BUILDING CHEMICAL TOOLS FROM THE INDOLMYCIN BIOSYNTHETIC PATHWAY

by

ELESHA RENE HOFFARTH

B.Sc., The University of Calgary, 2014

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Chemistry)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2021

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Building chemical tools from the indolmycin biosynthetic pathway.

submitted by Elesha Rene Hoffarth in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

in Chemistry

Examining Committee:

Dr. Katherine S. Ryan, Associate Professor, Chemistry, University of British Columbia
Supervisor

Dr. Stephen G. Withers, Professor, Chemistry, University of British Columbia
Supervisory Committee Member

Dr. Laurel L. Schafer, Professor, Chemistry, University of British Columbia
University Examiner

Dr. Natalie Strynadka, Professor, Biochemistry and Molecular Biology, University of British Columbia
University Examiner

Dr. Erika Plettner, Professor, Chemistry, Simon Fraser University
External Examiner

Additional Supervisory Committee Members:

Dr. Mark MacLachlan, Professor, Chemistry, University of British Columbia
Supervisory Committee Member

Dr. Martin Tanner, Professor, Chemistry, University of British Columbia
Supervisory Committee Member

Abstract

Natural products are essential to the discovery of new drugs, including antibiotics. Industrial interest in natural products has declined since the 1980s, but advances in biocatalysis and biosynthetic knowledge have helped revive interest in natural products as these advances contribute to more feasible discovery, production and derivatization of natural products. To continue this industrial interest in natural products and their related compounds, work should be done to accumulate more biosynthetic knowledge to further improve methods of discovery, production and derivatization and facilitate more widespread use of biocatalysts. Indolmycin is a natural product with antibiotic activities against methicillin-resistant *Staphylococcus aureus*, *Helicobacter pylori* and *Plasmodium falciparum*, whose biosynthetic pathway is shown here to be a source of new biochemical tools.

First, in order to better understand the unique reactivity of the rare oxygen- and pyridoxal 5′-phosphate (PLP)-dependent arginine desaturases discovered from the indolmycin biosynthetic pathway, the first X-ray crystal structure of an arginine desaturase was solved. This structure showed an active site that was highly similar to the related oxygen- and PLP-dependent hydroxylases. Catalytic residues for the arginine desaturases were uncovered by creating mutagenic variants based on the crystal structure information. Second, sequence similarity analysis and side-product analysis were done, which further supported a higher similarity to the arginine hydroxylases than was originally predicted. Additionally, superoxide was shown to be an intermediate of the arginine oxidase mechanism for the first time through EPR and cytochrome c assays. Based on this information, a unified mechanistic hypothesis is proposed which suggests that desaturation and hydroxylation may be differentiated by the presence/position of water in the active site.

Third, the indolmycin biosynthetic enzymes are used in conjunction with a promiscuous tryptophan synthase and a three-step chemical synthesis to produce indolmycin and several novel halogenated derivatives. Derivatives with fluorinated indole substitutions showed a moderate bioactivity against *S. aureus* and could be useful in developing indolmycin for clinical use. Overall, this work uses the indolmycin biosynthetic enzymes to expand the known biocatalytic repertoire with the hope that it can contribute to more widespread use of biocatalysts in the production of natural product-derived molecules.

Lay Summary

Natural products are molecules from natural sources, such as bacteria, fungi and plants. The discovery and production of natural products is essential for drug development, as nearly half of all approved drugs are derived from natural products. Nature produces these molecules using biosynthetic enzymes. Advances in genomics and knowledge of biosynthetic enzymes has allowed biosynthetic enzymes to be more frequently harnessed by companies to make useful natural products. Here, the biosynthetic enzymes used by nature to make indolmycin, a natural product, are studied for their unique characteristics and their ability to make non-natural, indolmycin-like molecules. This study expands the current knowledge of biosynthetic enzymes, which helps make industrial use of biosynthetic enzymes more accessible.

