We first searched the SNB-035 genome to homologs to TglHI however, no homologs were identified. TglH annotates on BLAST to DUF692, so we expanded our HMM to enzymes annotating as DUF692. An HMM was built using DUF692 sequences which resembled those found in erythrobacters and also just with those described by

Ting et al. to share close homology with TglH (last 3 accession numbers). However, none of these searches indicated any homologs.

3.3 RNA-Seq analysis

Due to the reduced production of erythrazoles when SNB-035 is grown in TCG media (Figure 3.15), we hypothesized that the genes which produce the erythrazole may not be expressed in TCG media. This observation allows for analysis using RNA-Seq as a tool to see which genes are being differently expressed in our regular nutrient-rich seawater-based media, A1BFe+C, vs. the TCG media. RNA for RNA-Seq was extracted using Monarch Total RNA Miniprep Kit (NEB #T2010S and RNA Sequencing was done by Azenta Life Sciences. Initial analysis of the RNA showed that SNB-035 grown in A1BFe+C had a total of 166,218,273 reads with a mean quality score of 36.06. SNB-035 grown in TCG had 159,562,214 reads with a mean quality score of 36.12. Figure 3.5 shows the analysis of the runs performed by Azenta. Overall, the expression values between the two samples were similar (Figure 3.5a). However, the heat map (Figure 3.5c) shows the top 30 genes based on significant difference between groups. These differences are as large as 4.8 log2 fold change with a p-value of 0.014. In the volcano plot analysis (Figure 3.5b), each dot represents a gene. The red dots represent upregulated genes while blue dots represent downregulated genes. Dots further from the center

on either side indicate genes with greater log2 fold change and those higher on the y-axis have more significant p-values. Therefore, the dots higher and further from the center overall are those which vary the most significantly between media conditions and therefore may be of interest in erythrazole production. For my analysis, I annotated all genes with $p<0.05$ significance and examined whether they are genes that may be part of the pathways which we hypothesized were related to erythrazole biosynthesis. Additionally, regions where many genes had a change of expression were noted. Particular notice was given to genes that were downregulated in the TCG media. This information was used to support which gene clusters were chosen as hypothesized biosynthetic gene clusters.



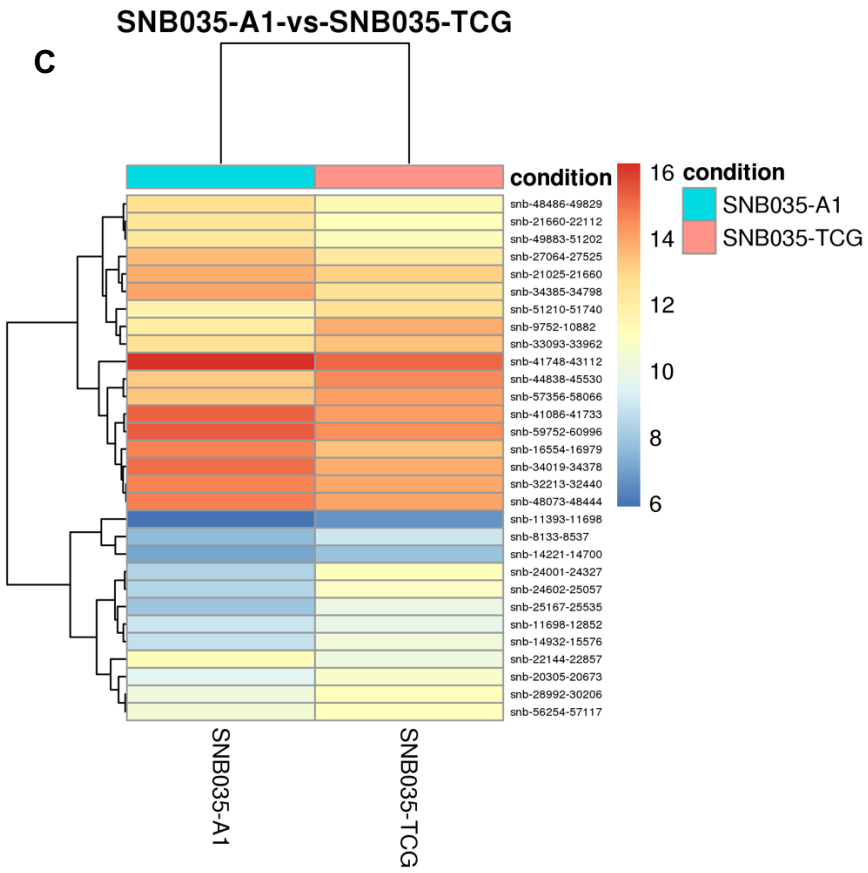


Figure 3.5 RNA-Seq analysis performed by Azenta. A. Normalized expression values of all of the genes from each group. **B.** Volcano plot showing the global transcriptional change across the A1BFe+C group (A1) and TCG group. **C.** Heat map showing the expression change between the top 30 genes with the greatest change in expression based on their p-value. The color scale indicates high relative expression (red) ranging to low relative expression (blue).

3.4 CRISPRi knockdown of genes in predicted biosynthetic gene clusters

3.4.1 Biosynthetic gene cluster predictions for manipulation

Four gene clusters were identified within the SNB-035 genome (Figure 3.6) which at minimum met some of the criteria which we defined as genes which encode for:

1. Enzymes capable of amino acid (cysteine and glycine) or peptide (glutathione) installation

2. Enzymes that play a role in terpene chemistry (isopentyl diphosphate synthases, prenyltransferases)

3. Shikimate pathway-like enzymes

4. Genes which were downregulated in the RNA-Seq analysis between A1BFe+C media and TCG media.

Additionally, genes which can do oxidative chemistry or are involved in ubiquinone biosynthesis were considered a plus.

Hypothetical BGC1 contains a phosphopantothenate-cysteine ligase (CoAB). CoAB is responsible for conjugating cysteine to 4'phosphopantothenate in coenzyme A biosynthesis[22]. This gene caught our

attention due to our interest in identifying an enzyme which can incorporate the cysteine of the erythrazoles. BGC1 also harbors ubiquinone biosynthesis *O*-methyltransferase (*ubiG*) and ubiquinone biosynthesis c-methyltransferase (*ubiE*), which perform methylation in the biosynthesis of ubiquinone[23,24]. This cluster also contains a glutathione s-transferase, which catalyse nucleophilic attack by glutathione on electrophilic groups[25]. The alcohol dehydrogenase iii and VOC family protein both have oxidation capabilities[26,27] which could be involved in the oxidative chemistry of erythrazole biosynthesis. Finally, slightly downstream of this cluster (~8.7kb) is a phenylalanine-4-hydroxylase, which catalyzes the hydroxylation of phenylalanine to form tyrosine[28]. This transformation closely resembles the hydroxylation of the aromatic ring of the erythrazoles and thereby fulfill the requirement for shikimate pathway related genes in the gene cluster. Overall, this cluster fulfills all of the requirements besides containing enzymes which account for the isoprene chain of the erythrazoles. This BGC also does not show a consistent decrease in expression for the TCG condition in the RNA-Seq experiments.

BGC2 also contains a glutathione S-transferase enzyme. BGC2 also contains a chorismate mutase, which catalyzes the Claisen rearrangement of chorismate to prephenate in the shikimate pathway enzyme[29]. The chorismate mutase is located directly adjacent to the geranylgeranyl diphosphate synthase (GGPPS), which is an isoprenyl diphosphate synthase that condenses 4 isoprene units (C_{20})[17]. A GGPS would be able to produce the C_{20} terpene chain

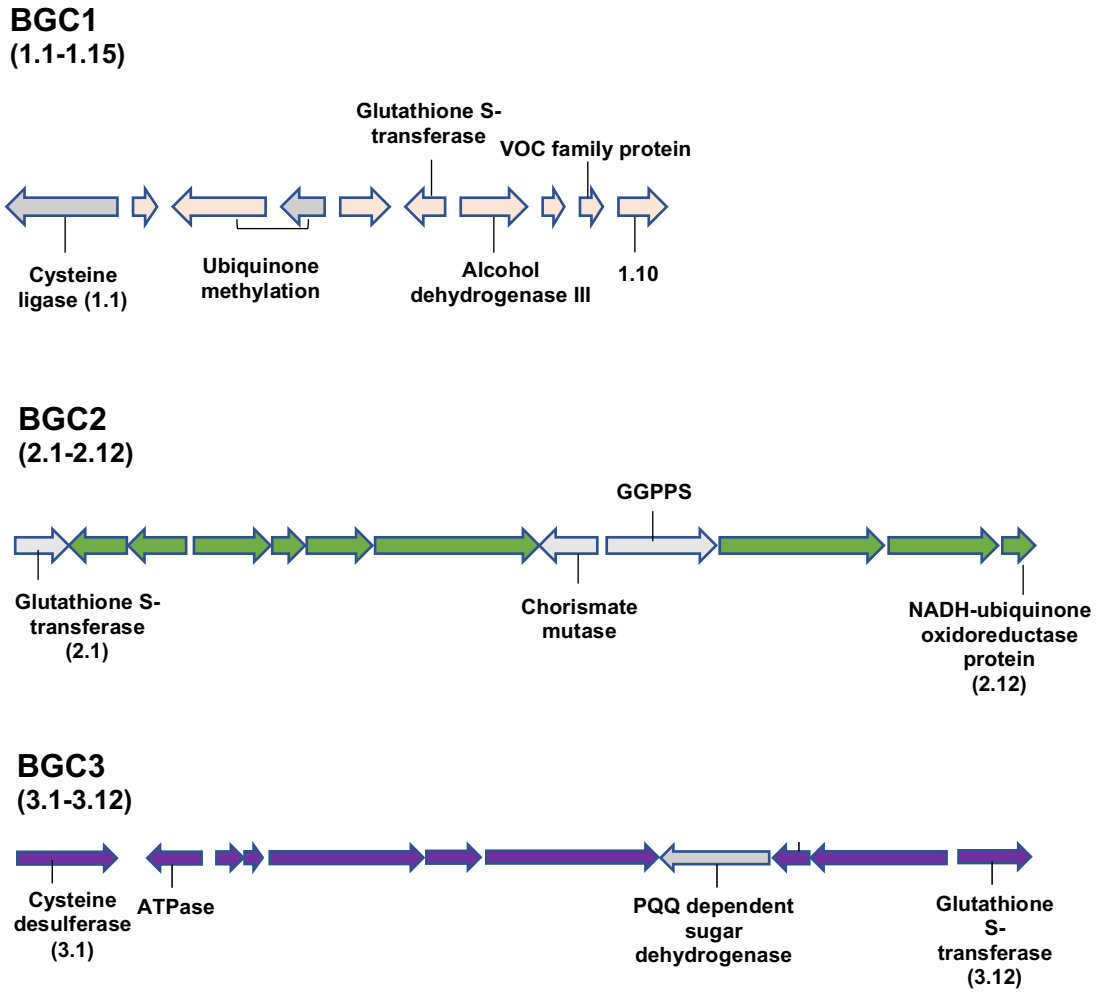
of EB. However, as we showed that EA and EB likely derive from oxidative cleavage of a longer terpene chain, a GGPS condensing only 4 isoprene units may not account for this. Finally, this gene cluster contains a NADH-ubiquinone oxidoreductase protein, which removes electrons from NADH and passes them to ubiquinone. Due to its reducing abilities and relationship with ubiquinone in addition to proximity to other enzymes of interest, we supposed that this could play a role in erythrazole biosynthesis. Overall, BGC2 fulfills all of the requirements for different pathway components assuming that cysteine and glycine are attached as glutathione rather than individually. All of the genes described here also show a decrease in expression when erythrazole production is decreased in the TCG media of the RNA-Seq experiments. Of note, the glutathione s-transferase in this cluster showed a 4-fold decrease in expression in the TCG media and showed one of the most significant changes in the RNA-Seq analysis.

BGC3 was mainly identified as a putative BGC due to the observed decrease in expression of many of the genes in this region in the RNA-Seq experiments. Several genes in this region were some of the most varied between the media conditions. In addition to the change of expression observed, this region also contains a cysteine desulferase, ATPase, pyrroloquinoline quinone (PQQ) dependent sugar dehydrogenase and a glutathione transferase. Cysteine desulferase was interesting to us due to the presence of cysteine in the erythrazoles. However, the chemistry that this

enzyme is known to do is to convert cysteine to alanine[30]. The ATPase may have a variety of functions but could be useful to provide energy or cofactors to a secondary metabolism mechanism. More interestingly in this cluster is the PQQ-dependent sugar dehydrogenase, which transfers electrons to ubiquinone during the oxidation of d-glucose. Electrons are then subsequently transferred to ubiquinol oxidase in the respiratory chain[31]. The involvement of these enzymes in ubiquinone metabolism may indicate that it could also be involved in erythrazole metabolism. Finally, BGC3 also contains a glutathione S-transferase. This gene cluster lacks terpene chemistry and shikimate pathway enzymes. However, we still wanted to probe it based on the observed expression change in our RNA-Seq experiments.

BGC4 contains several enzymes of interest in addition to important regulatory elements and oxidative enzymes. This gene cluster is also the BGC which we predicted would biosynthesize erythrazole C (EC), as many of the moieties of EC which drew us to the cluster are also in the structures of erythrazoles A-B. Thiazole synthase (ThiG) is an enzyme involved in the thiazole ring formation of thiamine, which occurs through the intermediate, 1-deoxy-D-xylulose 5-phosphate (DXP). The involvement in thiazole biosynthesis in addition to working on DXP, which is an intermediate in the MEP pathway, makes it an intriguing potential for involvement in erythrazole biosynthesis. BGC4 also contains a cysteine desulferase and a glycine ligase. Glycine ligase, or glycinamide ribonucleotide synthetase (GARS), catalyzes the second step

in purine biosynthesis where glycine, ATP and phosphoribosylamine are ligated together[32]. This was interesting to us as erythrazole biosynthesis may also require an enzyme which can condense glycine with other precursors. This gene cluster incorporates the amino acids including the thiazole formation. However, it lacks enzymes related to the shikimate pathway and terpene biosynthesis enzymes. This area also does not show consistent change in expression in the RNA-Seq experiments.



BGC4
(4.1-4.22)

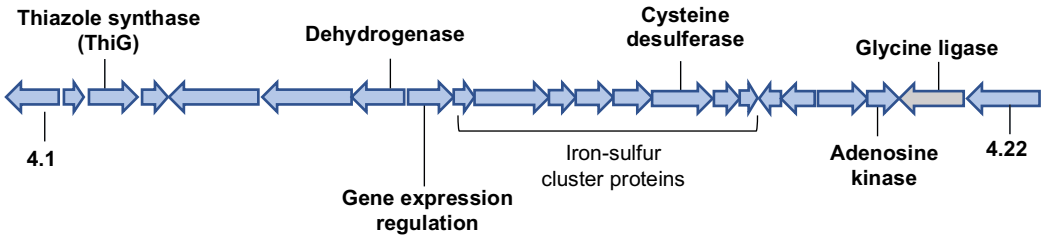


Figure 3.6 Biosynthetic gene cluster predictions. Numbers refer to genes described in table 3.4. Genes in grey were knocked down in CRISPRi studies.

Gene #	BLAST ID	Closest homolog accession number	Proposed Function
1.1	bifunctional phosphopantothenoylcysteine decarboxylase/phosphopantothenate synthase	WP_279351846.1	Catalyzes cysteine ligation and decarboxylation in coenzyme A biosynthesis
1.2	DUF4345 family protein	WP_160745426.1	Unknown
1.3	2-polyprenylphenol 6-hydroxylase (ubiI)	WP_279351788.1	Catalyzes the first hydroxylation in ubiquinone biosynthesis
1.4	class I SAM-dependent methyltransferase (ubiG)	WP_279351787.1	Catalyzes *O*-methylation in ubiquinone biosynthesis
1.5	bifunctional DNA-formamidopyrimidine glycosylase/DNA-(apurinic or apyrimidinic site) lyase	WP_279351786.1	Removes oxidation damaged bases and cleaves 3' and 5'-phosphodiester bonds of the resulting apurinic/apyrimidinic site
1.6	Glutathione S-transferase family protein	WP_279351785.1	Conjugates glutathione to xenobiotic substrates for detoxification
1.7	S-(hydroxymethyl)glutathione dehydrogenase/class III alcohol dehydrogenase	WP_279351782.1	High formaldehyde dehydrogenase activity in the

			presence of glutathione and catalyzes the oxidation of alcohols
1.8	Cupin domain-containing protein	WP_279351781.1	Variety of functions
1.9	Vicinal oxygen chelate (VOC) family protein	WP_279351780.1	Bidentate coordination to a divalent metal center by a substrate or intermediate through vicinal oxygen atoms
1.10	S-formylglutathione hydrolase	WP_279351779.1	Hydrolyzes S-formylglutathione in formaldehyde detoxification
2.1	Glutathione binding-like protein	WP_279351932.1	Glutathione import
2.2	Uracil-DNA glycosylase family protein	WP_279351931.1	DNA repair
2.3	ABC-F family ATP-binding cassette domain-containing protein, partial	WP_279351622.1	Membrane transport
2.4	SIMPL domain-containing protein	WP_279351621.1	Unknown
2.5	PepSY domain-containing protein	WP_279351620.1	Unknown
2.6	response regulator transcription factor	WP_279351619.1	Transcription factor
2.7	HAMP domain-containing sensor histidine kinase	WP_279351628.1	Sensor protein for signal transduction
2.8	Chorismate mutase	WP_279351618.1	Converts chorismate to prephenate in the shikimate pathway of aromtic amino acid biosynthesis
2.9	Polyprenyl synthetase family protein (FPPS - 89.35% in Erythrobacter sp. Accession MAM37649.1	WP_279351617.1	Condenses isoprenes to form terpenes
2.10	ATP-dependent helicase HrpB	WP_279351616.1	Catalyzes hydrolysis of ATP to ADP
2.11	ATP-dependent helicase	WP_279351616.1	Genome maintenance
2.12	ETC complex I subunit	WP_279351615.1	Mitochondrial electron transport
3.1	Cysteine desulfurase (EC 2.8.1.7)	WP_279350608.1	Converts cysteine to alanine
3.2	YifB family Mg chelatase-like AAA ATPase	WP_234091799.1	Cellular processes
3.3	ribonuclease P protein component	WP_279350603.1	Catayzes hydrolysis of the 5'leader sequence of precursor tRNA
3.4	membrane protein insertion efficiency factor YidD	WP_279350602.1	Insertion of integral membrane proteins into the membrane

3.5	membrane protein insertase YidC	WP_279350601.1	Insertion of integral membrane proteins
3.6	ribosome biogenesis GTP-binding protein YihA/YsxC	WP_279350600.1	Ribosome biogenesis
3.7	OmpA family protein	WP_279350599.1	Resistance to environmnetal stress
3.8	S9 family peptidase	WP_279350598.1	Serine dependent peptidase
3.9	PQQ-dependent sugar dehydrogenase	WP_279350597.1	Reduction of ubiquinone to ubiquinol
3.10	VOC family protein	WP_279350594.1	Bidentate coordination to a divalent metal center by a substrate or intermediate through vicinal oxygen atoms
3.11	von Willebrand factor type A domain protein	WP_279350593.1	Cellular processes
3.12	Glutathione S-transferase family protein	WP_279350592.1	Conjugates glutathione to xenobiotic substrates for detoxification
4.1	MerC domain-containing protein	WP_279351131.1	Mercury transport
4.2	sulfur carrier protein ThiS	WP_279351048.1	Donates sulfur in the synthesis of thiazole in thiamine phosphate biosynthesis
4.3	Hypothetical protein	WP_279351049.1	Unknown
4.4	Long-chain-fatty-acid--CoA ligase	WP_279351132.1	Fatty acid metabolism
4.5	Gamma-glutamyltranspeptidase	WP_279351050.1	Catalyzes transfer of gamma-glutamul of glutathione to an accpetor. Plays role in synthesis and degradation of glutathione
4.6	quinone-dependent dihydroorotate dehydrogenase	WP_279351051.1	Converts the conversion of dihydroorotate to orotate with quinone as an electron acceptor
4.7	helix-turn-helix domain-containing protein	WP_279351052.1	Regulatory - can be transcriptional activator or repressor
4.8	SUF system Fe-S cluster assembly regulator	WP_279351053.1	Can be involved in electron transfer, redox and non-redox

			catalysis, gene regulation, or sensors of oxygen and iron
4.9	Fe-S cluster assembly protein SufB	WP_279351054.1	The SufBCD complex stimulates cysteine desulferase activity. Contributes to repair of oxygen-labile iron-sulfur custers under oxidative stress
4.10	DUF559 domain-containing protein	WP_279351056.1	Unknown
4.11	Fe-S cluster assembly ATPase SufC	WP_279351057.1	ATPase in SUF system
4.12	SufD family Fe-S cluster assemby protein	WP_279351059.1	The SufBCD complex stimulates cysteine desulferase activity. Contributes to repair of oxygen-labile iron-sulfur custers under oxidative stress
4.13	SufD family Fe-S cluster assembly protein	WP_279351059.1	The SufBCD complex stimulates cysteine desulferase activity. Contributes to repair of oxygen-labile iron-sulfur custers under oxidative stress
4.14	Cysteine desulfurase (EC 2.8.1.7)	WP_279351060.1	Converts cysteine to alanine
4.15	SUF system Fe-S cluster assembly protein	WP_279351061.1	Iron-sulfur complex assembly
4.16	Iron-sulfur cluster assembly accessory protein	WP_279351062.1	Iron-sulfur complex assembly
4.17	Hypothetical protein	WP_279351063.1	Unknown
4.18	EI24 domain-containing protein	WP_279351064.1	Apoptosis
4.19	Adenosine kinase	WP_279351065.1	Phosphorylates adenosine and other nucleosides
4.20	Phytanoyl-CoA dioxygenase family protein	WP_279351066.1	Phytanic acid oxidation
4.21	Phosphoribosylamine-glycine ligase	WP_279351067.1	Catalyzes the formation of N(1)-(5-phospho-D-ribosyl)glycinamide from 5-phospho-D-ribosylamine and

			glycine in purine biosynthesis
4.22	Exodeoxyribonuclease 7 large subunit (EC 3.1.11.6) (Exodeoxyribonuclease VII large subunit)	WP_279351068.1	DNA degredation

Table 3.4 Descriptions of genes in proposed biosynthetic gene clusters. Accession numbers provided of closest homolog. All genes had 90% or greater sequence similarity to genes in *Erythrobacter litoralis* (NCBI: txid39960). Gene functions were provided by InterPro[33].

Overall, BGC2 fills the most amount of the requirements which we set up to identify the erythrazole BGC while also showing a consistent decrease of expression in the TCG condition in the RNA-Seq experiment. However, since several other regions in the genome contained other genes of interest or showed an interesting change in expression in our RNA-Seq experiments, we wanted to explore multiple regions. Additionally, if the biosynthetic genes are not clustered, looking at more than one area may be beneficial. Based on these hypotheses, I went forward to explore these areas of the SNB-035 genome using CRISPRi.

3.4.2 Establishing Mobile-CRISPRi in Erythrobacter, SNB-035

In order to investigate whether any of the genes which were predicted to be involved in biosynthesis of the erythrazoles are in fact playing this role, I utilized CRISPR interference (CRISPRi) technology to reduce transcription of these genes. CRISPRi uses an inactive Cas9 nuclease (dCas9) with a single guide RNA (sgRNA) to form an RNA-protein complex which sterically blocks the transcription of certain genes[34]. The method I chose to use was based on

that described by Banta et al. in the protocol “Programmable Gene Knockdown in Diverse Bacteria Using Mobile-CRISPRi”[34]. The use of Mobile CRISPRi allows for the components of the CRISPRi to be easily exchangeable. These components include a dCas9, an sgRNA, and an antibiotic resistance marker for selection in recipient bacteria. Additionally, Mobile-CRISPRi is inducible and titratable, allowing the targeting of essential genes[35,36].

Since SNB-035 is a Gram-negative bacteria, I used the Tn7 system, which transfers to the recipient bacteria via triparental mating. Tn7 is a transposon that inserts at high frequency into a specific site called its attachment site (*attTn7*) in the genome of many Gram-negative bacteria. The *attTn7* site is located downstream of the *glmS* gene, which encodes a highly conserved and essential protein. *E. coli* donor strains are used for the triparental mating where one donor carries a plasmid with Tn7 transposon ends that flank the CRISPRi components. Another donor carries a plasmid that expresses Tn7 transposase genes. The transposase is then able to integrate the Tn7 transposon carrying the CRISPRi machinery into the chromosome at the *attTn7* site[37]. This processis shown in Figure 3.7.

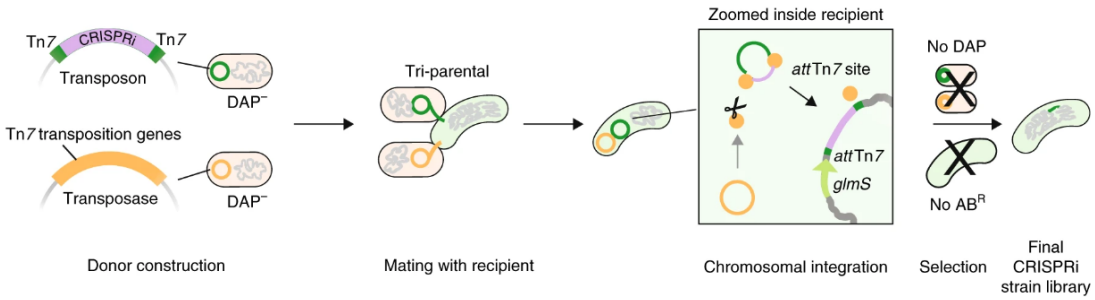


Figure 3.7 Mobile CRISPRi components integration into the bacterial chromosome through the Tn7 system. Figure from Peters, J.M, Koo, B.M. Patino, R. *et al*. Enabling genetic analysis of diverse bacteria with Mobile-CRISPRi. *Nat Microbiol* **2019**, *4*, 244-250. https://doi.org/10.1038/s41564-018-0327-z.

The authors of the protocol also developed test vectors for the integration of the CRISPRi system into various hosts. These test systems incorporate a fluorescent marker (monomeric red fluorescent protein (mRFP) or superfolder green fluorescent protein (sfGFP)) into the host genome along with the CRISPRi machinery with either a targeting sgRNA to suppress expression of the fluorescent marker or without a targeting sgRNA. This can then determine how well the system is working by comparing the fluorescence of the marker between the targeting or non-targeting strains. These vectors also include two versions of the dCas9; Spy::3xmyc and HSA Spy::3xmyc. These dCas9 versions have varying efficacy in different hosts. Specifically, HSA Spy::3xmyc has been seen to sometimes work better in high GC organisms. Therefore, I chose to use try test vectors with each dCas9s and with either targeting sgRNA for a fluorescent marker or without the targeting

sgRNA (pJMP1187, pJMP1189, pJMP2822, pJMP2834). These vectors are described in the table below, which is adapted from Banta et al[34].

Plasmid name	Addgene no.	Reporter	Antibiotic resistance	dCas9	Description
pJMP1187	119256	mRFP	amp^R, kan^R	HSA Spy::3Xmyc	mRFP "test" (+ sgRNA)
pJMP1189	119257	mRFP	amp^R, kan^R	HSA Spy::3Xmyc	mRFP "test" (- sgRNA)
pJMP2822	160670	sfGFP	amp^R, kan^R	Spy::3Xmyc	sfGFP "test" (+ sgRNA)
pJMP2834	160673	sfGFP	amp^R, kan^R	Spy::3Xmyc	sfGFP "test" (- sgRNA)
pJMP1039	119239	None	amp^R, kan^R	HSA Spy::3Xmyc	Expresses Tn7 transposase
pJMP1339	119271	None	amp^R, kan^R	HSA Spy::3Xmyc	For cloning new sgRNAs

Table 3.4 Mobile-CRISPRi vectors used for test system and knockdown construction.

Strains were grown in quadruplicate in either an induced or uninduced condition. Fluorescence was determined at days 2, 3 and 4 with optical density (OD) also measured to normalize the data. Data was analyzed using the

formula: (Mutant fluorescence/OD) - (WT fluorescence/OD). Overall, inhibition of the fluorescent marker was successful for the mRFP marker but not for the sfGFP marker (Figure 3.9). This indicated that the HSA Spy::3Xmyc dCas9 worked best in SNB-035 over the Spy::3Xmyc dCas9. Additionally, inhibition was greatest of mRFP expression at day 4. By day 4, the non-targeting strain had greatly increased mRFP detection while the targeting strain continued to show no expression. This informed me that for future CRISPRi experiments in SNB-035 that I should use vectors with the Spy::3Xmyc dCas9 and that knockdown may be best observed at later time points.

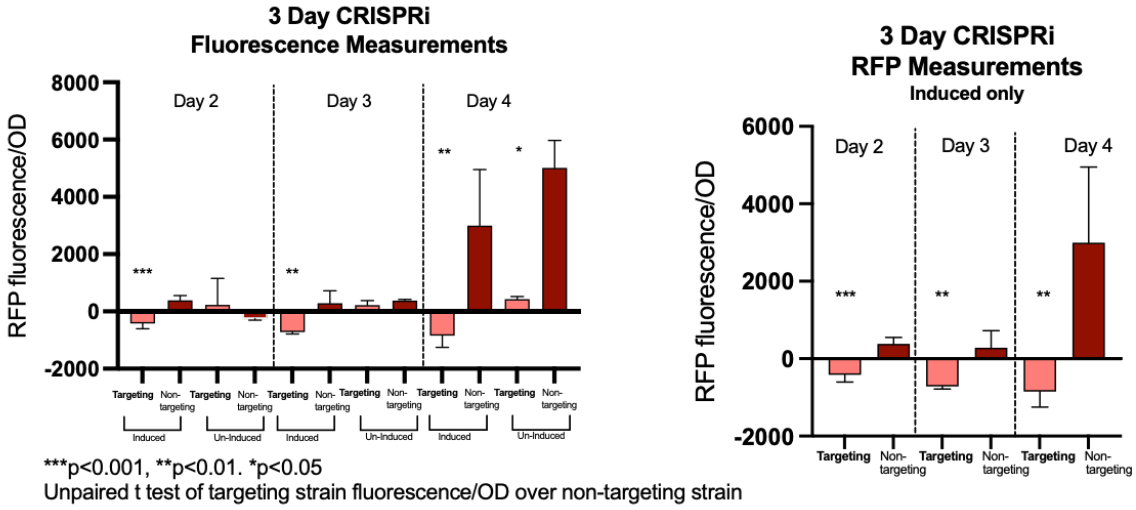


Figure 3.8 CRISPRi repression of mRFP with Spy::3Xmyc dCas9. A. Fluorescent measurements on days 2-4 comparing both targeting and non-targeting strains and induced vs. un-induced controls. **B.** Fluorescence measurements comparing induced only controls.

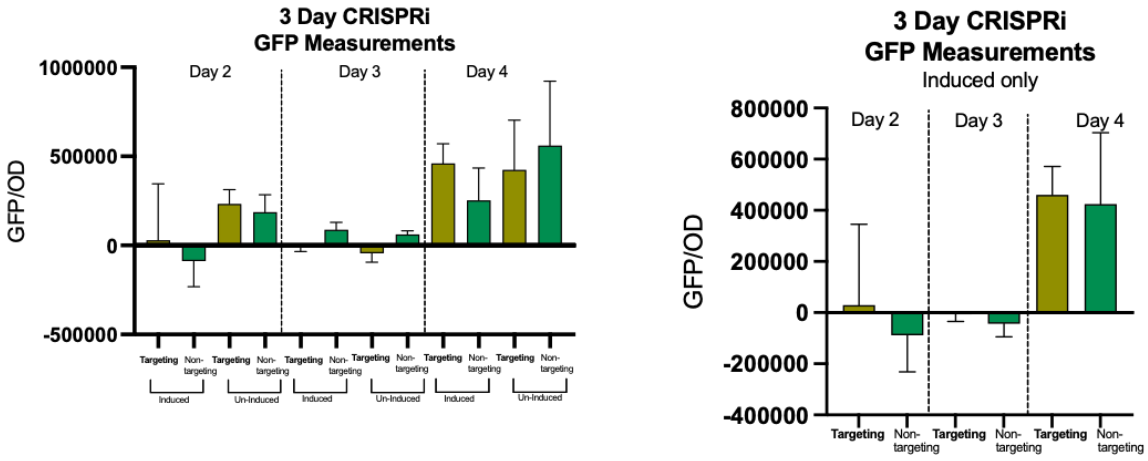


Figure 3.9 CRISPRi repression of sfGFP with HSA Spy::3Xmyc dCas9A. Fluorescent measurements on days 2-4 comparing both targeting and non-targeting strains and induced vs. un-induced controls. **B.** Fluorescence measurements comparing induced only controls.

3.4.3 CRISPRi of candidate biosynthetic genes

Genes to target were chosen based on the hypothesized biosynthetic gene clusters described above. In total, 11 knockdown strains were constructed, which target 7 different genes with 1-2 guides for each gene (Table 3.4). The genes targeted and their abbreviations used here are chorismate mutase (CM), ubiquinone biosynthesis *O*-methyltransferase (UbiG), geranylgeranyl diphosphate synthase (GGP), phosphopantothenate-cysteine ligase (CoAB), glycine ligase, glutathione S-transferase (GST), and pyrroloquinoline quinone-dependent sugar dehydrogenase (PQQ). The pJMP1189 vector contains the CRISPRi cassette but has no guide RNA, which was used in the fluorescence testing. This vector was conjugated into the SNB-

035 genome to serve as a control strain (1189), which has had the CRISPRi cassette integrated into the genome but is not targeting any genes for knockdown. We thought this would serve as a more appropriate control rather than the wild-type SNB-035 as it accounts for any disruptions in metabolism that may result from manipulating the bacterial chromosome. Gene name abbreviation and its offset are used to describe the strains which were constructed.

Gene	Offset	Cluster
Chorismate Mutase 1	10	2
Chorismate Mutase 1	19	2
UbiG	12	1
Geranylgernayl pyrophosphate synthase	41	2
CoAB	29	1
Glycine Ligase	9	4
Glycine Ligase	15	4
Glutathione S-transferase 2	0	2
Glutathione S-transferase 2	21	2
PQQ-dependent sugar dehydrogenase	13	3
PQQ-dependent sugar dehydrogenase	-29	3
1189		Control

Table 3.5 Genes targeted with CRISPRi with their corresponding spacers for guide RNAs. These targets are found in one of the four gene clusters described in 3.4.1. The offset is defined by the distance from the PAM site and is optimized to selectively target the gene of interest. Abbreviations in further results are referred to as gene abbreviation and offset. i.e. Chorismate mutase with 10 offset is referred to as CM10.

Successful integration of the CRISPRi cassette into the SNB-035 genome was confirmed by kanamycin selection and PCR. The PCR test used to confirm successful conjugation was adapted from Banta et al. (2020) which amplifies the region flanking the *attTn7* insertion site[38]. To verify that the genes were successfully knocked down, RT-PCR was performed on a few of the genes of interest (Figure 3.10). Both the Gene of interest and a control (16S) were amplified in the knockdown strains and compared to that in the 1189 control. Band intensity was determined with ImageLab software and fold change in expression was calculated based on the difference in expression between the target genes of interest in the knockdown strains vs. the control strain. This was run in duplicate where we saw an average of a 21 fold decrease in expression of glycine ligase with the GlyLi9 knockdown strains, 41 fold decrease in the GlyLi15 strain, a 594 fold decrease in the GST21 strain, and a 4.5 fold decrease in the UbiG knockdown. While the magnitude in the change of expression varied amongst the strains, overall this demonstrated that we successful blocked the transcription of target genes using CRISPRi without affecting whole genome transcription as shown by the 16S control.

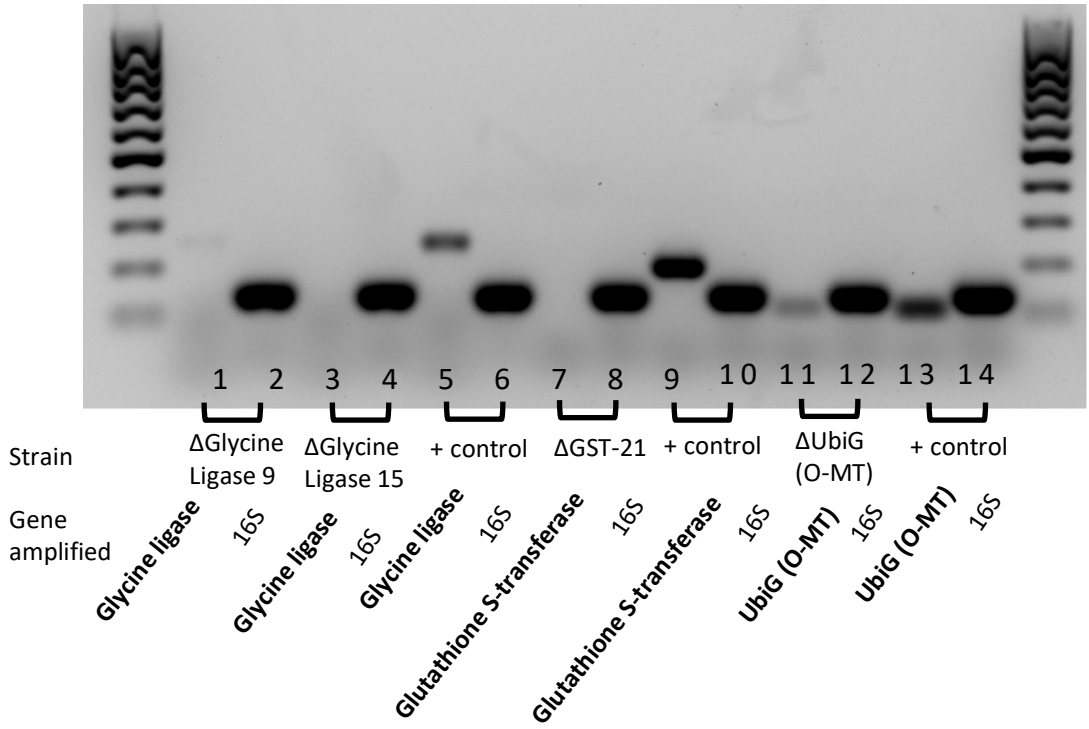


Figure 3.10 RT-PCR results of several knockdown strains. Knockdown gene of interest in addition to the 16S gene were amplified and compared to the 1189 positive control.

3.4.4 Knockdown effect on erythrazole production

To test the production of the erythrazoles in the knockdown strains, cultures were grown in triplicate to an OD of 2-2.5 before being induced with 1mM isopropyl β-d-1-thiogalactopyranoside (IPTG). Strains were then grown for 7 days shaking 200rpm at 30°C. They were extracted using liquid:liquid partitioning according to procedures described. Erythrazole production was determined with area under the curve and normalized to an internal standard of 1μM chloramphenicol (CAM). These results are shown in Figure 3.11.

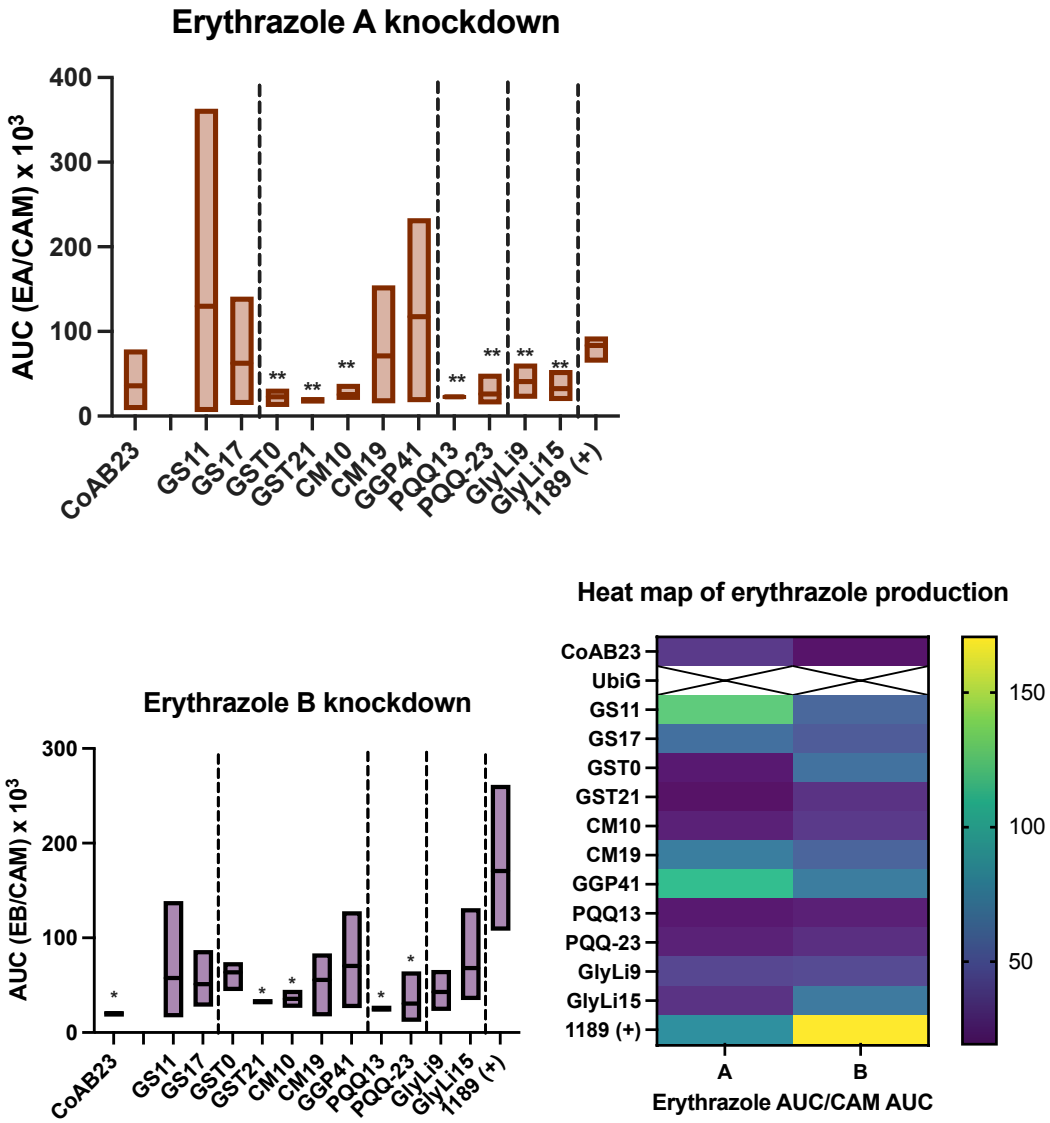


Figure 3.11 Erythrazole production by CRISPRi strains. AUC was normalized to internal standard of 1μM chloramphenicol (CAM) with n=3 per group. Δ*ubiG* was excluded here for visualization purposes but is shown in Figure 3.13. **p<0.05, *p<0.01 for Welch T-test between control (1189) and knockdown strains. Line in the center of the boxes indicates the mean with the minimum and maximum values depicted by the edges of the box. Heat map indicates high to low AUC values for EA and EB (also normalized to 1μM CAM).

Overall, these data show that many of the knockdown strains decreased erythrazole production. While these experiments did not indicate a definitive BGC or specific enzymes that are responsible for erythrazole biosynthesis, the results did provide further insight. Welch unpaired t-tests were performed between the normalized AUC of EA or EB of the knockdown strains and those in the control strain (1189). Significance of these t-tests is indicated with **$p<0.05$ or *$p<0.01$. T-tests of the EB data showed less significance, likely because the 1189 control showed greater variability in EB production. For erythrazole A production, GST0, GST21, CM10, PQQ13, PQQ-23, GlyLi9, and GlyLi15 all show signifant decrease with $p<0.05$. CoAB23, GST21, CM10, PQQ13, and PQQ-23 show a decrease with $p<0.1$ for EB production. Altogether, these experients show a significant decrease in both erythrazole production in strains which target glutathione s-transferase (GST), chorismate mutase (CM), pyrroloquinoline quinone-dependent sugar dehydrogenase (PQQ), and glycine ligase (GlyLi). Interestingly, these are genes from three of the four hypothesized BGCs. BGC1 showed the least amount of decrease when targeted for knockdown, suggesting that this cluster can be ruled out as being related to erythrazole biosynthesis. Optical density of the cultures was also measured on day 7 prior to extraction (Figure 3.12). These results indicated that none of the knockdowns significantly impacted growth of the microbe.

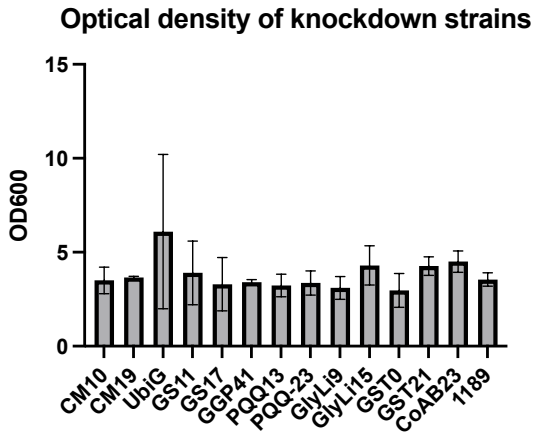


Figure 3.12 OD_{600} of knockdown strains measured at day 7 of growth prior to extraction.

One particularly interesting result of these studies was that the Δ*ubiG* knockdown results in a higher production of the erythrazoles. The S-adenosylmethionine (SAM)-dependent methyltransferase, *ubiG*, catalyzes the two *O*-methylation steps in CoQ_{10} biosynthesis[23,24,39]. Mutations of *ubiG* have been shown to hinder bacterial growth[40] and the level of *ubiG* expression has been shown have a direct effect on CoQ_{10} biosynthesis[41–44]. The *ubiG* knockdown resulted in a decrease of 3-4 fold expression of UbiG (Figure 3.10). While this knockdown is not hugely changing expression, this may have prevented any significant harm to bacterial growth. Interestingly, the decrease in expression resulted in an average of a 9-fold increase in erythrazole A production and a 3-fold increase in erythrazole B (Figure 3.13). Additionally, Δ*ubiG* actually showed varying OD values, often being higher than the other

strains (Figure 3.12). This indicated that the knockdown may be increasing growth rather than inhibiting it, which has been seen previously.

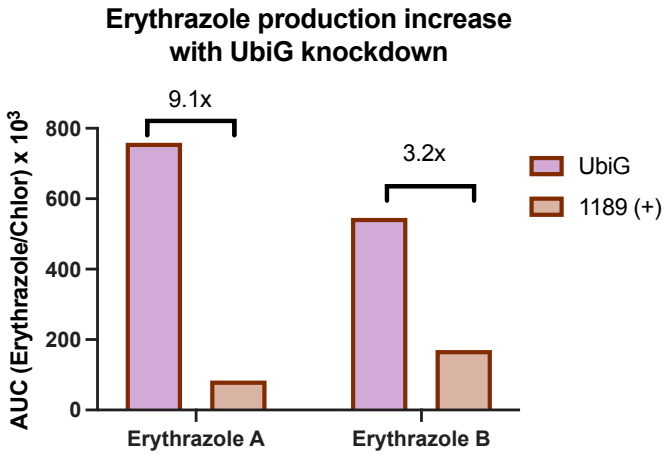


Figure 3.13 *ubiG* knockdown results in increased erythrazole production.

Knockdowns of *ubiG* have been shown to cause an accumulation of the intermediates, 2-polyprenylphenol and 2-polyprenyl-3-methyl-6-methoxy-1,4-benzoquinone[45]. Therefore, the increased erythrazole production when *ubiG* expression is decreased suggests that 2-polypreylphenol is an intermediate of erythrazole biosynthesis. This suggests that this may be where the erythrazole pathway deviates from ubiquinone, rather than what we originally proposed at $DDMQ_{10}$ (Figure 3.10).

We also wanted to compare the changes we saw in the glycine ligase knockdowns vs. the glutathione S-transferase knockdowns as our isotope labeling study results were inconclusive regarding the incorporation of glycine vs. glutathione. When initially looking at the knockdown strains' production of

the erythrazoles (Figure 3.11), they look equal. However, when we considered the expression change in these strains, we realized that GST expression is much more significantly effected. To visually compare this, we plotted GlyLi9, GlyLi15, and GST21 erythrazole production and gene expression (Figure 3.15). GST0 was not included as GST21 had significantly greater impact on expression while GlyLi9 and GlyLi15 were comparable.

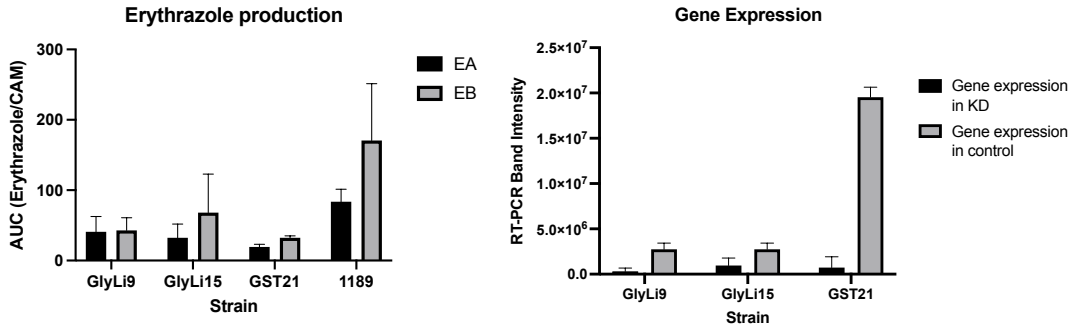


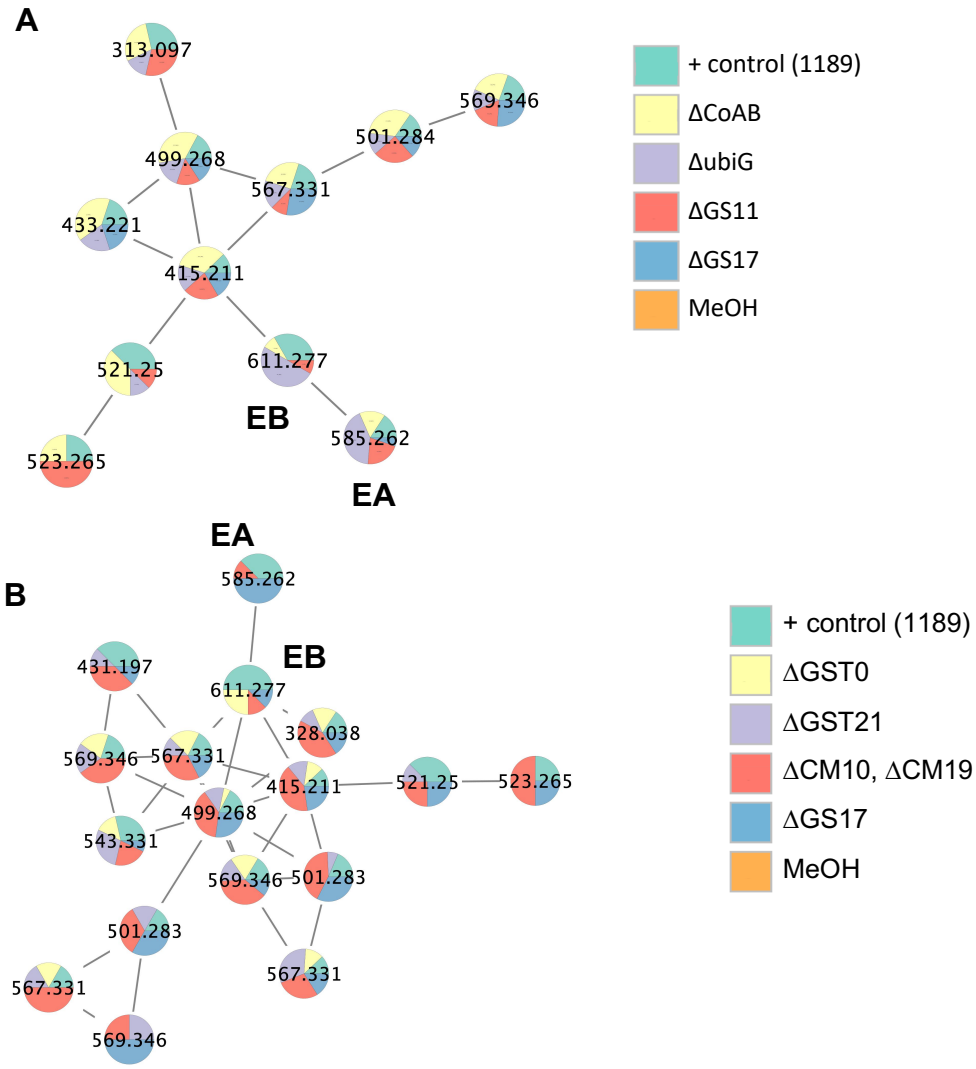
Figure 3.14 Comparison of glycine ligase and glutathione S-transferase knockdowns in both erythrazole production and gene expression.

While the expression analysis of these strains showed that GlyLi9 only caused a decreased average expression of glycine ligase of 21-fold while GST21 had an average decreased expression of glutathione S-transferase (GST) of 594-fold. These values were derived from the band intensity in the RT-PCR experiment (Figure 3.10). Quantitative RT-PCR (qPCR) attempts were unsuccessful to quantify these values as there was no detection of the targets in many of the knockdown strains. These data show that while glutathione S-transferase expression was more greatly inhibited, glycine ligase

showed a greater decrease of erythrazole production compared to erythrazole intermediates. Overall, these strains production of the erythrazoles was approximately equal. Therefore, the reduced erythrazole production in the glycine ligase knockdowns is due to a more minor change in gene expression than with the GST21 knockdown. This observation suggests that glycine ligase is more likely responsible for erythrazole production than the GST which was targeted. If the targeted GST were involved in erythrazole biosynthesis we would expect to see practically depleted erythrazole production in the GST21 strain.