Preface

Some parts of **Chapter 1** were published as a review article: Hoffarth, E. R., Rothchild, K. W. & Ryan, K. S. Emergence of oxygen- and pyridoxal phosphate-dependent reactions. *FEBS J.* **287**, 1403–1428 (2020). E. R. H. (the author) wrote most of the sections based on the manuscript and all other section in **Chapter 1**. **Sections 1.4.3.3**, **1.4.4.1** and **1.4.4.2** were originally drafted by K. W. R. and parts of **Section 1.3.1** were originally drafted by K. S. R (PhD supervisor). The manuscript and this chapter were also reviewed and edited by K. S. R.

A version of **Chapters 2** and **3**, combined, has been submitted for publication: Hoffarth, E. R., Haaviet, K. C., Kuatsjah, E., MacNeil, G. A., Garcia-Borràs, M., Saroya, S., Walsby, C. J., Eltis, L. D., Houk, K. N., Ryan, K. S. A shared mechanistic pathway for pyridoxal phosphate-dependent arginine oxidases (in revision). E. R. H. wrote most of these chapters and completed most experiments and data analyses. E. R. H. and K. S. R designed this research project. Some of the Ind4 homolgs described in **Chapters 2** and **3** were partially purified by S. S. In **Chapter 2**, **Sections 2.2.5** and **2.3.1.3,** the DFT calculations were completed by M. G.-B. and the methods describing these calculations in **Section 2.2.5** were originally drafted by K. C. H, both under the supervision of K. N. H. Experiments and data analysis to determine the kinetic parameters of Plu4 described in **Chapter 3**, **Sections 3.2.5** and **3.3.2.1** were completed by E. K under the supervision of L. D. E. Under the supervision of C. J. W., EPR data in **Section 3.3.2.3** was collected and processed by G. A. M., who also wrote the EPR instrumentation methods in **Section 3.2.6**. The DFT calculations from the submitted manuscript that are briefly summarized in **Section 3.3.3**, were completed by K. C. H under the supervision of K. N. H. The manuscript and these chapters were reviewed and edited by K. S. R.

A version of **Chapter 4** has been submitted for publication: Hoffarth, E. R., Kong, S., He, H.-Y., Ryan, K. S. An engineered biosynthetic-synthetic platform for production of halogenated indolmycin antibiotics (in revision). E. R. H. wrote this chapter and completed most experiments and data analyses. E. R. H. and K. S. R designed this research project. H.-Y. H. helped design and set up the synthetic aspects of the project. S. K. and H.-Y. H. helped isolate some of the indolmycenic acid derivatives and completed synthetic steps for some of the indolmycin derivatives. NMR data was collected by staff at the NMR Facility in the Department of Chemistry at the University of British Columbia. HR-MS data was collected by staff at the University of British Columbia Mass Spectrometry Center. Disk diffusion and MIC bioactivity

data was collected by Jessie Chen at the Biolocal Services Laboratory in the Department of Chemistry at the University of British Columbia. The manuscript and this chapter was reviewed and edited by K. S. R.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AADC: amino acid decarboxylase

AARS: aminoacyl-tRNA synthetase

AAS: amino acid aldehyde synthase

AEE: L-Ala-D/L-Glu epimerases

AMD: α-methylDOPA

amd: α-methyldopa hypersensitive gene

AMP: adenosine monophosphate

AONS: 8-amino-7-oxononanoate synthase

CarU: uridine 5′-carboxoamide

CDC: Center for Disease Control and Prevention

CMH: cyclic hydroxylamine 1-hydroxy-3-methoxycarbonyl-2,2,5,5-tetramethylpyrrolidine