3.5 Global Natural Products Social Molecular Networking analysis

Global Natural Products Social Molecular Networking (GNPS) detects structurally similar molecules based on MS/MS patterns. The platform does this based on the assumption that similar molecules will fragment in similar ways[46]. This networking is a good way to identify intermediates or to discover similar molecules in a crude mixture. As described here, the erythrazoles distinctively fragment their glycine and acetylglycine groups. Since none of the knockdowns had full loss of erythrazole production, we thought to use GNPS to look in the strains with reduced erythrazole production for intermediates that may be accumulating. This could tell us which part of the pathway is being hindered in each CRISPRi strain.


A
313.097
569.346
501.284
499.268
567.331
433.221
415.211
521.25
611.277
EB
585.262
523.265
EA
+ control (1189)
ΔCoAB
ΔubiG
ΔGS11
ΔGS17
MeOH
B
EA
585.262
EB
431.197
611.277
328.038
569.346
567.331
415.211
521.25
523.265
499.268
543.331
569.346
501.283
501.283
567.331
567.331
569.346
+ control (1189)
ΔGST0
ΔGST21
ΔCM10, ΔCM19
ΔGS17
MeOH

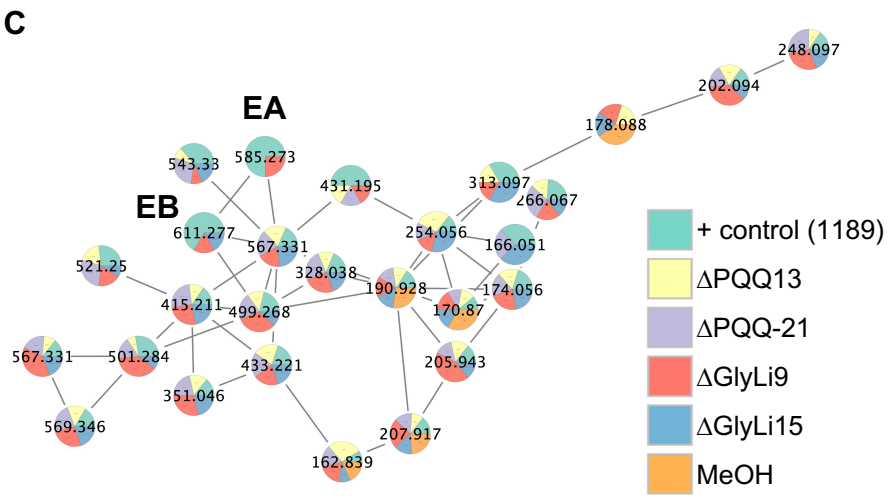


Figure 3.15 GNPS molecular networking of knockdown strains. Analysis was done separately with **A** showing BGC1 knockdowns, **B** showing BGC2 knockdowns, and **C** showing BGC3 and 4 knockdowns. Nodes showing EA and EB are indicated. Pie graphs of nodes indicate the percentage which each group contributes to the feature shown on the node. MeOH control is included to show background features. Nodes which were majority MeOH were excluded from the figures.

This molecular networking identified 4 molecules which consistently clustered with EA (m/z 585.26) and EB (m/z 611.28). The m/z of these molecules are 567.33, 499.27, 415.21, and 328.04. Of all the knockdown strains, CoAB, chorismate mutase, and GlyLi9 show the greatest accumulation of these related molecules. All of these knockdowns also showed a significant decrease in production of EA and EB other than CoAB, which only showed a significant decrease in EB. When these strains were combined and analyzed through GNPS, they all showed equal accumulation of these intermediates. While the cysteine ligase (CoAB) and glycine ligase (GlyLi) may target the

amino acid portion of the erythrazoles, chorismate mutase would affect the aromatic ring as it plays a role in shikimate biosynthesis. These results were therefore overall inconclusive.

We attempted to hypothesize the structures of some of these intermediates based on the observed m/z in negative mode $(M-H)^-$. None of these features have an acetylglycine fragment besides 328.04. This molecule would have an odd molecular weight, which differs than the erythrazoles which have even molecular weights. This indicates that according to the nitrogen rule[47], there must be one more or one less nitrogen present in the molecule than in the erythrazoles. Therefore, while the molecule with m/z = 328.04 may have an acetylglycine moiety it does not have the nitrogen from cysteine or potentially has an additional nitrogen. The biosynthetic logic of the erythrazoles would incorporate cysteine prior to glycine, so the identity of this intermediate remains unknown.

The feature with m/z = 567.33 showed exciting potential to be related to the glycine ligation due to its high abundance in the glycine ligase knockdowns (Figure 3.17) and its lack of an acetylglycine fragment. We proposed that it may be the benzothiazine intermediate prior to glycine installment, though we are unsure of the structure. We proposed something that would look like (**1**), though this exact structure is not correct as it does not align with the correct mass. Therefore, we tried to isolate this compound using similar techniques to those which we used for the erythrazoles to perform structure elucidation. However,

upon isolation, the molecule degraded and we were not left with enough material for NMR experiments.

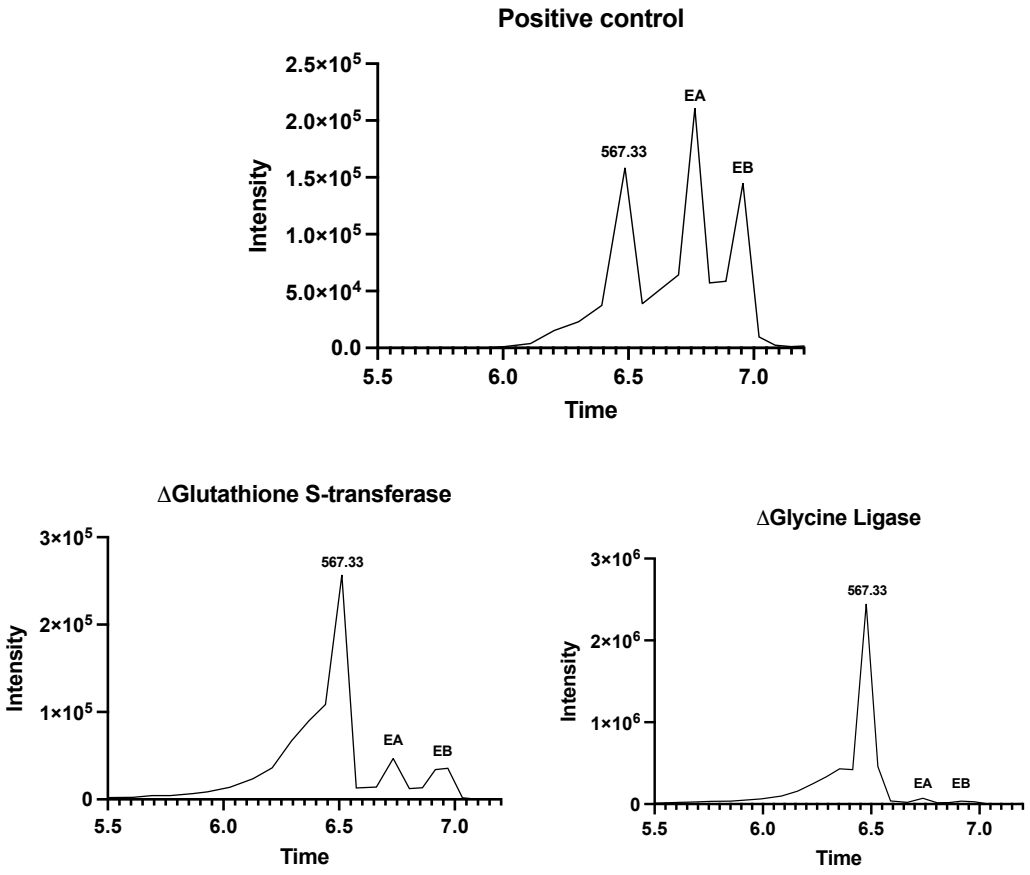
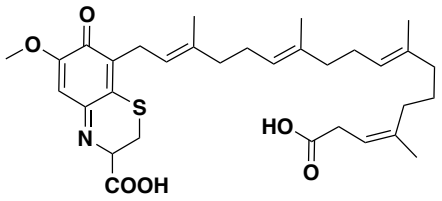


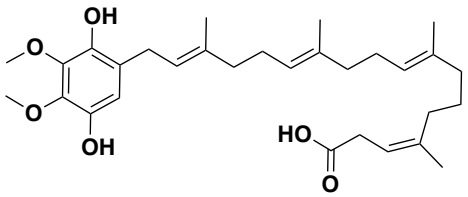
Figure 3.16 Extracted ion chromatograms (EIC) for the m/z 567.3, 585.26 (EA) and 611.28 (EB). ΔGlutathione S-transferase EIC is shown for the GST21 strain and ΔGlycine ligase for the GlyLi9 strain.

We were able to propose a structure which aligned with the correct mass for the observed feature with m/z of 499.27 (Figure 3.17). This structure shows a putative structure of EB prior to amino acid installment and with an additional methoxy group. This structure seems reasonable given its accumulation in knockdowns which aimed to interrupt amino acid incorporation. However,

overall the identity of this molecule and the other putative intermediates remains elusive.



1
Exact mass = 469.28



2
Exact mass = 500.31

Figure 3.17 Proposed structure of intermediates. 1 is the proposed structure prior to glycine installment and 2 is the proposed structure prior to any amino acid incorporation. 2 aligns with a feature detected to be related to the erythrazoles while 1 does not.

3.6 Discussion

Overall, the SNB-035 genome has been extensively studied and examined for erythrazole biosynthetic genes using a variety of different methods. We started genome mining for an erythrazole BGC with antiSMASH though we acknowledged that while the erythrazoles have a very unique structure compared to previously known natural products this platform may be limited by its preference for BGCs with more typical architecture. As we expected, this yielded no significant discoveries and neither did the employment of the deep learning strategy DeepBGC.

We then acknowledged that a more manual approach to genome mining may be more applicable to these unusual molecules, which would seek to identify a biosynthetic handle that may lead to the discovery of the other

biosynthetic genes. This first involved identifying the pathways which seemed related to erythrazole production within the genome of SNB-035 and comparing them to those in other alphaproteobacterial species. This analysis showed that SNB-035 had relatively standard organization of the genes in these pathways other than the clustering of the MEP pathway genes DXP and undecaprenyl diphosphate synthase and the presence of an additional *ubiG* gene in the SNB-035 genome.

Subsequently, to look for more discrete genes that may be a part of erythrazole biosynthesis, we employed custom built profile HMMs to query the SNB-035 for genes of interest. These HMMs were designed both to look for homologs of enzymes with activity similar to what we would expect to happen in erythrazole biosynthesis. To expand our search, profile HMMs were built with genes in similar organisms to SNB-035 and those which were more divergent but showed more similar biosynthetic mechanisms. The ability to tailor these searches for both gene function and in specific phyla is a great benefit of using this method rather than using more standard genome mining platforms. This method identified several genes of well-known families and function which could have putative involvement in erythrazole production (IDSs, prenyltransferases, CYP450, O-MT). However, similarly to what was observed thus far, no clustering of these genes was observed. Homology to genes with more specific function like carotenoid cleavage dioxygenases for the oxidative cleavage of terpenes and TglHI, which shows a similar β-carbon cleavage

mechanism as what we proved is occurring in chapter 2 was not identified in the SNB-035 genome.

Another idea which was inspired by the results of chapter 2 was to determine the global expression differences between SNB-035 growing in media where it shows high erythrazole production and that which SNB-035 shows nearly depleted erythrazole production (Figure 3.15). By performing RNA-Seq on SNB-035 grown in these different conditions, we were able to establish more basis for which we could propose certain genes' involvement in erythrazole production. This experiment in addition to the previously discussed research helped to develop hypotheses for four different putative BGCs.

The putative BGC's were investigated using CRISPRi knockdowns. CRISPRi of target genes was overall successful, showing the first instance of manipulation of an *Erythrobacte*r or *Qipengyuania* species genome to our knowledge. We were able to detect up to a 594-fold decrease in expression in these knockdown strains. A major benefit of using CRISPRi is that the knockdowns are titratable and do not involve any deletions in the genome. Rather, machinery is added to the genome which can block transcription of specified genes. This allowed us to manipulate the SNB-035 genome without the same concern of downstream effects as a full gene deletion may have. However, adding elements into the genome may be metabolically expensive which is why we used a strain which had these added elements but did not target genes for knockdown as our control.

The results of the knockdowns told us several things regarding erythrazole biosynthesis. First, we were able to eliminate one entire gene proposed BGC as it did not show significantly decreased erythrazole production. Unexpectedly, the knockdown of the ubiquinone biosynthesis *O*-methyltransferase, *ubiG*, showed increased erythrazole production. This was an extremely significant result as it supports our hypothesis that the erythrazoles are derived from the same intermediates which are part of the ubiquinone biosynthetic pathway. We were not able to quantify ubiquinone production due to its instability and solubility. However, as previous studies have showed that *ubiG* knockdowns hinder ubiquinone production but cause an increase of the precursors 2-polyprenylphenol and 2-polyprenyl-3-methyl-6-methoxy-1,4-benzoquinone[45], we can deduce that this is what is happening in our knockdown. This accumulation of 2-polyprenylphenol can then be repurposed as an erythrazole precursor, which could cause the observed increase.

While this result provides is useful insight for our work, it also shows the ability to metabolically engineer SNB-035 to increase erythrazole production. Natural product research and use is often limited by low production in the host organism[48,49]. Therefore, this knockdown strain may be useful for future studies which aim to isolate EA or EB. The fact that a relatively small decrease in expression (average 4.5 fold, Figure 3.10) could cause a 3-9 fold increase in

erythrazole production, suggests that future optimization of this knockdown may be able to increase production to even greater amounts.

Finally, the comparison of the glycine ligase knockdown to the glutathione s-transferase knockdown in relation to their change in expression suggested that the targeted GST is not related to erythrazole biosynthesis. However, as there are several GSTs in the SNB-035 genome we cannot outrule involvement of those. Regardless, the relatively low change in expression of glycine ligase leading to significant decrease in erythrazole production does support the idea that glycine ligase is in fact resposible for the glycine installment on the erythrazoles. This was also supported by the increase in intermediates present in the extract of the glycine ligase knockdowns.

Overall, these results led to an updated biosynthetic proposal (Figure 3.19). We propose that 2-polyprenylphenol (**1**) undergoes the same hydroxylation followed by an *O*-methylation, followed by another hydroxylation as that which occurs in CoQ10 biosynthesis. Due to the results of the knockdown, we can assume a different enzyme is responsible for the *O*-methylation. Then, selective oxidative cleavage occurs, resulting in the 10 isoprenes of **3** to be cleaved to an isoprene chain comprised of 20 (EA) or 22 (EB) carbons (**4**, **5**). Next, cysteine is incorporated followed by the β-carbon cleavage. Finally, glycine is attached to **7**, likely by glycine ligase.

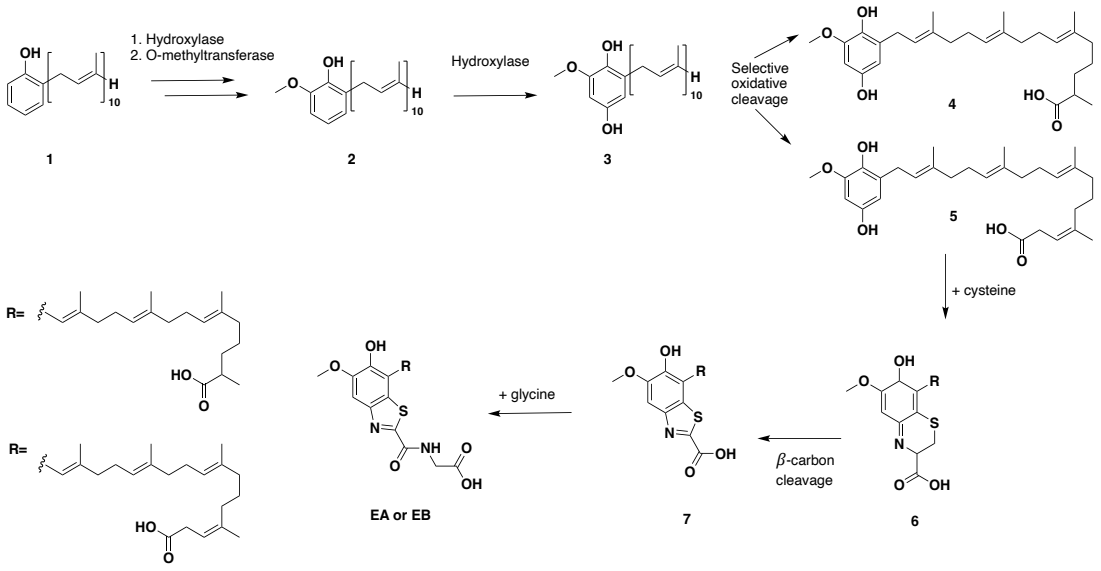


Figure 3.18 Updated proposed biosynthesis of erythrazoles A and B incorporating ubiquinone pathway intermediates suggested by the results from *ubiG* knockdown.

While the we data collected supports this hypothesis, a few pieces remain undetermined. We cannot definitively state that the oxidative cleavage occurs after the hydroxylation and *O*-methylation steps. While the amino acid fragmentation was diagnostic for other intermediates, these steps likely occurred before amino acid installation and therefore will not show the distinct amino acid fragmentation. Of note, upon isolation of the erythrazoles, we noticed a MS feature with a m/z 30 greater than each erythrazole with similar MS/MS fragmentation to the erythrazoles (Figure 3.15) which indicates that an additional *O*-methylation occurs at the carbon 5 position, similar to in ubiquinone. This shows that the enzymes we described may also act upon this scaffold.

Future work should investigate genes in proximity to glycine ligase. However, at this point I believe that the biosynthetic genes are not clustered. If our proposed pathway is correct, the only genes necessary for biosynthesis would be an *O*-methyltransferase, perhaps a hydroxylase (if not repurposed from *ubiI*), an oxidative cleavage enzyme, an enzyme for cysteine incorporation, an enzyme for its subsequent β-carbon cleavage, and glycine ligase. This is at minimum 5 biosynthetic genes, most of which are not found in the SNB-035 genome with proximity to glycine ligase. I think it is likely that the other *ubiG* gene found in the genome is the *O*-methyltransferase involved. This *ubiG* is located ~200kb apart from glycine ligase. Similarly, about halfway between these genes I recently discovered the presence of a luciferase-like monooxygenase (LLM). LLMs have been shown to perform ring contractions[50], which could be involved in the β-carbon cleavage occurring in the benzothiazole formation of the erythrazoles. Future work should use the methods described here to probe these genes.

3.7 Chapter three references

(1) Medema, M.; Blin, K.; Cimermancic, P.; Zakrzewski, P.; Fischbach, M.; Weber, T.; Takano, E.; Breitling, R. antiSMASH: Rapid Identification, Annotation and Analysis of Secondary Metabolite Biosynthesis Gene Clusters in Bacterial and Fungal Genome Sequences. *Nucleic Acids Res.* **2011**, *39*, W339–W346.

(2) Hannigan, G. D.; Prihoda, D.; Britton, D. A. A Deep Learning Genome-Mining Strategy for Biosynthetic Gene Cluster Prediction. *Nucleic Acids Res.* **2019**, *47* (18), e110.

(3) Barra, L.; Awakawa, T.; Abe, I. Noncanonical Functions of Enzyme Cofactors as Building Blocks in Natural Product Biosynthesis. *JACS Au* **2022**, *2* (9), 1950–1963. https://doi.org/10.1021/jacsau.2c00391.

(4) Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **2000**, *28* (1), 27–30. https://doi.org/10.1093/nar/28.1.27.

(5) Frank, A.; Groll, M. The Methylerythritol Phosphate Pathway to Isoprenoids. *Chem. Rev.* **2017**, *117* (8), 5675–5703. https://doi.org/10.1021/acs.chemrev.6b00537.

(6) Zucko, J.; Dunlap, W. C.; Shick, J. M.; Cullum, J.; Cercelet, F.; Amin, B.; Hammen, L.; Lau, T.; Williams, J.; Hranueli, D.; Long, P. F. Global Genome Analysis of the Shikimic Acid Pathway Reveals Greater Gene Loss in Host-Associated than in Free-Living Bacteria. *BMC Genomics* **2010**, *11* (1), 628. https://doi.org/10.1186/1471-2164-11-628.

(7) Lee, Y. S.; Lee, D.-H.; Kahng, H.-Y.; Kim, E. M.; Jung, J. S. Erythrobacter Gangjinensis Sp. Nov., a Marine Bacterium Isolated from Seawater. *Int. J. Syst. Evol. Microbiol.* **2009**, *60* (6), 1413–1417.

(8) Yoon, B.-J.; Lee, D.-H.; Oh, D.-C. Erythrobacter Jejuensis Sp. Nov., Isolated from Seawater. *Int. J. Syst. Evol. Microbiol.* **2013**, *63*, 1421–1426.

(9) Harashima, K.; Nakada, H. Carotenoids and Ubiquinone in Aerobically Grown Cells of an Aerobic Photosynthetic Bacterium, Erythrobacter Species OCh114. *Agric. Biol. Chem.* **1983**, *47* (5), 1057–1063.

(10) Eddy, S. R. Hidden Markov Models. *Curr. Opin. Struct. Biol.* **1996**, *6* (3), 361–365. https://doi.org/10.1016/s0959-440x(96)80056-x.

(11) Krogh, A.; Brown, M.; Mian, I. S.; Sjölander, K.; Haussler, D. Hidden Markov Models in Computational Biology. Applications to Protein Modeling. *J. Mol. Biol.* **1994**, *235* (5), 1501–1531. https://doi.org/10.1006/jmbi.1994.1104.

(12) Scesa, P. D.; Lin, Z.; Schmidt, E. W. Ancient Defensive Terpene Biosynthetic Gene Clusters in the Soft Corals. *Nat. Chem. Biol.* **2022**, *18* (6), 659–663. https://doi.org/10.1038/s41589-022-01027-1.

(13) Yamada, Y.; Kuzuyama, T.; Komatsu, M.; Shin-ya, K.; Omura, S.; Cane, D. E.; Ikeda, H. Terpene Synthases Are Widely Distributed in Bacteria. *Proc. Natl. Acad. Sci.* **2015**, *112* (3), 857–862. https://doi.org/10.1073/pnas.1422108112.

(14) Wilson, K.; Rond, T. de; Burkhardt, I.; Steele, T. S.; Schäfer, R. J. B.; Podell, S.; Allen, E. E.; Moore, B. S. Terpene Biosynthesis in Marine Sponge Animals. *Proc. Natl. Acad. Sci.* **2023**, *120* (9), e2220934120. https://doi.org/10.1073/pnas.2220934120.

(15) Madeira, F.; Pearce, M.; Tivey, A. R. N.; Basutkar, P.; Lee, J.; Edbali, O.; Madhusoodanan, N.; Kolesnikov, A.; Lopez, R. Search and Sequence Analysis Tools Services from EMBL-EBI in 2022. *Nucleic Acids Res.* **2022**, *50* (W1), W276—W279. https://doi.org/10.1093/nar/gkac240.

(16) Eddy, S. R. Accelerated Profile HMM Searches. *PLOS Comput. Biol.* **2011**, *7* (10), 1–16. https://doi.org/10.1371/journal.pcbi.1002195.

(17) Oldfield, E.; Lin, F.-Y. Terpene Biosynthesis: Modularity Rules. *Angew. Chem. Int. Ed Engl.* **2012**, *51* (5), 1124–1137. https://doi.org/10.1002/anie.201103110.

(18) Zhang, X.; Guo, J.; Cheng, F.; Li, S. Cytochrome P450 Enzymes in Fungal Natural Product Biosynthesis. *Nat. Prod. Rep.* **2021**, *38* (6), 1072–1099.

(19) Dufossé, L.; Echanove, M. C. de. The Last Step in the Biosynthesis of Aryl Carotenoids in the Cheese Ripening Bacteria Brevibacterium Linens ATCC 9175 (Brevibacterium Aurantiacum Sp. Nov.) Involves a Cytochrome P450-Dependent Monooxygenase. *Food Res. Int.* **2005**, *38* (8), 967–973. https://doi.org/10.1016/j.foodres.2005.02.017.

(20) Us-Camas, R.; Aguilar-Espinosa, M.; Rodriguez-Campos, J.; Rivera-Madrid, R. Identifying Bixa Orellana L. New Carotenoid Cleavage Dioxygenases 1 and 4 Potentially Involved in Bixin Biosynthesis. *Front Plant Sci* **2022**, *13*.

(21) Ting, C.P; Funk, M. A.; Halaby, S. L.; Zhang, Z.; Gonen, T.; van der Donk, W. A. Use of a Scaffold Peptide in Thebiosynthesis of Amino Acid–Derivednatural Products. *Science* **2019**, *365* (6450), 280–284.

(22) Daugherty, M.; Polanuyer, B.; Farrell, M.; Scholle, M.; Lykidis, A.; Crécy-Lagard, V. de; Osterman, A. Complete Reconstitution of the Human Coenzyme A Biosynthetic Pathway via Comparative Genomics*. *J. Biol. Chem.* **2002**, *277* (24), 21431–21439. https://doi.org/10.1074/jbc.M201708200.

(23) Abby, S. S.; Kazemzadeh, K.; Vragniau, C.; Pelosi, L.; Pierrel, F. Advances in Bacterial Pathways for the Biosynthesis of Ubiquinone. *Biochim. Biophys. Acta BBA - Bioenerg.* **2020**, *1861* (11), 148259. https://doi.org/10.1016/j.bbabio.2020.148259.

(24) Meganathan, R. Ubiquinone Biosynthesis in Microorganisms. *FEMS Microbiol. Lett.* **2001**, *203* (2), 131–139. https://doi.org/10.1111/j.1574-6968.2001.tb10831.x.

(25) Allocati, N.; Federici, L.; Masulli, M.; Di Ilio, C. Glutathione Transferases in Bacteria. *FEBS J.* **2009**, *276* (1), 58–75. https://doi.org/10.1111/j.1742-4658.2008.06743.x.

(26) Lin, G.-H.; Hsieh, M.-C.; Shu, H.-Y. Role of Iron-Containing Alcohol Dehydrogenases in Acinetobacter Baumannii ATCC 19606 Stress Resistance and Virulence. *Int. J. Mol. Sci.* **2021**, *22* (18). https://doi.org/10.3390/ijms22189921.

(27) He, P.; Moran, G. R. Structural and Mechanistic Comparisons of the Metal-Binding Members of the Vicinal Oxygen Chelate (VOC) Superfamily. *J. Inorg. Biochem.* **2011**, *105* (10), 1259–1272. https://doi.org/10.1016/j.jinorgbio.2011.06.006.

(28) Lin, Y.; Sun, X.; Yuan, Q.; Yan, Y. Engineering Bacterial Phenylalanine 4-Hydroxylase for Microbial Synthesis of Human Neurotransmitter Precursor 5-Hydroxytryptophan. *ACS Synth. Biol.* **2014**, *3* (7), 497–505. https://doi.org/10.1021/sb5002505.

(29) Khanjin, N. A.; Snyder, J. P.; Menger, F. M. Mechanism of Chorismate Mutase: Contribution of Conformational Restriction to Catalysis in the Claisen Rearrangement. *J. Am. Chem. Soc.* **1999**, *121* (50), 11831–11846. https://doi.org/10.1021/ja992453d.

(30) Mihara, H.; Esaki, N. Bacterial Cysteine Desulfurases: Their Function and Mechanisms. *Appl. Microbiol. Biotechnol.* **2002**, *60* (1), 12–23. https://doi.org/10.1007/s00253-002-1107-4.

(31) Elias, M.; Tanaka, M.; Sakai, M.; Toyama, H.; Matsushita, K.; Adachi, O.; Yamada, M. C-Terminal Periplasmic Domain of Escherichia Coli Quinoprotein Glucose Dehydrogenase Transfers Electrons to Ubiquinone*. *J. Biol. Chem.* **2001**, *276* (51), 48356–48361. https://doi.org/10.1074/jbc.M107355200.

(32) Wang, W.; Kappock, T. J.; Stubbe, J.; Ealick, S. E. X-Ray Crystal Structure of Glycinamide Ribonucleotide Synthetase from Escherichia

Coli,. *Biochemistry* **1998**, *37* (45), 15647–15662. https://doi.org/10.1021/bi981405n.

(33) Paysan-Lafosse, T.; Blum, M.; Chuguransky, S.; Grego, T.; Pinto, B. L.; Salazar, G. A.; Bileschi, M. L.; Bork, P.; Bridge, A.; Colwell, L.; others. InterPro in 2022. *Nucleic Acids Res.* **2023**, *51* (D1), D418–D427.

(34) Banta, A. B.; Ward, R. D.; Tran, J. S.; Bacon, E. E.; Peters, J. M. Programmable Gene Knockdown in Diverse Bacteria Using Mobile-CRISPRi. *Curr. Protoc. Microbiol.* **2020**, *59* (1), e130. https://doi.org/10.1002/cpmc.130.

(35) Li, X.; Jun, Y.; Erickstad, M. J.; Brown, S. D.; Parks, A.; Court, D. L.; Jun, S. tCRISPRi: Tunable and Reversible, One-Step Control of Gene Expression. *Sci. Rep.* **2016**, *6* (1), 39076. https://doi.org/10.1038/srep39076.

(36) Peters, J. M.; Koo, B.-M.; Patino, R.; Heussler, G. E.; Hearne, C. C.; Qu, J.; Inclan, Y. F.; Hawkins, J. S.; Lu, C. H. S.; Silvis, M. R.; Harden, M. M.; Osadnik, H.; Peters, J. E.; Engel, J. N.; Dutton, R. J.; Grossman, A. D.; Gross, C. A.; Rosenberg, O. S. Enabling Genetic Analysis of Diverse Bacteria with Mobile-CRISPRi. *Nat. Microbiol.* **2019**, *4* (2), 244–250. https://doi.org/10.1038/s41564-018-0327-z.

(37) Peters, J. E. Targeted Transposition with Tn7 Elements: Safe Sites, Mobile Plasmids, CRISPR/Cas and Beyond. *Mol. Microbiol.* **2019**, *112* (6), 1635–1644. https://doi.org/10.1111/mmi.14383.

(38) Banta, A. B.; Enright, A. L.; Siletti, C.; Peters, J. M. A High-Efficacy CRISPR Interference System for Gene Function Discovery in Zymomonas Mobilis. *Appl. Environ. Microbiol.* **2020**, *86* (23), e01621-20. https://doi.org/10.1128/AEM.01621-20.

(39) Aussel, L.; Pierrel, F.; Loiseau, L.; Lombard, M.; Fontecave, M.; Barras, F. Biosynthesis and Physiology of Coenzyme Q in Bacteria. *18th Eur.*

Bioenerg. Conf. 2014 Lisbon Port. **2014**, *1837* (7), 1004–1011. https://doi.org/10.1016/j.bbabio.2014.01.015.

(40) Wu, G.; Williams, H. D.; Zamanian, M.; Gibson, F.; Poole, R. K. Isolation and Characterization of Escherichia Coli Mutants Affected in Aerobic Respiration: The Cloning and Nucleotide Sequence of ubiG. *Microbiology*, 1992, *138*, 2101–2112. https://doi.org/10.1099/00221287-138-10-2101.

(41) Lu, W.; Shi, Y.; He, S.; Fei, Y.; Yu, K.; Yu, H. Enhanced Production of CoQ10 by Constitutive Overexpression of 3-Demethyl Ubiquinone-9 3-Methyltransferase under Tac Promoter in Rhodobacter Sphaeroides. *Biochem. Eng. J.* **2013**, *72*, 42–47. https://doi.org/10.1016/j.bej.2012.12.019.

(42) Zhu, X.; Yuasa, M.; Okada, K.; Suzuki, K.; Nakagawa, T.; Kawamukai, M.; Matsuda, H. Production of Ubiquinone in Escherichia Coli by Expression of Various Genes Responsible for Ubiquinone Biosynthesis. *J. Ferment. Bioeng.* **1995**, *79* (5), 493–495. https://doi.org/10.1016/0922-338X(95)91268-A.

(43) Zhang, D.; Shrestha, B.; Li, Z.; Tan, T. Ubiquinone-10 Production Using Agrobacterium Tumefaciens Dps Gene in Escherichia Coli by Coexpression System. *Mol. Biotechnol.* **2007**, *35* (1), 1–14. https://doi.org/10.1385/MB:35:1:1.

(44) Burgardt, A.; Moustafa, A.; Persicke, M.; Wendische, V. F. Coenzyme Q10 Biosynthesis Established in the Non-Ubiquinone Containing Corynebacterium Glutamicum by Metabolic Engineering. *Front. Bioeng. Biotechnol.* **2021**, *9*.

(45) Stroobant, P.; Young, I. G.; Gibson, F. Mutants of *Escherichia Coli* K-12 Blocked in the Final Reaction of Ubiquinone Biosynthesis: Characterization and Genetic Analysis. *J. Bacteriol.* **1972**, *109* (1), 134–139. https://doi.org/10.1128/jb.109.1.134-139.1972.

(46) Wang, M. et al.; Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34* (8), 828–837. https://doi.org/10.1038/nbt.3597.

(47) Pellegrin, V. Molecular Formulas of Organic Compounds: The Nitrogen Rule and Degree of Unsaturation. *J. Chem. Educ.* **1983**, *60* (8), 626.

(48) Atanasov, A. G. et al; the International Natural Product Sciences Taskforce. Natural Products in Drug Discovery: Advances and Opportunities. *Nat. Rev. Drug Discov.* **2021**, *20* (3), 200–216. https://doi.org/10.1038/s41573-020-00114-z.

(49) Carroll, A. R.; Copp, B. R.; Davis, R. A.; Keyzers, R. A.; Prinsep, M. R. Marine Natural Products. *Nat. Prod. Rep.* **2022**, *39* (6), 1122–1171.

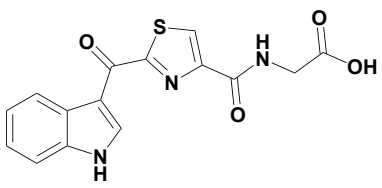
(50) Qin, Z.; Devine, R.; Booth, T. J.; Farrar, E. H.; Grayson, M. N.; Hutchings, M. I.; Wilkinson, B. Formicamycin Biosynthesis Involves a Unique Reductive Ring Contraction. *Chem. Sci.* **2020**, *11* (31), 8125–8131.

CHAPTER FOUR

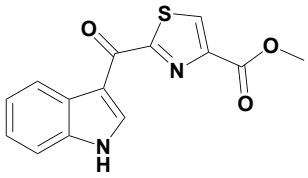
BIOSYNTHESIS AND CHARACTERIZATION OF ERYTHRAZOLE C AS AN ARYL HYDROCARBON LIGAND

4.1 Discovery of erythrazole C; putative aryl hydrocarbon receptor ligand

The MacMillan lab isolated both 1'H-indole-3'-carbonyl)- thiazole-4-carboxylic acid methyl ester (ITE) and a similar compound, erythrazole C (Figure 4.1) from SNB-035. We predicted that erythrazole C will activate the arylhydrocarbon receptor (AhR) due to its indole properties and similar structure to ITE, which is a known AhR agonist. Initial investigation of erythrazole C as an agonist to AhR was confirmed via measurement of two AhR target genes (CYP1A1 and CYP1B1) in collaboration with the Roth lab at University of Texas, Southwestern. This suggested that erythrazole C may activate the AhR in a similar manner as ITE, therefore playing a key role in maintaining intestinal homeostasis. Erythrazole C appears to be a small peptide, comprised of the amino acids cysteine, and glycine, and an indole resembling tryptophan. These components drove genome mining efforts for a biosynthetic pathway.



erythrazole C



ITE

Figure 4.1 Structures of erythrazole C and ITE.

4.1.1 The aryl hydrocarbon receptor and ITE binding

Signaling of the aryl hydrocarbon receptor (AhR) plays a key role in mucosal immune responses, making it an important regulator in intestinal barrier function, intestinal immune cells, and intestinal homeostasis. The AhR is a ligand-inducible and ligand-dependent transcription factor that is part of the basic helix-loop-helix family[1–3]. Various types of intestinal immune cells express the AhR which can regulate the immune cells to act as a line of defense in the interaction between the external environment and host[3,4]. The AhR was originally identified as a mediator of environmental toxins such as polychlorinated dioxins, dibenzofurans, polychlorinated biphenyls (PCBs), and polycyclic aromatic hydrocarbons (PAHs)[5] . However, recent research has shown that AhR activation plays a role in a variety of important physiological roles in major organs and tissues, such as maintaining intestinal homeostasis[1–3,6]. Studies on AhR knockout animals have shown numerous developmental and functional abnormalities, highlighting the importance of AhR activation[7]. The AhR is characterized as a promiscuous receptor due to its ability to bind a variety of exogenous and endogenous ligands with varying structures and characteristics.

One of the first endogenous AhR ligands identified was ITE. ITE was isolated from a porcine lung and competes for binding with 2,3,7,8-

[3H]tetrachlorodibenzo-p-dioxin (TCDD), a known toxic ligand for the AhR. Upon discovery of ITE, radiolabeled ITE was used in a saturation binding assay to determine specific binding[5] (Figure 4.2). The Gasiewicz lab further evaluated the AhR activity of ITE and confirmed that it is a potent AhR agonist in cell extracts, cultured cells, and intact animals but does not induce the toxicity associated with other known xenobiotic ligands, such as TCDD[5]. Later studies demonstrated that ITE does ameliorate colitis in mouse models[8,9]. ITE has been shown to compete for AhR binding in human, murine, and fish AhRs, indicating that recognition of this ligand is highly conserved[7].

AhR ligands have been found to come from a variety of sources including from environmental toxins, dietary sources, and through metabolism of tryptophan by microbes in the gut microbiome. Additionally, many tryptophan derivatives, like erythrazole C and ITE, have been found to be produced by intestinal gut microbes to form other endogenous AhR ligands. While ITE was originally isolated in a pig lung, homologs of ITE were later isolated from *Arabidopsis thaliana*[10], a myxobacterial strain[11], and two marine *Pseudoalteromonas* strains[12]. This suggests that perhaps these indoles are in fact produced by bacteria in the mammalian microbiome, thereby maintaining intestinal homeostasis via AhR activation. Bacterial species in the gut microbiome play important roles in metabolism, immunity, development and behavior[6]. Due to the small concentration of ITE that was isolated from the porcine lung (20μg/35kg)[7], we hypothesize that this finding was due to some

contamination between organs and that it is actually produced microbially by a member of the host's microbiome.

Additionally, other studies have used AhR ligand producing microbes to ameliorate colitis in mouse models in the form of probiotics[13,14]. Considering the success of ITE in treating colitis in addition to the success of using other AhR ligand-producing microbes as probiotic treatments, we hypothesize that incorporating SNB-035 or another microorganism which produces erythrazole C into the intestinal microbiome will similarly reduce disease in colitis models. This presents an especially attractive potential as a treament if erythrazole C does in fact hold simliar binding properties to ITE. The ability of ITE to compete with toxic ligands (such as TCDD) also suggests added protective benefits of employing this ligand as a treatment. Observing the effect of this treatment in probiotic form rather than direct compound administration may provide a more safe and effective treatment for inflammatory bowel disorder (IBD) as the microbe would regulate how much of the ligand is produced, may involve less frequent dosing, and may have less side effects if the microbe naturally occurs in the microbiome. Existing treatments for inflammatory bowel disorders are based on broad immune suppression, causing a variety of unwanted side effects[13]. Therefore, there is an unmet need for effective treatments for intestinal inflammation without these unwanted side effects.

The objectives of this chapter were to characterize the biosynthesis of EC and identify a biosynthetic gene cluster, search for these gene clusters

within the human microbiome, determine if EC binds and activates the AhR, and ultimately test EC as a treatment for colitis in a mouse model. The overarching goal of this research was to expand biosynthetic knowledge, provide greater understanding of homeostatic regulation via microorganisms in the intestine, and to potentially provide evidence for a novel therapeutic for the treatment of inflammatory bowel disorders.

4.1.2 Proposed aims and experiments

There were two major purposes in studying EC; to identify and validate the biosynthetic pathway which is responsible for EC production and to test the effects of erythrazole C and erythrazole C-producing microbes *in vitro* and *in vivo*. In order to identify the proposed biosynthetic gene cluster (BGC), I first used genome mining and then aimed to validate this proposed BGC with knockouts and heterologous expression.

My first objective was to validate a BGC through knockout studies and herterologous expression. Determining the BGC of EC would not only shine light on biosynthetic logic but also would allow the detection of this BGC and therefore prevalence of EC production in the human microbiome. The NIH Human Microbiome Project (HMP1) contains sequenced genomic data from healthy human gut microbiomes which would allow me to search for this gene cluster using this database. Additionally the Integrative Human Microbiome Project (iHMP) contains data from different cohort studies of human conditions,

which would allow taking this a step further to look for correlations between EC producing microbes in the gut microbiome and disease states.

While there is strong evidence suggesting that erythrazole C will bind to the aryl hydrocarbon receptor, this needs to be confirmed both *in vitro* and *in vivo*. First, I planned to confirm whether purified erythrazole C can activate the AhR *in vitro*. To observe direct binding of EC to the AhR, I planned to employ a competition binding experiment with the high-affinity photoaffinity ligand (PAL) assay. This assay has previously been used to determine sepecific binding of other ligands to the AhR[15] and quantifies binding of a ligand to its receptor through decreasing detection of the photoaffinity ligand with increasing concentrations of ligand. This decrease implies that the increasing ligand concentration is able to compete for AhR binding with the PAL assay. This method enables assessing direct binding and determining an IC50 value[5].

To determine AhR activation, I planned to treat human colonic epithelial cell line, CaCo2, and a mouse AhR reporter cell line, Hepa1.1, with erythrazole C and use qPCR to measure the expression of Cyp1A1 and CYP1B1, genes which are under direct regulation of AhR signaling[2]. Since rodents and humans show different binding abilities and AhR activation to different ligands, it is important to make sure activation is consistent before using animal models. These assays would confirm the ability of erythrazole C to bind to the AhR, in addition to its activation due to increased expression of key genes in the canonical AhR signaling pathway. Next, I planned to repeat this study with

cultured SNB-035 as to determine if the microbe itself can activate the AhR. Regardless of the properties of erythrazole C, SNB-035 should be able to activate the AhR as it also produces ITE. As it has been shown that bacteria can be directly used to activate the AhR[13,14], we believe that SNB-035 should have similar effects.

Ultimately, the goal of these experiments was to to administer SNB-035 and/or another erythrazole C-producing microbes identified to live animals to determine AhR binding and affect on colitis. AhR activation will be monitored in healthy mice when treated with these microbes compared to control groups. To test the effect of EC in a disease state, I planned to use a dextran sodium sulfate (DSS) colitis model to induce disease and subsequently treat it using SNB-035 (or other erythrazole C-producing microbes). These objectives would work towards characterizing this molecule's biological significance and to provide insight into a new method of treatment for inflammatory bowel disorders. However, for reasons described below, only the first aim of determining the BGC for EC was explored.

4.2 Detection of erythrazole C in SNB-035 extract

4.2.1 Culture and extraction

Bacterium SNB-035 was cultured in 5, 1L cultures in fernbach flasks. SNB-035 was grown in seawater based medium (10g starch, 4g yeast extract, 2g peptone, 1g $CaCO_3$, 40mg $Fe_2(SO_4)3*4H_2O$, 100mg KBr) and shaken at

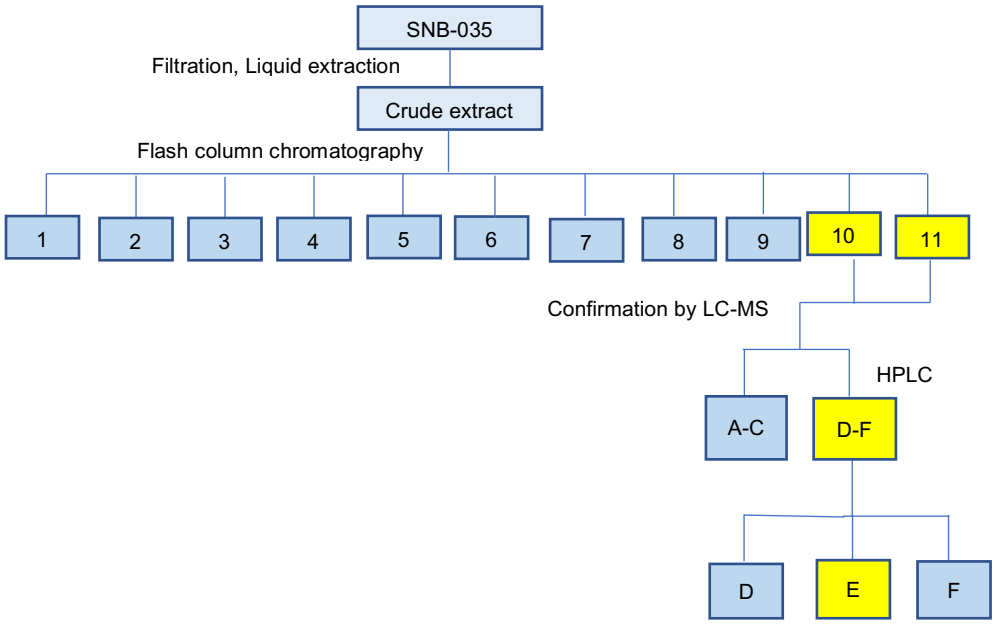
200 rpm at 30°C. After 5 days of cultivation, sterilized XAD-7-HP resin was added to the cultures and left shaking for 2 hours. The resin was then filtered through cheesecloth and submerged in acetone to shake overnight. The acetone soluble layer was then dried using a rotary evaporator. This resulted in 506.9mg of crude extract.

4.2.2 Isolation of erythrazole C

Crude extract was resuspended in MeOH and dried down with equal parts celite. The mixture was then fractionated by reversed phase flash column chromatography using Teledyne ISCO (Redisep column: C18 30g Gold, 35 mL/min, UV 200nm - 360 nm) using a gradient solvent system of MeOH and H_2O, from 5% to 100% of MeOH over 20 minutes. 25 fractions were collected and dried for LC-MS analysis. ElectroSpray Ionization (ESI) (G6130B Quadrupole LC/MS, Agilent 1200) was used to identify the presence of EC where *m/z* 330.1 $[M+H]^+$ with a retention time of 4.46 minutes. This which correlated to the exact mass of EC, which is 329.05. Fractions 10, 11 and 14 showed this *m/z* though fraction 14 showed significantly lower signal so fractions 10 and 11 were combined for further analysis. Fractions were then run on the Agilent HPLC (Semiprep C18, 250 x 10.0 mm, 1.5 mL/minute, UV=254nm). Peaks were collected and rerun for LC-MS analysis, which showed that fractions 4, 5 and 6 contained the mass for EC. These fractions

were then further purified by reverse phase HPLC (Phenomenex Luna, Phenyl-Hexyl, 250 x 10.0 mm, 2.5mL/min, 5um, UV=254nm), where 3 major peaks were collected from each (A-C and D-F, respectively) and it was determined that fraction E contained EC by LC-MS analysis. This isolation scheme can be seen in Figure 4.2.

A



B

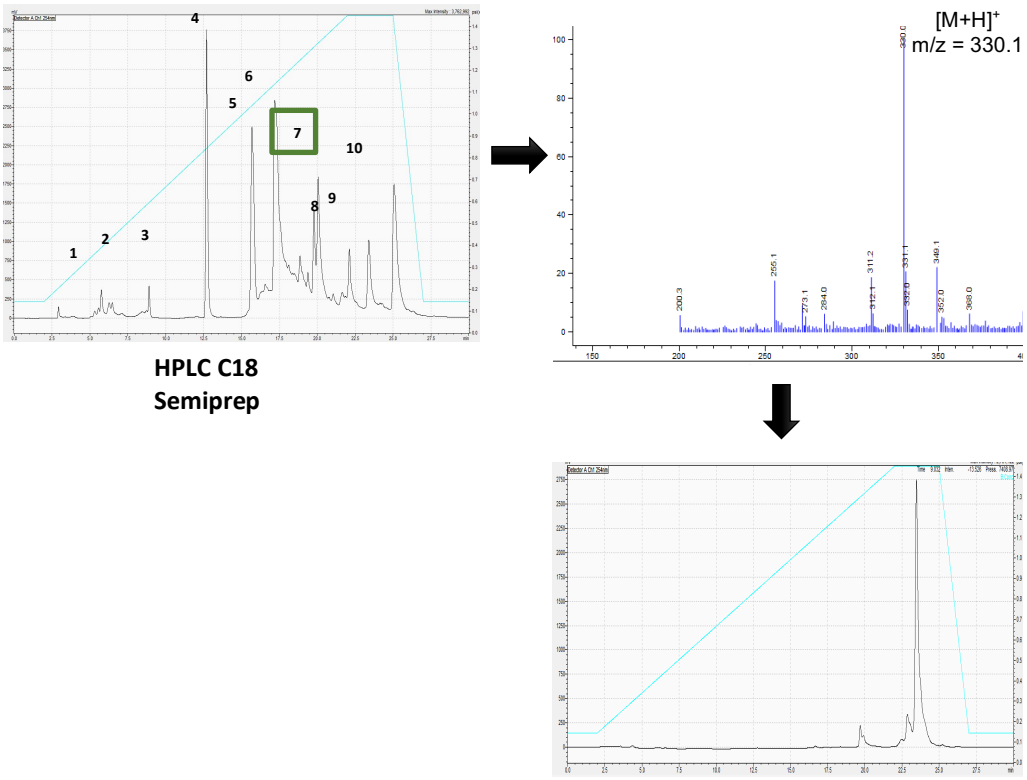


Figure 4.2 Isolation scheme for erythrazole C. Crude extract from SNB-035 was separated with Teledyne ISCO flash column chromatography. Then detection of EC was performed using LC-MS. Further purification was done with HPLC using a phenyl hexyl column. **A.** Overall isolation scheme with highlighted fractions which contain the *m/z* for EC. **B.** Chromatograms and mass spectrometry feature for EC from isolation work-through.

4.3 Proposed biosynthetic pathway of erythrazole C

Erythrazole C is comprised of the amino acids tryptophan, cysteine and glycine. However, this structure in combination with the genomic information of SNB-035 suggest that EC is not derived from an NRPS or RiPP. Therefore, we

were interested in identifying the genes and enzymes responsible for this amino acid derived natural product. The genome of SNB-035 was sequenced, and we employed standard genome mining efforts using computational programs such as antiSMASH. AntiSMASH did not indicate any biosynthetic gene clusters which correlated with EC. However, Rajniak et al. determined the biosynthesis of a structurally similar compound (B2) in *Arabidopsis thaliana* (Figure 4.3)[10].

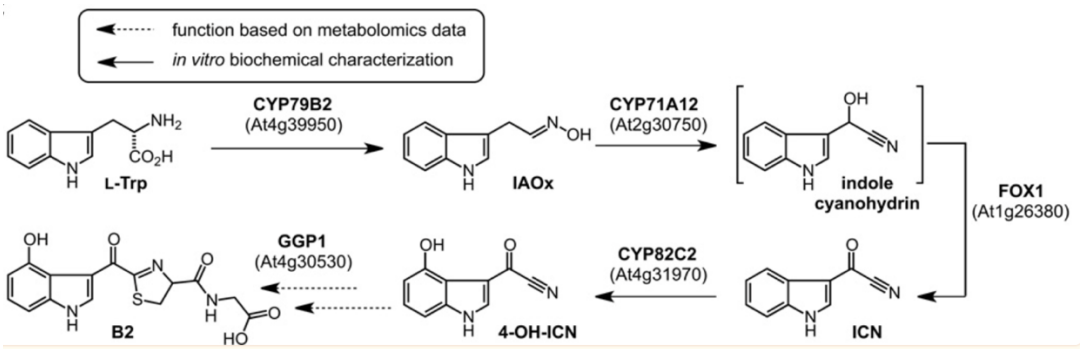


Figure 4.3 Proposed biosynthetic pathway of B2 (reduced and hydroxylated EC) in Arabidopsis thaliana. Figure reproduced with permission from Springer Nature from Rajniak, J.; Barco, B.; Clay, N.; and Sattely, E. A new cryogenic metabolite in Arabidopsis required for inducible pathogen defense. *Nature* **2015**, 525(7569), 376-379. https://doi.org/10.1038/nature14907.

Rajniak et al. showed that several cytochrome P450s and a flavin-dependent oxidoreductase (FOX1) are responsible for the biosynthesis of 4-hydroxyindole-3-carbonyl nitrile (4-OH-ICN), a key intermediate of B2. We searched for similar genes described in the SNB-035 genome, with particular attention to cytochrome P450's as there are several involved in this pathway.

However, we were not able to detect clustering of similar genes in SNB-035, and it is unclear how well that this biosynthesis pathway may be conserved in plants and bacteria. Therefore, I proposed a BGC with different genes but that still suggests some relationship to the structure of EC. By examining the sequenced genome, I proposed a BGC which contains a thiazole synthase, cysteine desulferase, and a glycine ligase which are present within a ~20kb region of the SNB-035 genome (Figure 4.4). Each of these genes are capable of similar chemistry as we would expect to occur in the biosynthesis of erythrazole C. Therefore, I went forward to attempt to validate this pathway by using molecular biology techniques such as making knockouts of genes and heterologous protein expression of enzymes in the proposed pathway.