DFT: density functional theory

DHPAA: 3,4-dihydroxyphenylacetaldehyde

DHPAAS: DHPAA synthase

DMSO: dimethyl sulfoxide

DOPA: 3,4-dihydroxyphenylalanine

DDC: DOPA decarboxylase

EFI-EST: Enzyme Function Initiative-Enzyme Similarity Tool

EFI-GNT: Enzyme Function Initiative-Genomic Neighbourhood Tool

EPR: electron paramagnetic resonance

ESI-MS: electrospray ionization-mass spectrometry

GAD: glutamate decarboxylase

GlyU: 5′-glycyluridine

G3P: glyceraldehyde-3-phosphate

HPLC: high-performance liquid chromatography

HR-MS: high resolution-mass spectrometry

L-HTP: L-hydroxytryptophan

HTS: high-throughput screening

IDT: Integrated DNA Technologies

IPTG: isopropyl β-D-1-thiogalactopyranoside

LB: Luria-Bertani

TOF-MS: time of flight-mass spectrometer

LC-MS: liquid chromatography-mass spectrometry

MHB: Mueller-Hinton Broth

MIC: minimum inhibitory concentration

MRSA: methicillin-resistant *Staphylococcus aureus*

NAAAR: *N*-acylamino acid racemase

NADH: nicotinamide adenine dinucleotide

NADPH: nicotinamide adenine dinucleotide phosphate

NEB: New England Bio Labs

NMR: nuclear magnetic resonance

ODC: ornithine decarboxylase

OPD: *o*-phenylenediamine

OSBS: *o*-succinylbenzoate synthase

PAAS: phenylacetaldehyde synthase

PCR: polymerase chain reaction

PHA: phenylacetaldehyde

PLP: pyridoxal 5′-phosphate

PMP: pyridoxamine 5′-phosphate

PTPS: 6-pyruvoyl-tetrahydropterin synthases

RMSD: root-mean-square deviation

SAM: *S*-adenosylmethionine

SDS-PAGE: sodium dodecyl sulfate-polyacrylamide gel electrophoresis

SOD: superoxide dismutase

SSN: sequence similarity network

TLS: translation-liberation-screw

TrpA: tryptophan synthase α-subunit

TrpB: tryptophan synthase β-subunit

TrpRS: tryptophanyl-tRNA synthetase

TrpS: tryptophan synthase

UV-Vis: ultraviolet-visible

Acknowledgements

I would like to thank my supervisor, Dr. Katherine Ryan, for constantly encouraging me throughout my graduate studies and being incredibly supportive of my career goals. She has taught me so much about the research process and I am constantly amazed at all of the expertise she had to share about communicating my work. Thank you to Melanie Higgins, Hai-Yan He, Yi-Ling Du and Lona Alkhalaf, who mentored me during my graduate studies and to all other Ryan lab members, past and present, that I have had the pleasure to work with. I would also like to thank my supervisory committee members for their encouragement and helpful suggestions to improve my thesis. Thank you to the many wonderful friends I have made during my time at the University of British Columbia, namely Elizabeth Fisher, Kelsi Lix, Phil Danby, Trevor Cross and Chloe Gerak, whom I am so grateful for.

I am so appreciative of the staff in the Chemistry Department at UBC, including Elena Polishchuck and Jessie Chen in the Biological Services Laboratory for being extremely helpful because their hard work was essential to making many of my research tasks run smoothly. Maria Ezhova in the NMR Facility was so valuable in obtaining and interpreting my NMR spectra. Ben Herring in the Shared Instrument Facility was so friendly and accommodating of my abundant LC-MS needs. Thank you to Anne Thomas who was an amazing teaching assistant supervisor to me and I appreciate that she kept me on in teaching positions that I have enjoyed very much throughout these last few years. I have also enjoyed our many discussions about the best food in the West End of Vancouver.

I am extremely grateful to my parents, Lisa and Lenard Hoffarth, who shaped my independence, confidence and curiosity that I have needed to complete my research. They have taught me how to work hard and persevere to achieve my goals and they have always supported my desire to further my education. Most of all, I would like to thank my fiancé, Jack, for his unwavering encouragement throughout my entire post-secondary and graduate education. He is especially appreciated for trying very hard to give me the quiet time that I needed to complete work and for occasionally picking me up from UBC when it was raining. He has comforted me through the hardest times and celebrated with me through the happiest moments.