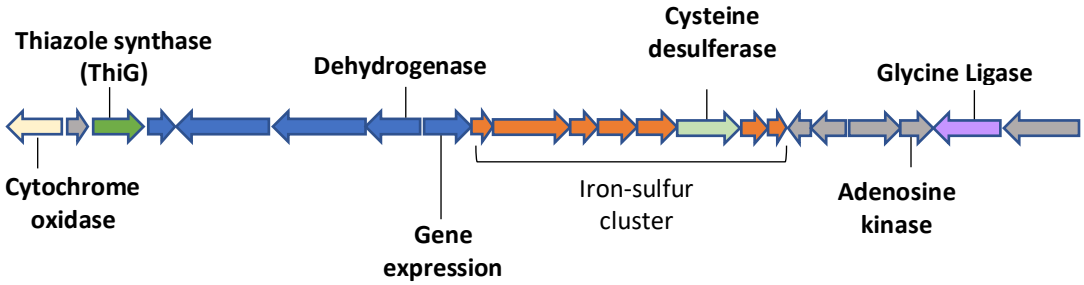


Figure 4.4 Proposed biosynthetic gene cluster for erythrazole C.

4.3.1 Homologous recombination for knockout of putative biosynthetic genes

Genes can be knocked out of bacterial genomes utilizing a suicide vector and homologous recombination. Using a protocol based on the Link et al. article[16] and modified by Dr. Chad Saltikov's lab at UCSC, I used the suicide vector, pLO3[17] (provided by Dr. Oliver Lenz, Berlin Institute of Technology). Conjugation of this vector into a host organism can cause gene deletion (Figure 4.6). Surrounding regions of the desired knockout (H1 and H2) are cloned into the vector, which is then incorporated into the host genome using conjugation. The homology of H1 and H2 in the genome will cause either a reversion to wild-type or adoption of the new sequence, which lacks the target gene between H1 and H2. The latter results in a successful knockout[18].

Therefore, I first amplified 1kb upstream and downstream of thiazole synthase (ThiG) and used crossover PCR to amplify the regions into one, 2kb fragment (referred to as ThsX). This fragment was then cloned into pLO3 and transformed into DH5λpir *E. coli* cells. The suicide vector containing ThsX was then extracted and cloned into a diaminopimelic acid (DAP) auxotroph *E. coli* conjugation strain, WM3064. This strain containing pLO3 + ThsX was then conjugated into SNB-035. Successful conjugation was confirmed by tetracycline selection and the exclusion of DAP from agar plates to prevent the *E. coli* from growing. Next, sucrose selection was used to remove the suicide vector due to the sucrose lethality gene, SacB, in pLO3, thereby inducing

homologous recombination. Individual colonies forming on plates with sucrose indicated that the colony no longer contained the SacB gene, and therefore had recombined. These colonies were then selected for polymerase chain reaction (PCR) analysis, which involved running a PCR of the thiazole synthase gene. Results of the PCR were sent for Sanger Sequencing.

During my experiments, no successful PCR products indicated loss of the thiazole synthase gene. The gain of tetracycline resistance indicated that the conjugation of the vector into SNB-035 was successful. However, the target gene being present still means that no recombination had occurred post-sucrose exposure. These colonies would have reverted to their wild-type sequence rather than recombining - one of two possible outcomes with homologous recombination (Figure 4.6). If these had been successful, the knockout strain would be cultured and the extract will be tested for the presence of erythrazole C by mass using LC-MS.

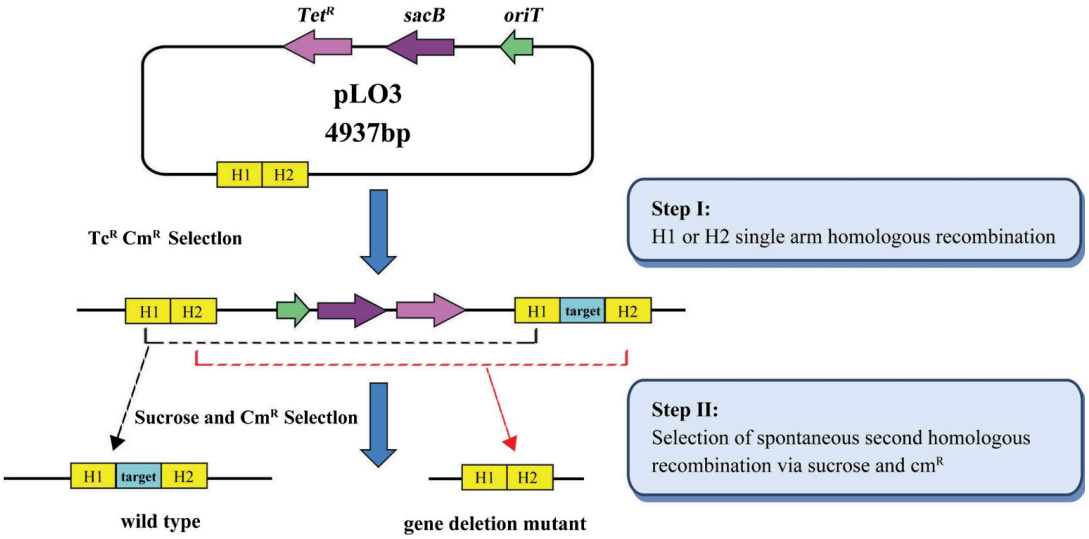


Figure 4.5 Homologous recombination using suicide vector, pLO3. Figure reproduced with permission from Elsevier from Wu, M. et al. Enhancement of transparents hydrogel sanxan production in *Sphingomonas sanxanigenens* NX-2 via rational and random gene manipulation. Carbohydrate Polymers **2018**, 189, 210-217. https://doi.org/10.1016/j.carbpol.2018.02.027.

4.3.2 Glycine Ligase as a hook for BGC identification

Synthetic routes of erythrazole C and its precursors were described in Rajniak et al. 2015[10]. Of special interest to us, was the intermediate A5 which only lacks the glycine moiety. This intermediate could be harnessed to see if the glycine ligase of the proposed BGC for EC does in fact incorporate the glycine moiety of EC. The confirmation of this enzyme's role in the biosynthesis of EC would help to confirm the gene cluster. To examine this, a postdoctoral scholar in the lab, Dr. Rahul Shingare, synthesized the intermediate A5 and I isolated the glycine ligase enzyme. These would then be combined in an *in*

vitro assay to determine the role of glycine ligase. This procedure is summarized in Figure 4.6.

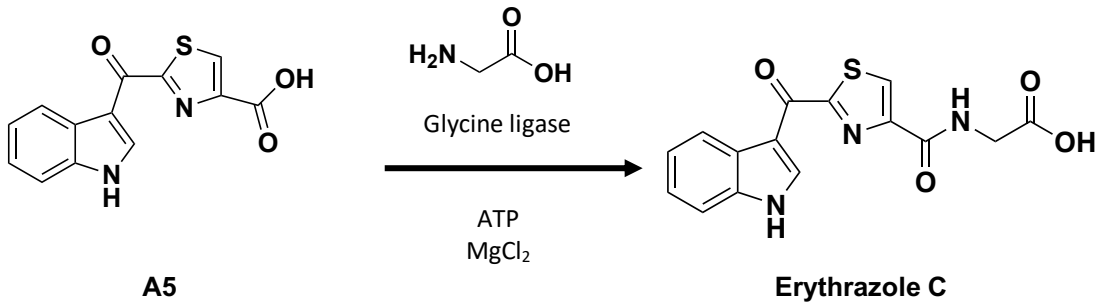


Figure 4.6 Scheme for *in vitro* glycine ligase assay. Intermediate A5 will be synthesized by Dr. Rahul Shingare. Then, glycine ligase will be heterologously expressed and purified. Our hypothesis is that the purified glycine ligase, with the addition of glycine and cofactors (ATP, $MgCl_2$) will act upon A5 to form erythrazole C.

First, Dr. Rahul Shingare synthesized both erythrazole C and intermediate, A5 (Figure 4.7). The synthesized erythrazole C would serve as a control in the *in vitro* assay. Synthesis of erythrazole C was started from the condensation of cysteine methyl ester hydrochloride and indole-3-carbonyl nitrile (ICN) **1** to afford thiazolinederivative **2** in 77% yield.[19] Compound **2** on treatment with bromotrichloromethane in combination with 1,5-diazabicyclo[5.4.0]undecane (DBU) at -20 °C in methylene dichloride undergoes dehydrogenations to give thiazole **3** as a sole product in excellent yield.[20] The methyl ester was hydrolyzed using lithium hydroxide to afford **A5** in quantitative yields. Finally, A5 was coupled with an ethyl ester of glycine under standard 1-ethyl-3-(3-diamethylaminopropyl)carbodiimide (EDC)

coupling conditions followed by ester hydrolysis giving erythrazole C in good yield. Total yield resulted in 63 mg A5 and 50 mg erythrazole C. Confirmation of these structures was determined using NMR (Figures 4.11-4.14).

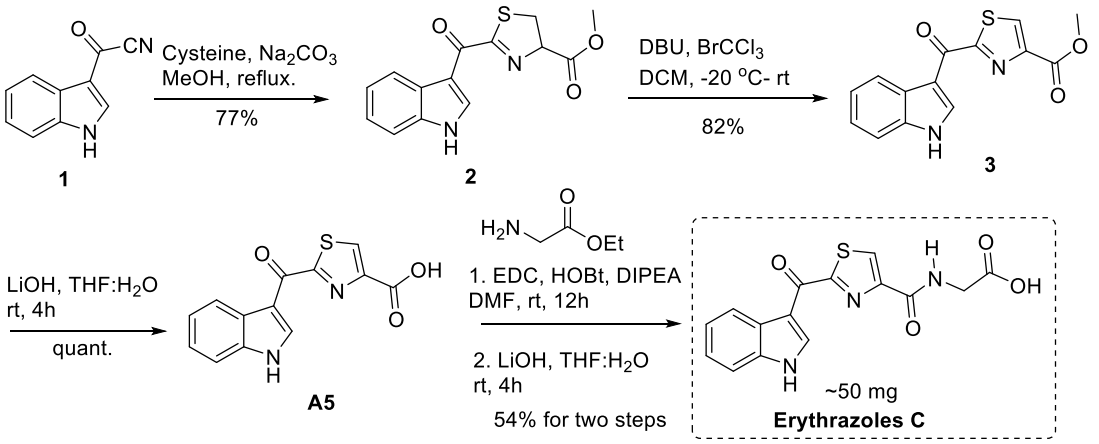
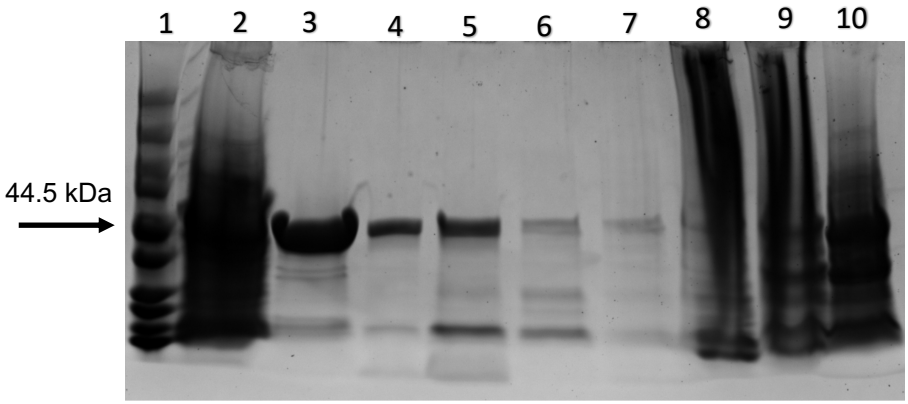


Figure 4.7 Synthetic route of the A5 intermediate for *in vitro* enzyme assay and for erythrazole C.

Next, I heterologously expressed the glycine ligase enzyme according to standard procedures in *E. coli* strain BL21 (experimental). Fast protein liquid chromatography (FPLC) was then used to purify the protein using a 6x-His tag on a Ni-NTA column. Fractions were collected and analyzed on a 10% SDS-PAGE gel, which showed that glycine ligase was successfully eluted as seen by a strong band corresponding to the size of the protein (44.5 kDa) (Figure 4.8). However, we noticed that the majority of the protein eluted in the wash waste. This indicates that the His-tagged protein was easily eluted from the

column at 10% buffer B. Since the waste was saved we were able to still collect the eluted protein.

Fractions containing glycine ligase were pooled and concentrated to 3.5 mL. These were then desalted on a PD-10 column. Bradford assay was performed to determine protein concentration, which was 24.7 mg/mL in 3.5 mL. This showed that the 2 L growth had yielded a total of 86.45 mg of glycine ligase (43.225 mg/L). Purified glycine ligase was then stored in GF buffer (50 mM HEPES pH 8, 300 mM KCl), aliquoted, and flash frozen in 10% glycerol. *In vitro* enzyme assays were performed with the purified glycine ligase, synthesized A5 intermediate, glycine, ATP and $MgCl_2$ as the cofactor for the enzyme. Two different conditions (experimental), neither of which resulted in detection of EC formation by LC-MS.



1. Ladder
2. Load waste
3. Wash waste
4. F4
5. F7
6. F10
7. F16
8. Lysed cells (pre-spin)
9. Pellet
10. Pre-fplc pellet

Figure 4.8 SDS-Page (10% acrylamide) gel image of glycine ligase protein preparation. Molecular weight of glycine ligase is 44.5 kDa. A strong, corresponding band can be seen in the wash waste, and fractions up to fraction 7. Fractions 1-8 were pooled in addition to the wash waste.

4.4 Crawford Lab publishes non-enzymatic biosynthesis of erythrazole C

In June 2020, the Crawford lab at Yale University published an article identifying a family of metabolites called the indolokines, including indolokine A3 (5)[21], which has an identical structure to that which we were referring to as erythrazole C (Figure 4.9). The authors were able to determine the biosynthetic

pathway and show AhR binding of these molecules, overall showing that the indolokines (including erythrazole C) are widely produced by various organisms as a stress response and have similar AhR activity as ITE (Figure 4.11B).

They detected these molecules in *E. coli* by comparing metabolite production in a BW25113 (wild-type) strain and a transaminase (*aspC*) mutant strain. *AspC* is a pyridoxal-5' phosphate (PLP)-dependent amino acid transaminase, which converts amino acids to their α-ketoacid forms. This gene also has been linked to a core antibiotic stress response. A *tyrB* mutant was also used in these experiments as it is also a transaminase which can accept similar substrates as *aspC*. These mutants showed decreased production in several metabolites with indole-like chromophores, which were then classified as the indolokines.

These compounds were then found as being produced by a variety of other bacteria and in mouse fecal samples, suggesting their production by bacteria in the gut microbiome. They were also found in the commensalistic *E. coli* strain *E. coli* Nissle 1917, which has been used to treat IBD[22]. These results align with our hypothesis that EC producing microbes could be used as probiotic treatment for IBD.

The Crawford group determined the biosynthetic origin of these molecules starts with a precursor molecule, indole-3-pyruvic acid (I3P). I3P is produced by PLP-dependent enzymes, *aspC*- and *tyrB*- mediated transamination of L-Trp. I3P can then spontaneously react with **L**-Cys or **L**-Cys-

Gly to produce the indolokines, including indolokine A3 (5) or erythrazole C (Figure 4.10A). This proposal suggests a non-enzymatic coupling pathway, which the authors explain is consistent with their widespread production in various bacteria. The authors were able to support the non-enzymatic reaction by incubating I3P in media in the absence of bacteria. The role of the PLP-dependent enzymes, *aspC*- and *tyrB*- was supported by observing the loss of production in knockout strains.

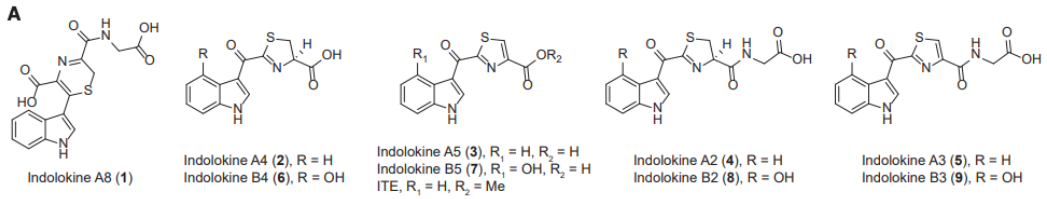


Figure 4.9 Identification of indolokines including erythrazole C (5). Reprinted from *Cell Chemical Biology*, 27(6), Kim et al., Cellular Stress Upregulates Indole Signaling Metabolites in Escherichia coli, 698-707, **2020** with permission from Elsevier. https://doi.org/10.1016/j.chembiol.2020.03.003.

Kim et al. also evaluated the AhR activity of the indolokines on an AhR reporter cell line using a beetle luciferase gene downstream of dioxin/xenobiotic response elements (Figure 4.10B). As predicted, all of the indolokines showed AhR activation at low concentrations. However, the indolokine whose structure matched that of EC (5) had moderate AhR activity, with less activity than ITE and indolokine A4 but higher activity than the other indolokines.

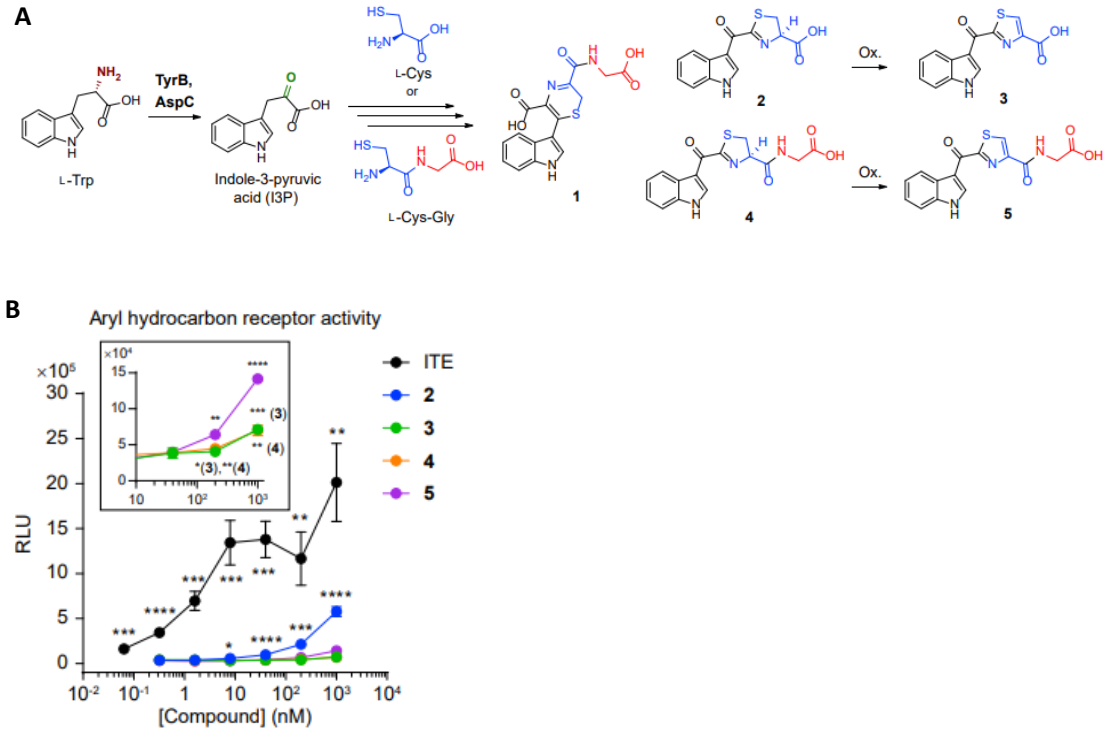


Figure 4.10 Biosynthesis of indolokines and their AhR activity. A. biosynthesis of the indolokines involves TyrB or AspC transamination of tryptophan to form intermediate I3P. I3P can then spontaneously react with L-Cys or L-Cys-Gly to form indolokines A8, A4, and A2, which can be oxidized to form indolokines A3 and A5. **B.** AhR reporter cell line was treated with the indolokines, showing the varying AhR activity of the indolokines and ITE. Reprinted from *Cell Chemical Biology*, 27(6), Kim et al., Cellular Stress Upregulates Indole Signaling Metabolites in Escherichia coli, 698-707, **2020** with permission from Elsevier. https://doi.org/10.1016/j.chembiol.2020.03.003.

4.5 Discussion

While the publication of the activity and biosynthesis of the indolokines was disappointing, this work did confirm many of our hypotheses. This work proved that we were incorrect about our proposed biosynthesis of the molecule.

However, we correctly predicted that EC would bind to the AhR though ITE and indolokine A4 (2) showed greater binding ability (Figure 4.10B). Our hypothesis that these molecules were produced in the microbiome to help maintain homeostasis was also proven correct. While this resulted in the termination of this project, it was satisfying to see that some of hypotheses were correct and were proved using some of the methods which we proposed.

The work described here showed success in conjugating a suicide vector into SNB-035 which may be applicable for future work. Additionally, we successfully expressed and purified the enzyme, glycine ligase and synthesized an EC intermediate and EC itself. After seeing this publication, we changed directions to looking at erythrazoles A and B. Erythrazoles A-B have some overlapping structural similarities to EC and are produced by the same organism, therefore some of the strategies described here could be applicable in future work.

4.6 Chapter four references

(1) Mulero-Navarro, S.; Fernandez-Salguero, M. New Trends in Aryl Hydrocarbon Receptor Biology. *Front Cell Dev Biol* **2016**, *4*, 45.

(2) Shinde, R.; McGaha, T. L. The Aryl Hydrocarbon Receptor: Connecting Immunity to the Microenvironment. *Trends Immunol.* **2018**, *39* (12), 1005–1020.

(3) Lamas, B. Aryl Hydrocarbon Receptor and Intestinal Immunity. *Mucosal Immunol.* **2018**, *11* (4), 1024–1038.

(4) Esser, C.; Rannug, A.; Stockinger, B. The Aryl Hydrocarbon Receptor in Immunity. *Trends Immunol.* **2009**, *30* (9), 447–454.

(5) Henry, E. C. A Potential Endogenous Ligand for the Aryl Hydrocarbon Receptor Has Potent Agonist Activity in Vitro and in Vivo. *Arch Biochem Biophys* **2006**, *450* (1), 67–77.

(6) Levy, M. Microbiome, Metabolites and Host Immunity. *Curr. Opin. Microbiol.* **2017**, *35*, 8–15.

(7) Song, J. A Ligand for the Aryl Hydrocarbon Receptor Isolated from Lung. *Proc Natl Acad Sci USA* **2002**, *99* (23), 14694–14699.

(8) Goettel, J. A. AHR Activation Is Protective against Colitis Driven by T Cells in Humanized Mice. *Cell Rep.* **2016**, *17* (5), 1318–1329.

(9) Abron, J. D.; Singh, N. P.; Mishra, M. K.; Price, R. L.; Nagarkatti, M.; Nagarkatti, P. S.; Singh, U. P. An Endogenous Aryl Hydrocarbon Receptor Ligand, ITE, Induces Regulatory T Cells and Ameliorates Experimental Colitis. *Am. J. Physiol.-Gastrointest. Liver Physiol.* **2018**, *315* (2), G220–G230. https://doi.org/10.1152/ajpgi.00413.2017.

(10) Rajniak, J. A New Cyanogenic Metabolite in Arabidopsis Required for Inducible Pathogen Defence. *Nature* **2015**, *525* (7569), 376–379.

(11) Jansen, R. Indothiazinone, an Indolyl Thiazolyl Ketone from a Novel Myxobacterium Belonging to the Sorangiineae. *J. Nat. Prod.* **2014**, *77* (4), 1054–1060.

(12) Guo, H. Natural Products and Morphogenic Activity of Gamma-Proteobacteria Associated with the Marine Hydroid Polyp Hydractinia Echinata. *Bioorg. Med. Chem.* **2017**, *25* (22), 6088–6097.

(13) Fukumoto, S. Identification of a Probiotic Bacteria-Derived Activator of the Aryl Hydrocarbon Receptor That Inhibits Colitis. *Immunol. Cell Biol.* **2014**, *92* (5), 460–465.

(14) Takamura, T.; Harama, D.; Fukumoto, S.; Nakamura, Y.; Shimokawa, N.; Ishimaru, K.; Ikegami, S.; Makino, S.; Kitamura, M.; Nakao, A. Lactobacillus Bulgaricus OLL1181 Activates the Aryl Hydrocarbon Receptor Pathway and Inhibits Colitis. *Immunol. Cell Biol.* **2011**, *89* (7), 817–822.

(15) Muku, G. E.; Murray, I. A.; Espín, J. C.; Perdew, G. H. Urolithin A Is a Dietary Microbiota-Derived Human Aryl Hydrocarbon Receptor Antagonist. *Metabolites* **2018**, *8* (4). https://doi.org/10.3390/metabo8040086.

(16) Link, A. J. Methods for Generating Precise Deletions and Insertions in the Genome of Wild-Type Escherichia Coli: Application to Open Reading Frame Characterization. *J. Bacteriol.* **1997**, *179* (20), 6228–6237.

(17) Lenz, O.; Friedrich, B. A Novel Multicomponent Regulatory System Mediates H2 Sensing in Alcaligenes Eutrophus. *Proc Natl Acad Sci USA* **1998**, *95* (21), 12474–12479.

(18) Wu, M.; Shi, Z.; Ma, T. Enhancement of Transparent Hydrogel Sanxan Production in Sphingomonas Sanxanigenens NX02 via Rational and Random Gene Manipulation. *Carbohydr. Polym.* **2018**, *189*, 210–217.

(19) Tan, F.; Shi, B.; Li, J.; Wu, W.; Zhang, J. Design and Synthesis of New 2-Aryl-4,5-Dihydro-Thiazole Analogues: In Vitro Antibacterial Activities and Preliminary Mechanism of Action. *Molecules* **2015**, *20*, 20118–20130.

(20) Williams, D. R.; Lowder, P. D.; Gu, Y. G.; Brooks, D. A. Studies of Mild Dehydrogenations in Heterocyclic Systems. *Tetrahedron Lett.* **1997**, *38*, 331–334.

(21) Kim, C. sub; Li, J.-H.; Barco, B.; Piizi, G.; Clay, N. K.; Crawford, J. M. Cellular Stress Upregulates Indole Signaling Metabolites in Escherichia Coli. *Cell Chem. Biol.* **2020**, *27* (6), 698–707.

(22) Schultz, M. Clinical Use of E. Coli Nissle 1917 in Inflammatory Bowel Disease. *Inflamm. Bowel Dis.* **2008**, *14* (7), 1012–1018. https://doi.org/10.1002/ibd.20377.