Chapter 1: Introduction[1]

To begin exploring the development of biochemical tools from the indolmycin biosynthetic pathway, this section will review relevant background information. First, the history of natural products will be described, followed by an analysis of natural products as antibiotics and the importance of biocatalysis in producing natural product-derived compounds. Then, indolmycin will be introduced with a focus on its discovery, bioactivity, synthesis and biosynthesis. Pyridoxal 5′-phosphate (PLP)-dependent enzymes are an essential part of indolmycin biosynthesis and will be briefly described with a focus on biocatalytic uses to introduce the unique and rare oxygen- and PLP-dependent enzymes. The scope and current mechanistic understanding of the oxygen- and PLP-dependent enzymes, including Ind4, an indolmycin biosynthetic enzyme, will be reviewed. Finally, the aims of this thesis will be described as they relate to use of the indolmycin biosynthetic pathway for building biochemical tools.

1.1 Natural products

1.1.1 A brief history and current perspective

Natural products are a diverse collection of molecules derived from natural sources, typically from the secondary metabolism of plants, bacteria, fungi and marine organisms.[1,2] Secondary metabolites provide their producing organisms with a survival advantage in certain environments, but are not required for the organism's survival under all circumstances.[1,2] These molecules have been widely useful, not only to the producing organisms, but also for industrial, medicinal and scientific purposes because of their diverse properties, ranging from anticancer activity to phytotoxicity. Antibiotic development in particular has greatly benefited from the discovery of new natural products.[3,4] Some of the earliest examples of human use of natural products come from ancient Egypt and Mesopotamia and were primarily from plant sources.[2,5] Notable examples of natural products discovered from plants and developed for pharmaceutical use are salicin, from the willow tree (*Salix alba* L.), morphine from the opium poppy (*Papaver somniferum* L.) and artemisinin from *Artemisia annua* L., which all have roots in ancient

[1] Parts of this chapter have been published: Hoffarth, E. R., Rothchild, K. W. & Ryan, K. S. Emergence of oxygen- and pyridoxal phosphate-dependent reactions. *FEBS J.* **287**, 1403–1428 (2020).

medicine.[2,6] Evidence of microbial sources of disease-preventing therapies can be seen for thousands of years in the use of medicinal soils and mouldy bread to treat open wounds.[7] However, research into microbial sources of natural products were not of particular focus until the discovery of penicillin by Fleming in 1928, which spurred the "Golden Era" of antibiotic discovery and the discovery of other microbial natural products as a result (**Figure 1.1**).[2,5]

By the 1990s, the "Golden Era" had ended as novel structures were being discovered less frequently. Natural product discoveries from pharmaceutical companies were being discontinued in favour of synthetic libraries due to incompatibility of the natural product libraries with high-throughput screening (HTS) methods that were becoming popular.[8] The pharmaceutical environment and the blockbuster model of drug development necessitates accurate, quick and profitable projects via HTS as a result of pressure from shareholder expectations following success in the 1990s and early 2000s. The high structural complexity and high probability of redundant characterization are just some of the reasons natural product research has been largely discontinued.[8] However, natural chemical diversity is unmatched in its ability to produce leads and inspire new compounds with close to half of all small molecule drugs approved between 1981-2014 being natural products or related molecules.[2,9–11] More recently, advances in genomics, screening compatibility, analytical techniques, and biosynthetic knowledge have all made natural product development more feasible, which could potentially lead to a new revolution in natural product discovery.[1,2,6,8,12]