4.7 NMR Spectra

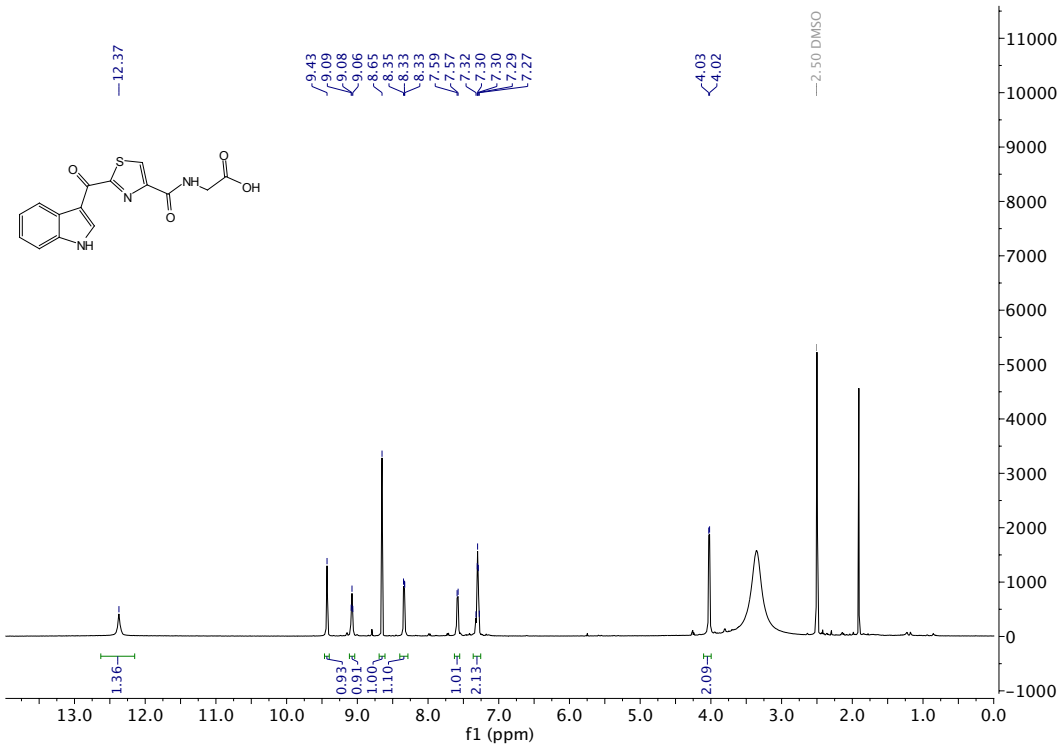


Figure 4.11 1H NMR of synthetic erythrazole C. MeOD, 500MHz.

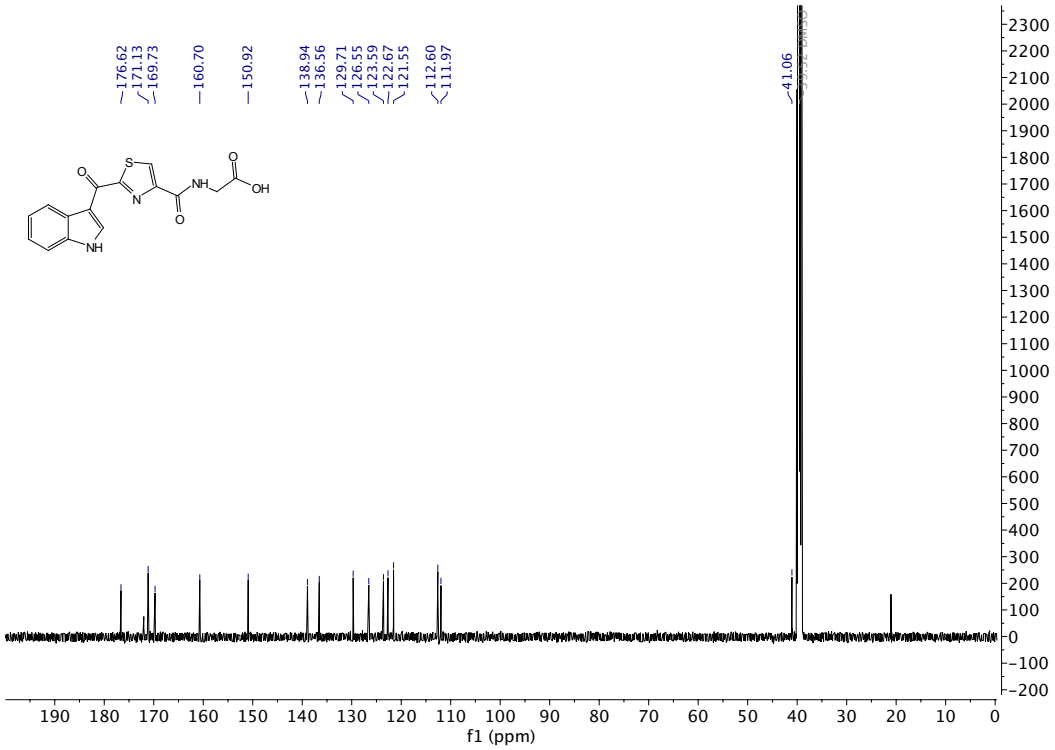


Figure 4.12 ^{13}C NMR of synthetic erythrazole C. MeOD, 500MHz.

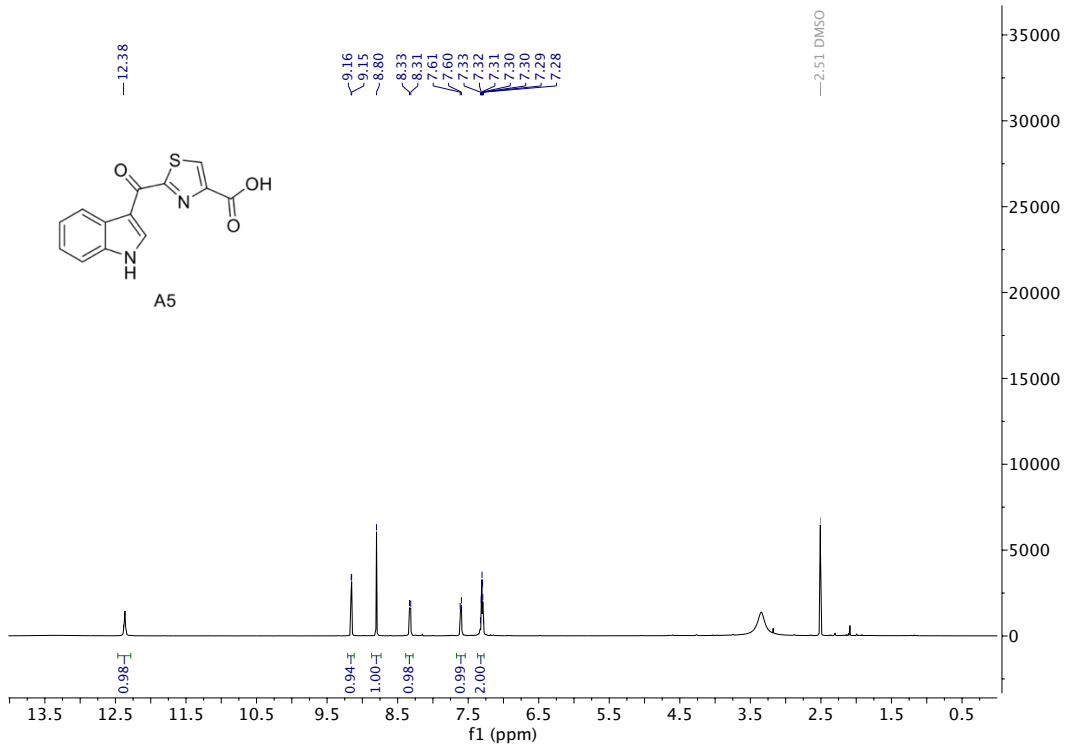


Figure 4.13 ^{1}H NMR of synthetic A5 intermediate for glycine ligase in vitro assays. MeOD, 500MHz.

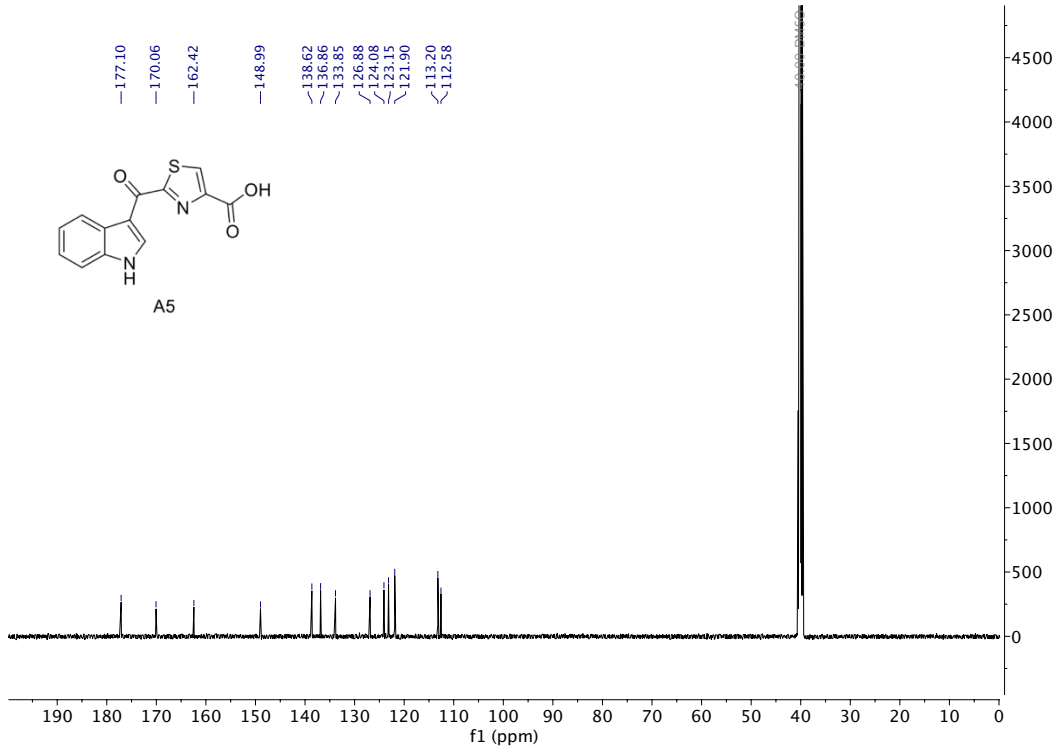


Figure 4.14 ^{13}C NMR of synthetic A5 intermediate for glycine ligase in vitro assays. MeOD, 500MHz.

EXPERIMENTAL

Cultivation and extraction

Bacterium SNB-035 was cultured in the seawater based medium, A1BFe+C (10g starch, 4g yeast extract, 2g peptone, 1g $CaCO_3$, 40mg $Fe_2(SO_4)3*4H_2O$, 100 mg KBr in 75% seawater). For large cultures (>1L), after 5 days of shaking at 200 rpm at 30°C, sterilized XAD-7-HP resin was added to the cultures and left shaking for 2 hours. The resin was then filtered through cheesecloth to remove the aqueous material. The collected resin was then submerged in acetone and left shaking overnight at 30°C, 200rpm. The acetone soluble material was then poured over a Whatman filter in a funnel and collected in a round bottom flask. This material was then dried using a rotary evaporator.

For smaller cultures (≤1L), a liquid-liquid partitioning method was used. Culture was added to a separatory funnel and an equal volume of ethyl acetate was added. The separatory funnel was then closed and shaken vigorously for ~30 seconds, periodically opening the valve to release pressure. The layers were then allowed to separate and the organic layer was collected in a round bottom flask. This was repeated with the aqueous culture 3 times and the organic material was then dried with a rotary evaporator.

Glycine ligase purification

Glycine ligase was amplified from the SNB-035 genome and cloned with a 6x-His tag into a pet-28a vector using Gibson assembly in DH5α *E. coli* cells.

This vector (pet28a+6x-His tag+GlyLi) was then transformed into BL21 E. coli for heterologous protein expression and purification. Then, the BL21 cells containing the vector were grown at a 2 liter scale for purification of glycine ligase. Purification was performed using fast protein liquid chromatography (FPLC) with a Ni-NTA column. After loading the sample onto the column, the column was equilibrated with 90% Ni buffer A (20 mM Tris pH 8, 1M NaCl, 30 mM imidazole) and 10% Ni buffer B (20 mM Tris pH 8, 1M NaCl, 150 mM imidazole). Buffer B was then run over the column at a gradient from 10% to 100% over 50 minutes at 2 mL/minute. The majority of the protein was eluted in the waste, indicating that a lower imidazole concentration in buffer B should be used in the future.

Glycine ligase enzyme assays:

Glycine ligase enzyme assays were performed with two different conditions:

1. The following reagents were combined and brought to a volume of 100 µL HEPES buffer (pH8): $MgCl_2$ (20mM), Glycine (5mM), ATP (1.5 mM), Glycine ligase (5 µM), A5 precursor (20 mM). 30 µL of the total reaction was removed and quenched for analysis at 2 hours, 5 hours, and 24 hours. Quenching involved adding 30 µL ACN and centrifuging at maximum speed for 10 minutes. Supernatant was removed and analyzed via LC-MS for confirmation of the final product.

2. The following reagents were combined and brought to a volume of 100 µL Tris-HCl: $MgCl_2$ (30mM), Glycine (30mM), ATP (10 mM), Glycine ligase (1 µM), A5 precursor (20 mM). Assays were performed at 30℃, 37℃ and room temperature. Time points were taken at 2 hours, 5 hours, and 24 hours. Quenching involved adding 30 µL ACN and centrifuging at maximum speed for 10 minutes. Supernatant was removed and analyzed via LC-MS for confirmation of the final product.

LC-HRMS/MS

Erythrazoles were detected from crude extracts of SNB-035 using either an Electrospray Ionization (ESI) Agilent 1200 or Orbitrap. The orbitrap system used Dionex UltiMate 3000 UHPLC paired with a Thermo Orbitrap Velos Pro LC-MS (Thermo Scientific). Extracts were resuspended in methanol at a concentration of 1 mg/mL and spun down to remove the insoluble material. 10-20µL of this solution was then injected and analyzed in negative mode. The chromatography method involved a mobile phase of 10 to 100% acetonitrile in water containing 0.1% formic acid for 10 minutes at 0.5mL/minute. We used a reverse-phase Luna Omega, 1.6µm C_{18} 100 Å, 50 x 2.1 mm column (Phenomenex). Excalibur software was used for instrument operation and data collection. Data was collected in centroid mode using an FTMS analyzer with collision energy set to 35.0 V.

Single ion monitoring was performed on the same instrument with the same column and LC method. Ions monitored were set to 585.15-586.15, 588.15-589.15, 611.15-612.15, and 614.15-615.15. Data was collected in centroid mode using an IonTrap analyzer with collision energy set to 35.0 V.

Low resolution ESI mass spectra were collected using an Agilent 1200 series with a reversed-phase Phenomenex Kinetex, 100mm x 2.1 mm, 2.6 um column. The mobile phase involved a gradient from 10-100% acetonitrile in water with 0.1% formic acid over 8 minutes at 0.3 mL/minute. 100% acetonitrile was held for 2 minutes before decreasing back to 10%. OpenLab software was used for instrument operation and data collection.

Stable isotope labeling

All stable isotope labeling studies were performed with the following conditions. Starter cultures were inoculated with 1 colony of SNB-035 and was grown in 50 mL seawater based medium, A1BFe+C. and shaken at 200 rpm at 30°C for 2 days. After 2 days, 250 µL of starter culture was used to inoculate a new flask containing fresh, sterile 50 mL of A1BFe+C media. Labeled material was dissolved in distilled water and immediately after inoculation of the bacteria was sterile filtered with a 0.45µm filter into the culture media for a final concentration of 1mM. These cultures were then shaken at 200 rpm at 30°C for 5 days before being extracted.

Extraction was performed using a 1:1 volume of ethyl acetate according to procedures described above. This yielded 20-30mg of crude extract per 50 mL culture. The crude extract was then dissolved in 1mg/mL methanol and centrifuged to remove any particulate matter before being analyzed using the LC-HRMS/MS method described above. Incorporation of stable isotope labeled compounds was confirmed by the shift of the m/z feature at 585.26 (EA) and 611.28 (EB) and by the fragmentation patterns of those shifted features.

CRISPRi vector cloning

The sgRNA spacers have only one target in the genome and have high efficiency. To design guides for the target genes, I used the code from https://github.com/ryandward/sgrna_design. This script designs spacers for the entire genome at once and determines guide specificity by taking into account the dCas9-sgRNA complex DNA-binding preferences[1,2]. This code produced various guides for each gene in the SNB-035 genome, all with the recommended specificity score of 39 and located on the anti-sense strand of DNA. Output included a gene identifier, offset, sequence, strand direction, and specificity score. sgRNA offset is defined as the distance between the 5' end of a target gene and the start of the sgRNA. Recommended offset of sgRNA's is -4 to 20bp[3]. While there were not sgRNA's recommended for all of the genes that I wished to target, I tried to choose guides that were as close to this range

as possible. In many cases I chose more than once guide to increase probability of successful targeting. Oligos were designed according to the sgRNA's chosen for each target gene (Table 3.4) with BsaI sticky ends and then were annealed and cloned into the BsaI cut site of the pJMP1339 plasmid. The result is the spacer located in the vector flanked by the sgRNA and lac operator. These components are then flanked by the Tn7 elements to allow them to be transposed by the transposase into the host chromosome. Vectors containing guides were transformed into DH5a cells (NEB) and successful cloning was confirmed with ampicillin selection and PCR.

CRISPRi vector conjugation

The constructs in addition to the Tn7 transposase vector (pJMP1039) were extracted with QIAprep Spin Miniprep Kit (#27104) and transformed into the diaminopimelic acid (DAP) auxotroph strain, *E. coli* WM6026. Conjugation was then performed between the two donor strains (*E. coli* WM6026 harboring either pJMP1039 or pJMP1339 with customized sgRNA spacers) and SNB-035. SNB-035 was grown in A1BFe+C media at 30°C for 2 days prior to conjugation. The *E. coli* donor strains were grown overnight at 37°C in LB broth supplemented with 300μM DAP and 100μg/mL ampicillin. After the cultures grew, 200μL of SNB-035 culture, 100μL of *E. coli* harboring pJMP1339 (transposon), and 100μL of *E. coli* harboring pJMP1039 (transposase) were added to 600μL of A1BFe+C media for a final volume of 1000 μL in a 2mL

Eppendorf tube. The tubes were then centrifuged for 2 minutes at 7000 x g. Supernatant was removed and the cells were resuspended in 1mL A1BFe+C prior to centrifugation at 6500 x g. Low centrifuge speeds are necessary to prevent excessive pipetting in the subsequent resuspension step, which can cause damage to the pili of the conjugation strains, which allow conjugation to occur. The supernatant was then removed and the cells were resuspended in 30 µL of fresh, sterile A1BFe+C media. This was then plated on a sterile cellulose filter placed on a LB + 300µM DAP plate. After 24 hours, the filter was removed from the plates and placed into a new Eppendorf tube with 1000µL A1BFe+C media. These were then vortexed until the culture was fully dislodged from the filter and resuspended in the media. 70µL of the resuspended cells was then plated on a A1BFe+C with 50 µg/mL kanamycin plate and incubated at 30°C for 48 hours. Once colonies started to form, a second round of selection was performed on new A1BFe+C with 50 µg/mL kanamycin plates. Colonies that grew after two rounds of selection were grown in liquid medium and frozen in 25% glycerol at -80°C.

Fluorescence measurements

To quantify CRISPRi repression, I used the fluorescent reporters according to the Banta et al. protocol[4]. After the test vectors and transposase were transformed into conjugation strain, WM3064, new strains containing *mRFP* or *sfGFP* either with or without sgRNA to target the fluorescent proteins

were cultured in 300μL of A1BFe+C in a 96 well plate in quadruplicate. Wild-type SNB-035 was also cultured in the same conditions. The protocol instructed to shake for 10 doubling times which was determined to be ~5 hours. The plate was covered with a breathable film and was then left shaking for 2 days at 30°C. Then, cultures were diluted (1:100 and 1:10,000) in duplicate, with one culture induced with 2mM IPTG and the other without. Measurements were read from the plate at days 2, 3 and 4 post-induction to detect OD600, *mRFP* expression through red fluorescence, and green fluorescence from the *sfGFP* protein.

These measurements were taken on an EnVision Platereader using Photometric600 (Abs. 600 nm) to detect OD600. To detect red fluorescence resulting from *mRFP* expression, Texas Red D595 (mirror) and Texas Red FP 555 (exc. filter) were used. To detect green fluorescence from the *sfGFP* protein, FITC (mirror) and FITC 485 (exc. filter). Fluorescence readings were normalized with the OD600 readings.

Knockdown strain growth and analysis

Glycerol stocks of knockdown strains were plated on A1BFe+C agar supplemented with 50μg/mL kanamycin (kan). Colonies took 5-7 days to appear on the plates, at which point individual colonies were picked and used to inoculate 50mL A1BFe+C starter cultures. After 48 hours, 250μL of starter culture was used to inoculate a new culture in 50mL A1BFe+C. OD600 was

read using a Nanodrop until an OD of 2-2.5 was reached (mid-exponential phase). This was typically achieved in 18-24 hours. When the target OD was reached, 1mM isopropyl β-d-1-thiogalactopyranoside (IPTG) was added to the culture to induce transcription of the CRISPRi machinery, which are regulated under the lac operon. Strains were then grown for 7 more days shaking at 200rpm, 30°C. They were extracted using liquid:liquid partitioning according to procedures described. Crude extracts were then resuspended in methanol, centrifuged to remove any particulate matter, and made at a 1mg/mL concentration. Erythrazole production was determined by the area under the curve and normalized to the area under the curve of the internal standard, 1μM chloramphenicol (CAM). Area under the curve was calculated using GNPS dashboard. This was done in triplicate for each strain and significance was determined by Welch's t-test between each knockdown strain and the control strain.

RT-PCR

Strains were grown to an OD of 2-2.5 before being induced with IPTG. They were then left shaking at 30°C for 7 days. After 7 days, 1mL was taken from each culture for RNA extraction using Monarch Total RNA Miniprep Kit (NEB T2010). Turbo DNA-free kit (Fisher AM1907) was then used to remove any background DNA before cDNA synthesis was performed using High-Capacity RNA-to-cDNA kit (Applied Biosystems 4388950). Oligonucleotides

were designed to amplify a 100-200bp region within each gene of interest via PCR.

GNPS analysis

Global Natural Products Social Molecular Networking (GNPS) was used to search for related erythrazole molecules and to analyze intermediates of knockdown strains. Orbitrap raw files were converted to mzML files using msConvert. For knockdown analysis, these files were grouped by strain with n=3 for each strain. The analysis was run independently for each predicted BGC other than BGC3 and BGC4 which only had a 2 strains each so were run together. These data were exported to Cytoscape where figures were created. Nodes which were shown to be derived majorly from the methanol wash (MeOH) controls were excluded from figures.

(1) Peters, J. M.; Koo, B.-M.; Patino, R.; Heussler, G. E.; Hearne, C. C.; Qu, J.; Inclan, Y. F.; Hawkins, J. S.; Lu, C. H. S.; Silvis, M. R.; Harden, M. M.; Osadnik, H.; Peters, J. E.; Engel, J. N.; Dutton, R. J.; Grossman, A. D.; Gross, C. A.; Rosenberg, O. S. Enabling Genetic Analysis of Diverse Bacteria with Mobile-CRISPRi. *Nat. Microbiol.* **2019**, *4* (2), 244–250. https://doi.org/10.1038/s41564-018-0327-z.

(2) Gilbert, L. A.; Horlbeck, M. A.; Adamson, B.; Villalta, J. E.; Chen, Y.; Whitehead, E. H.; Guimaraes, C.; Panning, B.; Ploegh, H. L.; Bassik, M. C.; Qi, L. S.; Kampmann, M.; Weissman, J. S. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **2014**, *159* (3), 647–661. https://doi.org/10.1016/j.cell.2014.09.029.

(3) Trevino, A. E.; Zhang, F. Chapter Eight - Genome Editing Using Cas9 Nickases. In *The Use of CRISPR/Cas9, ZFNs, and TALENs in Generating Site-Specific Genome Alterations*; Doudna, J. A., Sontheimer, E. J., Eds.; Methods in Enzymology; Academic Press, 2014; Vol. 546, pp 161–174. https://doi.org/10.1016/B978-0-12-801185-0.00008-8.

(4) Banta, A. B.; Ward, R. D.; Tran, J. S.; Bacon, E. E.; Peters, J. M. Programmable Gene Knockdown in Diverse Bacteria Using Mobile-CRISPRi. *Curr. Protoc. Microbiol.* **2020**, *59* (1), e130. https://doi.org/10.1002/cpmc.130.

Bibliography

Abby, S. S.; Kazemzadeh, K.; Vragniau, C.; Pelosi, L.; Pierrel, F. Advances in Bacterial Pathways for the Biosynthesis of Ubiquinone. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2020**, *1861* (11), 148259. https://doi.org/10.1016/j.bbabio.2020.148259.

Abron, J. D.; Singh, N. P.; Mishra, M. K.; Price, R. L.; Nagarkatti, M.; Nagarkatti, P. S.; Singh, U. P. An Endogenous Aryl Hydrocarbon Receptor Ligand, ITE, Induces Regulatory T Cells and Ameliorates Experimental Colitis. *American Journal of Physiology-Gastrointestinal and Liver Physiology* **2018**, *315* (2), G220–G230. https://doi.org/10.1152/ajpgi.00413.2017.

Alcántara, A. R.; Domínguez de María, P.; Littlechild, J. A.; Schürmann, M.; Sheldon, R. A.; Wohlgemuth, R. Biocatalysis as Key to Sustainable Industrial Chemistry. *ChemSusChem* **2022**, *15* (9), e202102709. https://doi.org/10.1002/cssc.202102709.

Allocati, N.; Federici, L.; Masulli, M.; Di Ilio, C. Glutathione Transferases in Bacteria. *The FEBS Journal* **2009**, *276* (1), 58–75. https://doi.org/10.1111/j.1742-4658.2008.06743.x.

Álvarez-Álvarez, R.; Botas, A.; Albillos, S. M.; Rumbero, A.; Martín, J. F.; Liras, P. Molecular Genetics of Naringeninbiosynthesis, a Typical Plant Secondary Metabolite Produced by Streptomyces Clavuligerus. *Microbial Cell Factories* **2015**, *14* (1), 178. https://doi.org/10.1186/s12934-015-0373-7.

Atanasov, A. G. et al. the International Natural Product Sciences Taskforce. Natural Products in Drug Discovery: Advances and Opportunities. *Nature Reviews Drug Discovery* **2021**, *20* (3), 200–216. https://doi.org/10.1038/s41573-020-00114-z.

Aussel, L.; Pierrel, F.; Loiseau, L.; Lombard, M.; Fontecave, M.; Barras, F. Biosynthesis and Physiology of Coenzyme Q in Bacteria. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2014**, *1837* (7), 1004–1011. https://doi.org/10.1016/j.bbabio.2014.01.015.

Bahrami, K.; Khodaei, M. M.; Naali, F. Mild and Highly Efficient Method for the Synthesis of 2-Arylbenzimidazoles and 2-Arylbenzothiazoles. *J. Org. Chem.* **2008**, *73* (17), 6835–6837. https://doi.org/10.1021/jo8010232.

Banta, A. B.; Enright, A. L.; Siletti, C.; Peters, J. M. A High-Efficacy CRISPR Interference System for Gene Function Discovery in Zymomonas Mobilis. *Applied and Environmental Microbiology* **2020**, *86* (23), e01621-20. https://doi.org/10.1128/AEM.01621-20.

Banta, A. B.; Ward, R. D.; Tran, J. S.; Bacon, E. E.; Peters, J. M. Programmable Gene Knockdown in Diverse Bacteria Using Mobile-CRISPRi. *Current Protocols in Microbiology* **2020**, *59* (1), e130. https://doi.org/10.1002/cpmc.130.

Barra, L.; Awakawa, T.; Abe, I. Noncanonical Functions of Enzyme Cofactors as Building Blocks in Natural Product Biosynthesis. *JACS Au* **2022**, *2* (9), 1950–1963. https://doi.org/10.1021/jacsau.2c00391.

Barra, L.; Awakawa, T.; Shirai, K.; Hu, Z.; Bashiri, G.; Abe, I. β-NAD as a Building Block in Natural Product Biosynthesis. *Nature* **2021**, *600* (7890), 754–758. https://doi.org/10.1038/s41586-021-04214-7.

Bell, E. L.; Finnigan, W.; France, S. P.; Green, A. P.; Hayes, M. A.; Hepworth, L. J.; Lovelock, S. L.; Niikura, H.; Osuna, S.; Romero, E.; Ryan, K. S.; Turner, N. J.; Flitsch, S. L. Biocatalysis. *Nature Reviews Methods Primers* **2021**, *1* (1), 46. https://doi.org/10.1038/s43586-021-00044-z.

Bertoldo, C.; Antranikian, G. Starch-Hydrolyzing Enzymes from Thermophilic Archaea and Bacteria. *Current Opinion in Chemical Biology* **2002**, *6* (2), 151–160. https://doi.org/10.1016/S1367-5931(02)00311-3.

Biermann, F.; Helfrich, E. J. N. Hidden Treasures: Microbial Natural Product Biosynthesis off the Beaten Path. *mSystems* **2021**, *6* (4), 10.1128/msystems.00846-21. https://doi.org/10.1128/msystems.00846-21.

Bode, H. B.; Müller, R. The Impact of Bacterial Genomics on Natural Product Research. *Angewandte Chemie International Edition* **2005**, *44* (42), 6828–6846. https://doi.org/10.1002/anie.200501080.

Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K. Engineering the Third Wave of Biocatalysis. *Nature* **2012**, *485* (7397), 185–194. https://doi.org/10.1038/nature11117.

Bruggink, A.; Roos, E. C.; de Vroom, E. Penicillin Acylase in the Industrial Production of β-Lactam Antibiotics. *Org. Process Res. Dev.* **1998**, *2* (2), 128–133. https://doi.org/10.1021/op9700643.

Brustad, E. M. New Recipes for Biocatalysis. *Nature Chemical Biology* **2014**, *10* (3), 170–171. https://doi.org/10.1038/nchembio.1457.

Burgardt, A.; Moustafa, A.; Persicke, M.; Wendische, V. F. Coenzyme Q10 Biosynthesis Established in the Non-Ubiquinone Containing Corynebacterium Glutamicum by Metabolic Engineering. *Frontiers in Bioengineering and Biotechnology* **2021**, *9*.

Butler, M. S. The Role of Natural Product Chemistry in Drug Discovery. *J. Nat. Prod.* **2004**, *67* (12), 2141–2153. https://doi.org/10.1021/np040106y.

Cane, D. E.; Walsh, C. T.; Khosla, C. Harnessing the Biosynthetic Code: Combinations, Permutations, and Mutations. *Science* **1998**, *282* (5386), 63–68. https://doi.org/10.1126/science.282.5386.63.

Carroll, A. R.; Copp, B. R.; Davis, R. A.; Keyzers, R. A.; Prinsep, M. R. Marine Natural Products. *Natural product reports* **2022**, *39* (6), 1122–1171.

Chakrabarty, S.; Romero, E. O.; Pyser, J. B.; Yazarians, J. A.; Narayan, A. R. H. Chemoenzymatic Total Synthesis of Natural Products. *Acc. Chem. Res.* **2021**, *54* (6), 1374–1384. https://doi.org/10.1021/acs.accounts.0c00810.

Chill, L.; Rudi, A.; Benayahu, Y.; Kashman, Y. Violatinctamine, a New Heterocyclic Compound from the Marine Tunicate Cystodytes Cf. Violatinctus. *Tetrahedron Letters* **2004**, *45* (42), 7925–7928. https://doi.org/10.1016/j.tetlet.2004.08.137.

Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H. Olefin Cyclopropanation via Carbene Transfer Catalyzed by Engineered Cytochrome P450 Enzymes. *Science* **2013**, *339* (6117), 307–310. https://doi.org/10.1126/science.1231434.

Cricchio, R.; Antonini, P.; Sartori, G. THIAZORIFAMYCINS. III BIOSYNTHESIS OF RIFAMYCINS P, Q AND VERDE, NOVEL METABOLITES FROM A MUTANT OF NOCARDIA MEDITERRANEA. *The Journal of Antibiotics* **1980**, *33* (8), 842–846.

Daugherty, M.; Polanuyer, B.; Farrell, M.; Scholle, M.; Lykidis, A.; Crécy-Lagard, V. de; Osterman, A. Complete Reconstitution of the Human Coenzyme A Biosynthetic Pathway via Comparative Genomics*. *Journal of*

Biological Chemistry **2002**, *277* (24), 21431–21439. https://doi.org/10.1074/jbc.M201708200.

Denner, E. B.; Vybiral, D.; Koblízek, M.; Kämpfer, P.; Busse, H.-J.; Velimirov, B. Erythrobacter Citreus Sp. Nov., a Yellow-Pigmented Bacterium That Lacks Bacteriochlorophyll a, Isolated from the Western Mediterranean Sea. *International journal of systematic and evolutionary microbiology* **2002**, *52* (5), 1655–1661.

Desriac, F.; Jégou, C.; Balnois, E.; Brillet, B.; Chevalier, P. L.; Fleury, Y. Antimicrobial Peptides from Marine Proteobacteria. *Marine Drugs* **2013**, *11* (10), 3632–3660. https://doi.org/10.3390/md11103632.

Di Donato, P.; Napolitano, A.; Prota, G. Metal Ions as Potential Regulatory Factors in the Biosynthesis of Red Hair Pigments: A New Benzothiazole Intermediate in the Iron or Copper Assisted Oxidation of 5-S-Cysteinyldopa. *Biochimica et Biophysica Acta (BBA)-General Subjects* **2002**, *1571* (2), 157–166.

Donadio, S.; Staver, M. J.; McAlpine, J. B.; Swanson, S. J.; Katz, L. Modular Organization of Genes Required for Complex Polyketide Biosynthesis. *Science* **1991**, *252* (5006), 675–679. https://doi.org/10.1126/science.2024119.

Dufossé, L.; Echanove, M. C. de. The Last Step in the Biosynthesis of Aryl Carotenoids in the Cheese Ripening Bacteria Brevibacterium Linens ATCC 9175 (Brevibacterium Aurantiacum Sp. Nov.) Involves a Cytochrome P450-Dependent Monooxygenase. *Food Research International* **2005**, *38* (8), 967–973. https://doi.org/10.1016/j.foodres.2005.02.017.

Eddy, S. R. Hidden Markov Models. *Curr Opin Struct Biol* **1996**, *6* (3), 361–365. https://doi.org/10.1016/s0959-440x(96)80056-x.

Eddy, S. R. Accelerated Profile HMM Searches. *PLOS Computational Biology* **2011**, *7* (10), 1–16. https://doi.org/10.1371/journal.pcbi.1002195.

Eddy, S. HMMER 3.1b2, 2015.

Elias, M.; Tanaka, M.; Sakai, M.; Toyama, H.; Matsushita, K.; Adachi, O.; Yamada, M. C-Terminal Periplasmic Domain of Escherichia Coli Quinoprotein Glucose Dehydrogenase Transfers Electrons to Ubiquinone*.

Journal of Biological Chemistry **2001**, *276* (51), 48356–48361. https://doi.org/10.1074/jbc.M107355200.

Esser, C.; Rannug, A.; Stockinger, B. The Aryl Hydrocarbon Receptor in Immunity. *Trends in immunology* **2009**, *30* (9), 447–454.

Fan, X.; He, Y.; Wang, Y.; Xue, Z.; Zhang, X.; Wang, J. A Novel and Practical Synthesis of 2-Benzoylbenzothiazoles and 2-Benzylbenzothiazoles. *Tetrahedron Letters* **2011**, *52* (8), 899–902.

Feng, X.-M.; Mo, Y.-X.; Han, L.; Nogi, Y.; Zhu, Y.-H.; Lv, J. Qipengyuania Sediminis Gen. Nov., Sp. Nov., a Member of the Family Erythrobacteraceae Isolated from Subterrestrial Sediment. *International Journal of Systematic and Evolutionary Microbiology*, 2015, *65*, 3658–3665. https://doi.org/10.1099/ijsem.0.000472.

Fiebig, A.; Varesio, L. M.; Alejandro Navarreto, X.; Crosson, S. Regulation of the Erythrobacter Litoralis DSM 8509 General Stress Response by Visible Light. *Mol Microbiol* **2019**, *112* (2), 442–460. https://doi.org/10.1111/mmi.14310.

Fischbach, M. A.; Walsh, C. T. Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. *Chem. Rev.* **2006**, *106* (8), 3468–3496. https://doi.org/10.1021/cr0503097.

Frank, A.; Groll, M. The Methylerythritol Phosphate Pathway to Isoprenoids. *Chemical Reviews* **2017**, *117* (8), 5675–5703. https://doi.org/10.1021/acs.chemrev.6b00537.

Fukumoto, S. Identification of a Probiotic Bacteria-Derived Activator of the Aryl Hydrocarbon Receptor That Inhibits Colitis. *Immunology Cell Biology* **2014**, *92* (5), 460–465.

Gao, X.; Liu, J.; Zuo, X.; Feng, X.; Gao, Y. Recent Advances in Synthesis of Benzothiazole Compounds Related to Green Chemistry. *Molecules* **2020**, *25* (7). https://doi.org/10.3390/molecules25071675.

Gilbert, L. A.; Horlbeck, M. A.; Adamson, B.; Villalta, J. E.; Chen, Y.; Whitehead, E. H.; Guimaraes, C.; Panning, B.; Ploegh, H. L.; Bassik, M. C.; Qi, L. S.; Kampmann, M.; Weissman, J. S. Genome-Scale CRISPR-Mediated

Control of Gene Repression and Activation. *Cell* **2014**, *159* (3), 647–661. https://doi.org/10.1016/j.cell.2014.09.029.

Goettel, J. A. AHR Activation Is Protective against Colitis Driven by T Cells in Humanized Mice. *Cell Reports* **2016**, *17* (5), 1318–1329.

Guengerich, F. P. Cytochrome P450 Enzymes in the Generation of Commercial Products. *Nature Reviews Drug Discovery* **2002**, *1* (5), 359–366. https://doi.org/10.1038/nrd792.

Guo, H. Natural Products and Morphogenic Activity of Gamma-Proteobacteria Associated with the Marine Hydroid Polyp Hydractinia Echinata. *Bioorganic Medicinal Chemistry* **2017**, *25* (22), 6088–6097.

Guo, J.; Hao, Y.; Ji, X.; Wang, Z.; Liu, Y.; Ma, D.; Li, Y.; Pang, H.; Ni, J.; Wang, Q. Optimization, Structure–Activity Relationship, and Mode of Action of Nortopsentin Analogues Containing Thiazole and Oxazole Moieties. *Journal of Agricultural and Food Chemistry* **2019**, *67*, 10018–10031.

Hannigan, G. D.; Prihoda, D.; Britton, D. A. A Deep Learning Genome-Mining Strategy for Biosynthetic Gene Cluster Prediction. *Nucleic Acids Research* **2019**, *47* (18), e110.

Harashima, K.; Nakada, H. Carotenoids and Ubiquinone in Aerobically Grown Cells of an Aerobic Photosynthetic Bacterium, Erythrobacter Species OCh114. *Agricultural and Biological Chemistry* **1983**, *47* (5), 1057–1063.

Harrison, P. J.; Bugg, T. D. H. Enzymology of the Carotenoid Cleavage Dioxygenases: Reaction Mechanisms, Inhibition and Biochemical Roles. *Archives of Biochemistry and Biophysics* **2014**, *544*, 105–111. https://doi.org/10.1016/j.abb.2013.10.005.

He, P.; Moran, G. R. Structural and Mechanistic Comparisons of the Metal-Binding Members of the Vicinal Oxygen Chelate (VOC) Superfamily. *Journal of Inorganic Biochemistry* **2011**, *105* (10), 1259–1272. https://doi.org/10.1016/j.jinorgbio.2011.06.006.

Helfrich, E. J. N.; Ueoka, R.; Dolev, A.; Rust, M.; Meoded, R. A.; Bhushan, A.; Califano, G.; Costa, R.; Gugger, M.; Steinbeck, C.; Moreno, P.; Piel, J. Automated Structure Prediction of Trans-Acyltransferase Polyketide Synthase Products. *Nature Chemical Biology* **2019**, *15* (8), 813–821. https://doi.org/10.1038/s41589-019-0313-7.

Helfrich, E. J.; Piel, J. Biosynthesis of Polyketides by Trans-AT Polyketide Synthases. *Natural product reports* **2016**, *33* (2), 231–316.

Henry, E. C. A Potential Endogenous Ligand for the Aryl Hydrocarbon Receptor Has Potent Agonist Activity in Vitro and in Vivo. *Arch Biochem Biophys* **2006**, *450* (1), 67–77.

Hoffmann, J.; Bóna-Lovász, J.; Beuttler, H.; Altenbuchner, J. In Vivo and in Vitro Studies on the Carotenoid Cleavage Oxygenases from Sphingopyxis Alaskensis RB2256 and Plesiocystis Pacifica SIR-1 Revealed Their Substrate Specificities and Non-Retinal-Forming Cleavage Activities. *The FEBS Journal* **2012**, *279* (20), 3911–3924. https://doi.org/10.1111/j.1742-4658.2012.08751.

Hosokawa, N.; Naganawa, H.; Hamada, M.; Iinuma, H.; Takeuchi, T.; Tsuchiya, K. S.; Hori, M. New Triene-Ansamycins, Thiazinotrienomycins F and G and a Diene-Ansamycin, Benzoxazomycin. *J Antibiot (Tokyo)* **2000**, *53* (9), 886–894. https://doi.org/10.7164/antibiotics.53.886.

Hu, Y.; Legako, A. G.; Espindola, A. P. D.; MacMillan, J. B. Erythrolic Acids A–E, Meroterpenoids from a Marine-Derived Erythrobacter Sp. *The Journal of Organic Chemistry* **2012**, *77* (7), 3401–3407.

Hu, Y.; MacMillan, J. B. Erythrazoles A–B, Cytotoxic Benzothiazoles from a Marine-Derived *Erythrobacter* Sp. *Org. Lett.* **2011**, *13* (24), 6580–6583. https://doi.org/10.1021/ol202944g.

Huffman, M. A.; Fryszkowska, A.; Alvizo, O.; Borra-Garske, M.; Campos, K. R.; Canada, K. A.; Devine, P. N.; Duan, D.; Forstater, J. H.; Grosser, S. T.; Halsey, H. M.; Hughes, G. J.; Jo, J.; Joyce, L. A.; Kolev, J. N.; Liang, J.; Maloney, K. M.; Mann, B. F.; Marshall, N. M.; McLaughlin, M.; Moore, J. C.; Murphy, G. S.; Nawrat, C. C.; Nazor, J.; Novick, S.; Patel, N. R.; Rodriguez-Granillo, A.; Robaire, S. A.; Sherer, E. C.; Truppo, M. D.; Whittaker, A. M.; Verma, D.; Xiao, L.; Xu, Y.; Yang, H. Design of an in Vitro Biocatalytic Cascade for the Manufacture of Islatravir. *Science* **2019**, *366* (6470), 1255–1259. https://doi.org/10.1126/science.aay8484.

Ikeda, H.; Ishikawa, J.; Hanamoto, A.; Shinose, M.; Kikuchi, H.; Shiba, T.; Sakaki, Y.; Hattori, M.; Ōmura, S. Complete Genome Sequence and Comparative Analysis of the Industrial Microorganism Streptomyces

Avermitilis. *Nature Biotechnology* **2003**, *21* (5), 526–531. https://doi.org/10.1038/nbt820.

Irfan, A.; Batool, F.; Zahra Naqvi, S. A.; Islam, A.; Osman, S. M.; Nocentini, A.; Alissa, S. A.; Supuran, C. T. Benzothiazole Derivatives as Anticancer Agents. *Journal of Enzyme Inhibition and Medicinal Chemistry* **2020**, *35* (1), 265–279. https://doi.org/10.1080/14756366.2019.1698036.

Isowa, Y.; Ohmori, M.; Ichikawa, T.; Mori, K.; Nonaka, Y.; Kihara, K.; Oyama, K.; Satoh, H.; Nishimura, S. The Thermolysin-Catalyzed Condensation Reactions of n-Substituted Aspartic and Glutamic Acids with Phenylalanine Alkyl Esters. *Tetrahedron Letters* **1979**, *20* (28), 2611–2612. https://doi.org/10.1016/S0040-4039(01)86363-2.

Jansen, R. Indothiazinone, an Indolyl Thiazolyl Ketone from a Novel Myxobacterium Belonging to the Sorangiineae. *Journal of Natural Products* **2014**, *77* (4), 1054–1060.

Jenke-Kodama, H.; Dittmann, E. Bioinformatic Perspectives on NRPS/PKS Megasynthases: Advances and Challenges. *Natural product reports* **2009**, *26* (7), 874–883.

Jhulki, I.; Chanani, P. K.; Abdelwahed, S. H.; Begley, T. P. A Remarkable Oxidative Cascade That Replaces the Riboflavin C8 Methyl with an Amino Group during Roseoflavin Biosynthesis. *J. Am. Chem. Soc.* **2016**, *138* (27), 8324–8327. https://doi.org/10.1021/jacs.6b02469.

Jin, X.; Zhang, Y.; Zhang, R.; Nguyen, K.-U.; Lindsey, J. S.; Miller, E. S. Identification of Putative Biosynthetic Gene Clusters for Tolyporphins in Multiple Filamentous Cyanobacteria. *Life* **2021**, *11* (8). https://doi.org/10.3390/life11080758.

Joint, I.; Mühling, M.; Querellou, J. Culturing Marine Bacteria – an Essential Prerequisite for Biodiscovery. *Microbial Biotechnology* **2010**, *3* (5), 564–575. https://doi.org/10.1111/j.1751-7915.2010.00188.x.

Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28* (1), 27–30. https://doi.org/10.1093/nar/28.1.27.

Keri, Rangappa S.; Patil, M.; Patil, S.; Budagumpi, S. A Comprehensive Review in Current Developments of Benzothiazole- Based

Molecules in Medicinal Chemistry. *European Journal of Medicinal Chemistry* **2015**, *89*, 207–251.

Khanjin, N. A.; Snyder, J. P.; Menger, F. M. Mechanism of Chorismate Mutase: Contribution of Conformational Restriction to Catalysis in the Claisen Rearrangement. *J. Am. Chem. Soc.* **1999**, *121* (50), 11831–11846. https://doi.org/10.1021/ja992453d.

Kim, C. sub; Li, J.-H.; Barco, B.; Piizi, G.; Clay, N. K.; Crawford, J. M. Cellular Stress Upregulates Indole Signaling Metabolites in Escherichia Coli. *Cell Chemical Biology* **2020**, *27* (6), 698–707.

Ko, S.-R.; Le, V. V.; Srivastava, A.; Kang, M.; Oh, H.-M.; Ahn, C.-Y. Algicidal Activity of a Novel Bacterium, Qipengyuania Sp. 3-20A1M, against Harmful Margalefidinium Polykrikoides: Effects of Its Active Compound. *Marine Pollution Bulletin* **2023**, *186*, 114397. https://doi.org/10.1016/j.marpolbul.2022.114397.

Koblížek, M.; Béjà, O.; Bidigare, R. R.; Christensen, S.; Benitez-Nelson, B.; Vetriani, C.; Kolber, M. K.; Falkowski, P. G.; Kolber, Z. S. Isolation and Characterization of Erythrobacter Sp. Strains from the Upper Ocean. *Archives of microbiology* **2003**, *180*, 327–338.

Komatsu, M.; Tsuda, M.; Omura, S.; Ikeda, H. Identification and Functional Analysis of Genes Controlling Biosynthesis of 2-Methylisoborneol. *PNAS 105* (21), 7422–7427.

Krogh, A.; Brown, M.; Mian, I. S.; Sjölander, K.; Haussler, D. Hidden Markov Models in Computational Biology. Applications to Protein Modeling. *J Mol Biol* **1994**, *235* (5), 1501–1531. https://doi.org/10.1006/jmbi.1994.1104.

Lamas, B. Aryl Hydrocarbon Receptor and Intestinal Immunity. *Mucosal Immunology* **2018**, *11* (4), 1024–1038.

Lee, Y. S.; Lee, D.-H.; Kahng, H.-Y.; Kim, E. M.; Jung, J. S. Erythrobacter Gangjinensis Sp. Nov., a Marine Bacterium Isolated from Seawater. *International Journal of Systematic and Evolutionary Microbiology* **2009**, *60* (6), 1413–1417.

Lenz, O.; Friedrich, B. A Novel Multicomponent Regulatory System Mediates H2 Sensing in Alcaligenes Eutrophus. *Proc Natl Acad Sci USA* **1998**, *95* (21), 12474–12479.

Levy, M. Microbiome, Metabolites and Host Immunity. *Current Opinions in Microbiology* **2017**, *35*, 8–15.

Li, B.; Su, J.; Tao, J. Enzyme and Process Development for Production of Nicotinamide. *Org. Process Res. Dev.* **2011**, *15* (1), 291–293. https://doi.org/10.1021/op100309g.

Li, X.; Jun, Y.; Erickstad, M. J.; Brown, S. D.; Parks, A.; Court, D. L.; Jun, S. tCRISPRi: Tunable and Reversible, One-Step Control of Gene Expression. *Scientific Reports* **2016**, *6* (1), 39076. https://doi.org/10.1038/srep39076.

Li, Y. Exogenous Stimuli Maintain Intraepithelial Lymphocytes via Aryl Hydrocarbon Receptor Activation.". *Cell* **2011**, *147* (3), 629–640.

Lin, G.-H.; Hsieh, M.-C.; Shu, H.-Y. Role of Iron-Containing Alcohol Dehydrogenases in Acinetobacter Baumannii ATCC 19606 Stress Resistance and Virulence. *Int J Mol Sci* **2021**, *22* (18). https://doi.org/10.3390/ijms22189921.

Lin, Y.; Sun, X.; Yuan, Q.; Yan, Y. Engineering Bacterial Phenylalanine 4-Hydroxylase for Microbial Synthesis of Human Neurotransmitter Precursor 5-Hydroxytryptophan. *ACS Synth. Biol.* **2014**, *3* (7), 497–505. https://doi.org/10.1021/sb5002505.

Link, A. J. Methods for Generating Precise Deletions and Insertions in the Genome of Wild-Type Escherichia Coli: Application to Open Reading Frame Characterization. *Journal of Bacteriology* **1997**, *179* (20), 6228–6237.

Liu, Y.; Pei, T.; Du, J.; Yao, Q.; Deng, M.-R.; Zhu, H. Comparative Genomics Reveals Genetic Diversity and Metabolic Potentials of the Genus Qipengyuania and Suggests Fifteen Novel Species. *Microbiol Spectr* **2022**, *10* (3), e0126421. https://doi.org/10.1128/spectrum.01264-21.

Lu, L.; Cai, L.; Jiao, N.; Zhang, R. Isolation and Characterization of the First Phage Infecting Ecologically Important Marine Bacteria Erythrobacter. *Virology Journal* **2017**, *14* (1), 1–15.

Lu, W.; Shi, Y.; He, S.; Fei, Y.; Yu, K.; Yu, H. Enhanced Production of CoQ10 by Constitutive Overexpression of 3-Demethyl Ubiquinone-9 3-Methyltransferase under Tac Promoter in Rhodobacter Sphaeroides.

Biochemical Engineering Journal **2013**, *72*, 42–47. https://doi.org/10.1016/j.bej.2012.12.019.

Lv, J.-M.; Gao, Y.-H.; Zhao, H.; Awakawa, T.; Liu, L.; Chen, G.-D.; Yao, X.-S.; Hu, D.; Abe, I.; Gao, H. Biosynthesis of Biscognienyne B Involving a Cytochrome P450-Dependent Alkynylation. *Angewandte Chemie International Edition* **2020**, *59* (32), 13531–13536. https://doi.org/10.1002/anie.202004364.

MacMillan, J. B.; Hu, Y. Erythrazoles A-B, Cytotoxic Benzothiazoles from a Marine-Derived Erythrobacter Sp. *Organic Letters* **2011**, *13* (24), 6580–6583.

Madeira, F.; Pearce, M.; Tivey, A. R. N.; Basutkar, P.; Lee, J.; Edbali, O.; Madhusoodanan, N.; Kolesnikov, A.; Lopez, R. Search and Sequence Analysis Tools Services from EMBL-EBI in 2022. *Nucleic acids research* **2022**, *50* (W1), W276—W279. https://doi.org/10.1093/nar/gkac240.

Marahiel, M. A.; Stachelhaus, T.; Mootz, H. D. Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem. Rev.* **1997**, *97* (7), 2651–2674. https://doi.org/10.1021/cr960029e.

Matsumura, H.; Takeyama, H.; Kusakabe, E.; Burgess, J. G.; Matsunaga, T. Cloning, Sequencing and Expressing the Carotenoid Biosynthesis Genes, Lycopene Cyclase and Phytoene Desaturase, from the Aerobic Photosynthetic Bacterium Erythrobacter Longus Sp. Strain Och101 in Escherichia Coli. *Gene* **1997**, *189* (2), 169–174.

Medema, M.; Blin, K.; Cimermancic, P.; Zakrzewski, P.; Fischbach, M.; Weber, T.; Takano, E.; Breitling, R. antiSMASH: Rapid Identification, Annotation and Analysis of Secondary Metabolite Biosynthesis Gene Clusters in Bacterial and Fungal Genome Sequences. *Nucleic Acids Research* **2011**, *39*, W339–W346.

Meganathan, R. Ubiquinone Biosynthesis in Microorganisms. *FEMS Microbiology Letters* **2001**, *203* (2), 131–139. https://doi.org/10.1111/j.1574-6968.2001.tb10831.x.

Mihara, H.; Esaki, N. Bacterial Cysteine Desulfurases: Their Function and Mechanisms. *Applied Microbiology and Biotechnology* **2002**, *60* (1), 12–23. https://doi.org/10.1007/s00253-002-1107-4.

Miller, J. H. M9 Minimal Medium (Standard). *Cold Spring harbor Protocols* **2010**.

Moghaddam, J. A.; Jautzus, T.; Alanjary, M.; Beemelmanns, C. Recent Highlights of Biosynthetic Studies on Marine Natural Products. *Organic & Biomolecular Chemistry* **2021**, *19* (1), 123–140.

Mostafa, Y. S.; Alrumman, S. A.; Otaif, K. A.; Alamri, S. A.; Mostafa, M. S.; Sahlabji, T. Production and Characterization of Bioplastic by Polyhydroxybutyrate Accumulating Erythrobacter Aquimaris Isolated from Mangrove Rhizosphere. *Molecules* **2020**, *25* (1), 179.

Muku, G. E.; Murray, I. A.; Espín, J. C.; Perdew, G. H. Urolithin A Is a Dietary Microbiota-Derived Human Aryl Hydrocarbon Receptor Antagonist. *Metabolites* **2018**, *8* (4). https://doi.org/10.3390/metabo8040086.

Mulero-Navarro, S.; Fernandez-Salguero, M. New Trends in Aryl Hydrocarbon Receptor Biology. *Front Cell Dev Biol* **2016**, *4*, 45.

Nakamura, H.; Schultz, E. E.; Balskus, E. P. A New Strategy for Aromatic Ring Alkylation in Cylindrocyclophane Biosynthesis. *Nature Chemical Biology* **2017**, *13* (8), 916–921. https://doi.org/10.1038/nchembio.2421.

Nguyen, T.; Ishida, K.; Jenke-Kodama, H.; Dittmann, E.; Gurgui, C.; Hochmuth, T.; Taudien, S.; Platzer, M.; Hertweck, C.; Piel, J. Exploiting the Mosaic Structure of Trans-Acyltransferase Polyketide Synthases for Natural Product Discovery and Pathway Dissection. *Nature Biotechnology* **2008**, *26* (2), 225–233. https://doi.org/10.1038/nbt1379.

Niero, H.; da Silva, M. A. C.; de Felicio, R.; Trivella, D. B. B.; Lima, A. O. de S. Carotenoids Produced by the Deep-Sea Bacterium Erythrobacter Citreus LAMA 915: Detection and Proposal of Their Biosynthetic Pathway. *Folia Microbiologica* **2021**, *66*, 441–456.

Nunes, J. E. S.; Duque, M. A.; de Freitas, T. F.; Galina, L.; Timmers, L. F. S. M.; Bizarro, C. V.; Machado, P.; Basso, L. A.; Ducati, R. G. Mycobacterium Tuberculosis Shikimate Pathway Enzymes as Targets for the Rational Design of Anti-Tuberculosis Drugs. *Molecules* **2020**, *25* (6). https://doi.org/10.3390/molecules25061259.

Oba, Y.; Yoshida, N.; Kanie, S.; Ojika, M.; Inouye, S. Biosynthesis of Firefly Luciferin in Adult Lantern: Decarboxylation of L-Cysteine Is a Key Step for Benzothiazole Ring Formation in Firefly Luciferin Synthesis. *PLoS One* **2013**, *8* (12), e84023. https://doi.org/10.1371/journal.pone.0084023.

Oldfield, E.; Lin, F.-Y. Terpene Biosynthesis: Modularity Rules. *Angew Chem Int Ed Engl* **2012**, *51* (5), 1124–1137. https://doi.org/10.1002/anie.201103110.

Olson, R. E.; Rudney, H. Biosynthesis of Ubiquinone; Aurbach, G. D., McCormick, D. B., Eds.; Vitamins & Hormones; Academic Press, 1983; Vol. 40, pp 1–43. https://doi.org/10.1016/S0083-6729(08)60431-8.

Orsi, E.; Beekwilder, J.; Peek, S.; Eggink, G.; Kengen, S. W. M.; Weusthuis, R. A. Metabolic Flux Ratio Analysis by Parallel 13C Labeling of Isoprenoid Biosynthesis in Rhodobacter Sphaeroides. *Metabolic Engineering* **2020**, *57*, 228–238. https://doi.org/10.1016/j.ymben.2019.12.004.

Pan, G.; Xu, Z.; Guo, Z.; null; Ma, M.; Yang, D.; Zhou, H.; Gansemans, Y.; Zhu, X.; Huang, Y.; Zhao, L.-X.; Jiang, Y.; Cheng, J.; Nieuwerburgh, F. V.; Suh, J.-W.; Duan, Y.; Shen, B. Discovery of the Leinamycin Family of Natural Products by Mining Actinobacterial Genomes. *Proceedings of the National Academy of Sciences* **2017**, *114* (52), E11131–E11140. https://doi.org/10.1073/pnas.1716245115.

Pellegrin, V. Molecular Formulas of Organic Compounds: The Nitrogen Rule and Degree of Unsaturation. *Journal of Chemical Education* **1983**, *60* (8), 626.

Peters, J. M.; Koo, B.-M.; Patino, R.; Heussler, G. E.; Hearne, C. C.; Qu, J.; Inclan, Y. F.; Hawkins, J. S.; Lu, C. H. S.; Silvis, M. R.; Harden, M. M.; Osadnik, H.; Peters, J. E.; Engel, J. N.; Dutton, R. J.; Grossman, A. D.; Gross, C. A.; Rosenberg, O. S. Enabling Genetic Analysis of Diverse Bacteria with Mobile-CRISPRi. *Nature Microbiology* **2019**, *4* (2), 244–250. https://doi.org/10.1038/s41564-018-0327-z.

Peters, J. E. Targeted Transposition with Tn7 Elements: Safe Sites, Mobile Plasmids, CRISPR/Cas and Beyond. *Mol Microbiol* **2019**, *112* (6), 1635–1644. https://doi.org/10.1111/mmi.14383.

Pyser, J. B.; Chakrabarty, S.; Romero, E. O.; Narayan, A. R. H. State-of-the-Art Biocatalysis. *ACS Cent. Sci.* **2021**, *7* (7), 1105–1116. https://doi.org/10.1021/acscentsci.1c00273.

Qin, Z.; Devine, R.; Booth, T. J.; Farrar, E. H.; Grayson, M. N.; Hutchings, M. I.; Wilkinson, B. Formicamycin Biosynthesis Involves a Unique Reductive Ring Contraction. *Chemical Science* **2020**, *11* (31), 8125–8131.

Rajniak, J. A New Cyanogenic Metabolite in Arabidopsis Required for Inducible Pathogen Defence. *Nature* **2015**, *525* (7569), 376–379.

Rassbach, J.; Hilsberg, N.; Haensch, V. G.; Dörner, S.; Gressler, J.; Sonnabend, R.; Semm, C.; Voigt, K.; Hertweck, C.; Gressler, M. Non-Canonical Two-Step Biosynthesis of Anti-Oomycete Indole Alkaloids in Kickxellales. *Fungal Biology and Biotechnology* **2023**, *10* (1), 19. https://doi.org/10.1186/s40694-023-00166-x.

Richter, T. K. S.; Hughes, C. C.; Moore, B. S. Sioxanthin, a Novel Glycosylated Carotenoid, Reveals an Unusual Subclustered Biosynthetic Pathway. *Environmental Microbiology* **2015**, *17* (6), 2158–2171. https://doi.org/10.1111/1462-2920.12669.

Rohmer, M.; Knani, M.; Simonin, P.; Sutter, B.; Sahm, H. Isoprenoid Biosynthesis in Bacteria: A Novel Pathway for the Early Steps Leading to Isopentenyl Diphosphate. *Biochemical Journal* **1993**, *295* (2), 517–524. https://doi.org/10.1042/bj2950517.

Rohmer, M. The Discovery of a Mevalonate-Independent Pathway for Isoprenoid Biosynthesis in Bacteria, Algae and Higher Plants. *Nat Prod Rep* **1999**, *16* (5), 565–574. https://doi.org/10.1039/a709175c.

Rohmer, M. Mevalonate-Independent Methylerythritol Phosphate Pathway for Isoprenoid Biosynthesis. Elucidation and Distribution. *Pure and Applied Chemistry* **2003**, *75* (2–3), 375–388. https://doi.org/doi:10.1351/pac200375020375.

Rohmer, M. Diversity in Isoprene Unit Biosynthesis: The Methylerythritol Phosphate Pathway in Bacteria and Plastids. *Pure and Applied Chemistry* **2007**, *79* (4), 739–751. https://doi.org/doi:10.1351/pac200779040739.

Rohmer, M.; Sutter, B.; Sahm, H. Bacterial Sterol Surrogates. Biosynthesis of the Side-Chain of Bacteriohopanetetrol and of a Carbocyclic Pseudopentose from 13C-Labelled Glucose in Zymomonas Mobilis. *J. Chem. Soc., Chem. Commun.* **1989**, No. 19, 1471–1472. https://doi.org/10.1039/C39890001471.

Scesa, P. D.; Lin, Z.; Schmidt, E. W. Ancient Defensive Terpene Biosynthetic Gene Clusters in the Soft Corals. *Nature Chemical Biology* **2022**, *18* (6), 659–663. https://doi.org/10.1038/s41589-022-01027-1.

Scherlach, K.; Hertweck, C. Mining and Unearthing Hidden Biosynthetic Potential. *Nature Communications* **2021**, *12* (1), 3864. https://doi.org/10.1038/s41467-021-24133-5.

Schultz, M. Clinical Use of E. Coli Nissle 1917 in Inflammatory Bowel Disease. *Inflammatory Bowel Diseases* **2008**, *14* (7), 1012–1018. https://doi.org/10.1002/ibd.20377.

Scott, T. A.; Piel, J. The Hidden Enzymology of Bacterial Natural Product Biosynthesis. *Nature Reviews Chemistry* **2019**, *3* (7), 404–425. https://doi.org/10.1038/s41570-019-0107-1.

Seijas, J. A.; Vázquez-Tato, M. P.; Carballido-Reboredo, M. R.; Crecente-Campo, J.; Romar-Lopez, L. Lawesson's Reagent and Microwaves: A New Efficient Access to Benzoxazoles and Benzothiazoles from Carboxylic Acids under Solvent-Free Conditions. *Synlett* **2007**, *2007* (02), 0313–0317.

Setiyono, E.; Heriyanto; Pringgenies, D.; Shioi, Y.; Kanesaki, Y.; Awai, K.; Brotosudarmo, T. H. P. Sulfur-Containing Carotenoids from a Marine Coral Symbiont Erythrobacter Flavus Strain KJ5. *Marine drugs* **2019**, *17* (6), 349.

Sheldon, R. A.; Brady, D. The Limits to Biocatalysis: Pushing the Envelope. *Chemical communications* **2018**, *54* (48), 6088–6104.

Sheldon, R. A.; Brady, D.; Bode, M. L. The Hitchhiker's Guide to Biocatalysis: Recent Advances in the Use of Enzymes in Organic Synthesis. *Chemical science* **2020**, *11* (10), 2587–2605.

SHIBA, T.; SIMIDU, U. Erythrobacter Longus Gen. Nov., Sp. Nov., an Aerobic Bacterium Which Contains Bacteriochlorophyll a. *International*

Journal of Systematic and Evolutionary Microbiology, 1982, *32*, 211–217. https://doi.org/10.1099/00207713-32-2-211.

Shin-ya, K.; Umeda, Y.; Chijiwa, S.; Furihata, K.; Furihata, K.; Hayakawa, Y.; Seto, H. Mevashuntin, a Novel Metabolite Produced by Inhibition of the Mevalonate Pathway in Streptomyces Prunicolor. *Tetrahedron Letters* **2005**, *46* (8), 1273–1276. https://doi.org/10.1016/j.tetlet.2004.12.134.

Shinde, R.; McGaha, T. L. The Aryl Hydrocarbon Receptor: Connecting Immunity to the Microenvironment. *Trends in Immunology* **2018**, *39* (12), 1005–1020.

Song, J. A Ligand for the Aryl Hydrocarbon Receptor Isolated from Lung. *Proc Natl Acad Sci USA* **2002**, *99* (23), 14694–14699.

Stierle, A. A.; Cardellina II, J. H.; Singleton, F. L. Benzothiazoles from a Putatitve Bacterial Symbiont of the Marine Sponge Tedania Ignis. *Tetrahedron letters* **1991**, *32* (37), 4847–4848.

Stroobant, P.; Young, I. G.; Gibson, F. Mutants of *Escherichia Coli* K-12 Blocked in the Final Reaction of Ubiquinone Biosynthesis: Characterization and Genetic Analysis. *Journal of Bacteriology* **1972**, *109* (1), 134–139. https://doi.org/10.1128/jb.109.1.134-139.1972.

Subramani, R.; Sipkema, D. Marine Rare Actinomycetes: A Promising Source of Structurally Diverse and Unique Novel Natural Products. *Marine Drugs* **2019**, *17* (5). https://doi.org/10.3390/md17050249.

Sufrin, J. R.; Finckbeiner, S.; Oliver, C. M. Marine-Derived Metabolites of S-Adenosylmethionine as Templates for New Anti-Infectives. *Marine Drugs* **2009**, *7* (3), 401–434. https://doi.org/10.3390/md7030401.

Sun, Y.; Jiang, H.; Wu, W.; Zeng, W.; Wu, X. Copper-Catalyzed Synthesis of Substituted Benzothiazoles via Condensation of 2-Aminobenzenethiols with Nitriles. *Organic letters* **2013**, *15* (7), 1598–1601.

Suzuki, H.; Shindo, K.; Ueno, A.; Miura, T.; Takei, M.; Sakakibara, M.; Fukamachi, H.; Tanaka, J.; Higa, T. S1319: A Novel B2-Adrenoceptor Agonist from a Marine Sponge Dysidea Sp. *Bioorganic & Medicinal Chemistry Letters* **1999**, *9* (10), 1361–1364. https://doi.org/10.1016/S0960-894X(99)00205-X.

Takaichi, S.; Furihata, K.; Ishidsu, J.; Shimada, K. Carotenoid Sulphates from the Aerobic Photosynthetic Bacterium, Erythrobacter Longus. *Phytochemistry* **1991**, *30* (10), 3411–3415. https://doi.org/10.1016/0031-9422(91)83219-B.

Takaichi, S.; Maoka, T.; Takasaki, K.; Hanada, S. Carotenoids of Gemmatimonas Aurantiaca (Gemmatimonadetes): Identification of a Novel Carotenoid, Deoxyoscillol 2-Rhamnoside, and Proposed Biosynthetic Pathway of Oscillol 2,2′-Dirhamnoside. *Microbiology*, 2010, *156*, 757–763. https://doi.org/10.1099/mic.0.034249-0.

Takaichi, S.; Shimada, K.; Ishidsu, J. Carotenoids from the Aerobic Photosynthetic Bacterium, Erythrobacter Longus: β-Carotene and Its Hydroxyl Derivatives. *Archives of microbiology* **1990**, *153*, 118–122.

Takamura, T.; Harama, D.; Fukumoto, S.; Nakamura, Y.; Shimokawa, N.; Ishimaru, K.; Ikegami, S.; Makino, S.; Kitamura, M.; Nakao, A. Lactobacillus Bulgaricus OLL1181 Activates the Aryl Hydrocarbon Receptor Pathway and Inhibits Colitis. *Immunology and cell biology* **2011**, *89* (7), 817–822.

Tan, D.; Jamieson, C. S.; Ohashi, M.; Tang, M.-C.; Houk, K. N.; Tang, Y. Genome-Mined Diels–Alderase Catalyzes Formation of the Cis-Octahydrodecalins of Varicidin A and B. *J. Am. Chem. Soc.* **2019**, *141* (2), 769–773. https://doi.org/10.1021/jacs.8b12010.

Tan, F.; Shi, B.; Li, J.; Wu, W.; Zhang, J. Design and Synthesis of New 2-Aryl-4,5-Dihydro-Thiazole Analogues: In Vitro Antibacterial Activities and Preliminary Mechanism of Action. *Molecules* **2015**, *20*, 20118–20130.

Tareen, S.; Risdian, C.; Müsken, M.; Wink, J. Qipengyuania Pacifica Sp. Nov., a Novel Carotenoid-Producing Marine Bacterium of the Family Erythrobacteraceae, Isolated from Sponge (Demospongiae), and Antimicrobial Potential of Its Crude Extract. *Diversity* **2022**, *14* (4), 295.

Tian, B.; Hua, Y. Carotenoid Biosynthesis in Extremophilic Deinococcus–Thermus Bacteria. *Trends in Microbiology* **2010**, *18* (11), 512–520. https://doi.org/10.1016/j.tim.2010.07.007.

Ting, C.P; Funk, M. A.; Halaby, S. L.; Zhang, Z.; Gonen, T.; van der Donk, W. A. Use of a Scaffold Peptide in Thebiosynthesis of Amino Acid–Derivednatural Products. *Science* **2019**, *365* (6450), 280–284.

Tongsook, C.; Uhl, M. K.; Jankowitsch, F.; Mack, M.; Gruber, K.; Macheroux, P. Structural and Kinetic Studies on RosA, the Enzyme Catalysing the Methylation of 8-Demethyl-8-Amino-d-Riboflavin to the Antibiotic Roseoflavin. *The FEBS Journal* **2016**, *283* (8), 1531–1549. https://doi.org/10.1111/febs.13690.

Trevino, A. E.; Zhang, F. Chapter Eight - Genome Editing Using Cas9 Nickases. In *The Use of CRISPR/Cas9, ZFNs, and TALENs in Generating Site-Specific Genome Alterations*; Doudna, J. A., Sontheimer, E. J., Eds.; Methods in Enzymology; Academic Press, 2014; Vol. 546, pp 161–174. https://doi.org/10.1016/B978-0-12-801185-0.00008-8.

Us-Camas, R.; Aguilar-Espinosa, M.; Rodriguez-Campos, J.; Rivera-Madrid, R. Identifying Bixa Orellana L. New Carotenoid Cleavage Dioxygenases 1 and 4 Potentially Involved in Bixin Biosynthesis. *Front. Plant Sci.* **2022**, *13*.

Walsh, C. T.; Fischbach, M. A. Natural Products Version 2.0: Connecting Genes to Molecules. *J. Am. Chem. Soc.* **2010**, *132* (8), 2469–2493. https://doi.org/10.1021/ja909118a.

Wang, H.; Fewer, D. P.; Holm, L.; Rouhiainen, L.; Sivonen, K. Atlas of Nonribosomal Peptide and Polyketide Biosynthetic Pathways Reveals Common Occurrence of Nonmodular Enzymes. *Proceedings of the National Academy of Sciences* **2014**, *111* (25), 9259–9264. https://doi.org/10.1073/pnas.1401734111.

Wang, M. et al. Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking. *Nature Biotechnology* **2016**, *34* (8), 828–837. https://doi.org/10.1038/nbt.3597.

Wang, W.; Kappock, T. J.; Stubbe, J.; Ealick, S. E. X-Ray Crystal Structure of Glycinamide Ribonucleotide Synthetase from Escherichia Coli,. *Biochemistry* **1998**, *37* (45), 15647–15662. https://doi.org/10.1021/bi981405n.

Williams, D. R.; Lowder, P. D.; Gu, Y. G.; Brooks, D. A. Studies of Mild Dehydrogenations in Heterocyclic Systems. *Tetrahedron Letters* **1997**, *38*, 331–334.

Wilson, K.; Rond, T. de; Burkhardt, I.; Steele, T. S.; Schäfer, R. J. B.; Podell, S.; Allen, E. E.; Moore, B. S. Terpene Biosynthesis in Marine Sponge Animals. *Proceedings of the National Academy of Sciences* **2023**, *120* (9), e2220934120. https://doi.org/10.1073/pnas.2220934120.

Wu, G.; Williams, H. D.; Zamanian, M.; Gibson, F.; Poole, R. K. Isolation and Characterization of Escherichia Coli Mutants Affected in Aerobic Respiration: The Cloning and Nucleotide Sequence of ubiG. *Microbiology*, 1992, *138*, 2101–2112. https://doi.org/10.1099/00221287-138-10-2101.

Wu, M.; Shi, Z.; Ma, T. Enhancement of Transparent Hydrogel Sanxan Production in Sphingomonas Sanxanigenens NX02 via Rational and Random Gene Manipulation. *Carbohydrate Polymers* **2018**, *189*, 210–217.

Wu, S.; Snajdrova, R.; Moore, J. C.; Baldenius, K.; Bornscheuer, U. T. Biocatalysis: Enzymatic Synthesis for Industrial Applications. *Angewandte Chemie International Edition* **2021**, *60* (1), 88–119. https://doi.org/10.1002/anie.202006648.

Xiao, R.; Hao, W.; Ai, J.; Cai, M.-Z. A Practical Synthesis of 2-Aminobenzothiazoles via the Tandem Reactions of 2-Haloanilines with Isothiocyanates Catalyzed by Immobilization of Copper in MCM-41. *Journal of Organometallic Chemistry* **2012**, *705*, 44–50.

Xie, C.-L.; Zhang, D.; Guo, K.-Q.; Yan, Q.-X.; Zou, Z.-B.; He, Z.-H.; Wu, Z.; Zhang, X.-K.; Chen, H.-F.; Yang, X.-W. Meroterpenthiazole A, a Unique Meroterpenoid from the Deep-Sea-Derived Penicillium Allii-Sativi, Significantly Inhibited Retinoid X Receptor (RXR)-α Transcriptional Effect. *Chinese Chemical Letters* **2022**, *33* (4), 2057–2059. https://doi.org/10.1016/j.cclet.2021.09.073.

Yamada, Y.; Kuzuyama, T.; Komatsu, M.; Shin-ya, K.; Omura, S.; Cane, D. E.; Ikeda, H. Terpene Synthases Are Widely Distributed in Bacteria. *Proceedings of the National Academy of Sciences* **2015**, *112* (3), 857–862. https://doi.org/10.1073/pnas.1422108112.

Yang, X.; Wu, W.; Li, H.; Zhang, M.; Chu, Z.; Wang, X.; Sun, P. Natural Occurrence, Bioactivity, and Biosynthesis of Triene-Ansamycins. *European Journal of Medicinal Chemistry* **2022**, *244*, 114815. https://doi.org/10.1016/j.ejmech.2022.114815.

Yi, D.; Bayer, T.; Badenhorst, C. P.; Wu, S.; Doerr, M.; Höhne, M.; Bornscheuer, U. T. Recent Trends in Biocatalysis. *Chemical Society Reviews* **2021**, *50* (14), 8003–8049.

Yoon, B.-J.; Lee, D.-H.; Oh, D.-C. Erythrobacter Jejuensis Sp. Nov., Isolated from Seawater. *International Journal of Systematic and Evolutionary Microbiology* **2013**, *63*, 1421–1426.

Youcai Hu and John B. MacMillan. Erythrazoles A B, Cytotoxic Benzothiazoles from a Marine-Derived Erythrobacter Sp. *Organic Letters* **2011**, *13* (21), 6580–6583.

Yurkov, V.; Stackebrandt, E.; Holmes, A.; Fuerst, J. A.; Hugenholtz, P.; Golecki, J.; Gad'on, N.; Gorlenko, V. M.; Kompantseva, E. I.; Drews, G. Phylogenetic Positions of Novel Aerobic, Bacteriochlorophyll a-Containing Bacteria and Description of Roseococcus Thiosulfatophilus Gen. Nov., Sp. Nov., Erythromicrobium Ramosum Gen. Nov., Sp. Nov., and Erythrobacter Litoralis Sp. Nov. *International Journal of Systematic and Evolutionary Microbiology* **1994**, *44* (3), 427–434.

Zhang, D.; Shrestha, B.; Li, Z.; Tan, T. Ubiquinone-10 Production Using Agrobacterium Tumefaciens Dps Gene in Escherichia Coli by Coexpression System. *Molecular Biotechnology* **2007**, *35* (1), 1–14. https://doi.org/10.1385/MB:35:1:1.

Zhang, R.; He, J.; Dong, Z.; Liu, G.; Yin, Y.; Zhang, X.; Li, Q.; Ren, Y.; Yang, Y.; Liu, W.; Chen, X.; Xia, W.; Duan, K.; Hao, F.; Lin, Z.; Yang, J.; Chang, Z.; Zhao, R.; Wan, W.; Lu, S.; Peng, Y.; Ge, S.; Wang, W.; Li, X. Genomic and Experimental Data Provide New Insights into Luciferin Biosynthesis and Bioluminescence Evolution in Fireflies. *Scientific Reports* **2020**, *10* (1), 15882. https://doi.org/10.1038/s41598-020-72900-z.

Zhang, X.; Guo, J.; Cheng, F.; Li, S. Cytochrome P450 Enzymes in Fungal Natural Product Biosynthesis. *Natural Product Reports* **2021**, *38* (6), 1072–1099.

Zheng, Q.; Lin, W.; Liu, Y.; Chen, C.; Jiao, N. A Comparison of 14 Erythrobacter Genomes Provides Insights into the Genomic Divergence and Scattered Distribution of Phototrophs. *Frontiers in Microbiology* **2016**, *7*. https://doi.org/10.3389/fmicb.2016.00984.

Zhou, D.; White, R. H. Early Steps of Isoprenoid Biosynthesis in Escherichia Coli. *Biochemical Journal* **1991**, *273* (3), 627–634. https://doi.org/10.1042/bj2730627.

Zhou, D.; White, R. H. Early Steps of Isoprenoid Biosynthesis in Escherichia Coli. *Biochemical Journal* **1991**, *273* (3), 627–634. https://doi.org/10.1042/bj2730627.

Zhu, X.; Yuasa, M.; Okada, K.; Suzuki, K.; Nakagawa, T.; Kawamukai, M.; Matsuda, H. Production of Ubiquinone in Escherichia Coli by Expression of Various Genes Responsible for Ubiquinone Biosynthesis. *Journal of Fermentation and Bioengineering* **1995**, *79* (5), 493–495. https://doi.org/10.1016/0922-338X(95)91268-A.

Ziemert, N.; Alanjary, M.; Weber, T. The Evolution of Genome Mining in Microbes–a Review. *Natural product reports* **2016**, *33* (8), 988–1005.