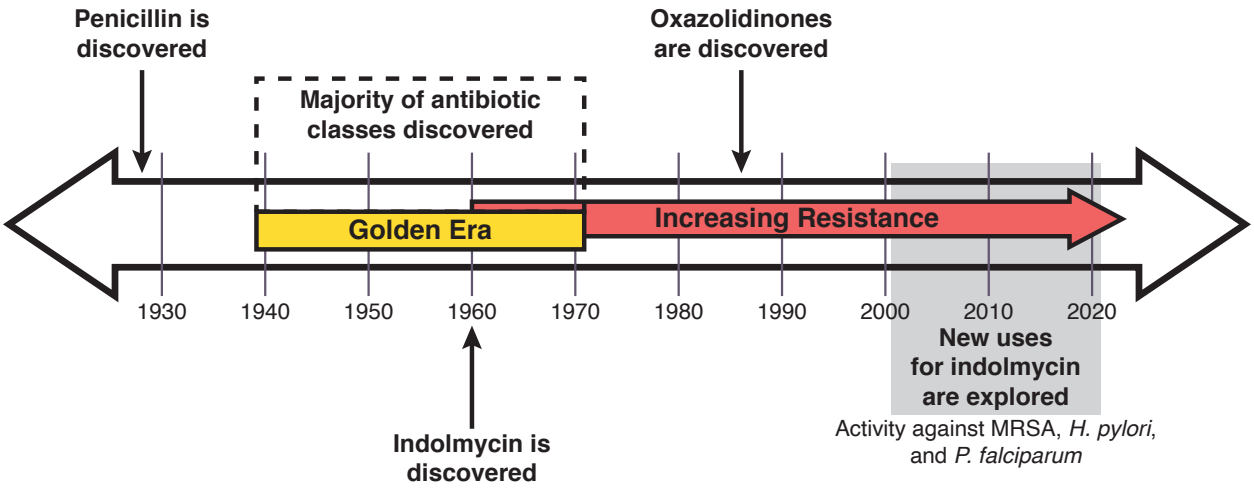


Figure 1.1: Timeline of antibiotic discovery and resistance beginning with penicillin discovery in 1928, through to 2020. Timeline is adapted from previous review articles.[7,15,16,18]

1.1.2 Antibiotics

Natural product research has been imperative to the discovery of new antibiotics. While examples of both natural and synthetic antibiotics exist, the importance of natural product discovery in the success of antibiotic development is not trivial.[13,14] The first natural product-based, and possibly most famous, antibiotic discovered was penicillin. In 1928, Fleming classically observed inhibited bacterial growth around a contaminating fungus in a petri dish. This inhibited growth was later shown to be the result of the antibiotic penicillin, whose beta-lactam structure was solved by Dorothy Hodgkin in 1945.[7,15,16] The discovery of penicillin inspired the exploration of many new antibiotics and other natural products during a period often referred to as the 'Golden Era' of antibiotic discovery (**Figure 1.1**), which ultimately saved countless lives and revolutionized modern medicine.[7,15,16] However, the initial rapid discovery during the "Golden Era" resulted in misuse of antibiotics, which caused a quick rise in antibiotic resistance.[7]

Antibiotic resistance is an ancient mechanism bacteria have evolved to defend themselves against their own and other antimicrobials.[15,17] Some antimicrobial resistance methods that bacteria have evolved include inactivation by an enzyme, efflux, mutations to decrease binding efficiency to the target, overproduction of the target, and decreased uptake.[16,18] Phylogenetic studies of some β-lactamases, the enzyme responsible for inactivating β-lactam antibiotics like penicillin, revealed that they may have originated more than two billion years ago.[17] Penicillinase, the enzyme responsible for inactivating penicillin in resistant strains, was discovered before penicillin was even put into clinical use.[16] While antibiotic resistance mechanisms existed long before human use of antibiotics, it is undeniable that human use has increased the environmental presence of antibiotic resistance through use in medicine, agriculture, aquaculture, and waste disposal.[16] Therefore, antimicrobial resistance is an inevitable consequence of antibiotic use that must be effectively managed.[16,18] It is predicted that millions of people will die each year from antibiotic-resistant infections within a few decades if the issue is not dealt with.[7,15]

Unfortunately, the antibiotic resistance problem has been developing concurrently with dwindling antibiotic approval, ultimately exacerbating the effects of resistance.[19] Reduced development of natural products following the "Golden Era" of antibiotics also resulted in reduced antibiotic discovery.[7] The last new antibiotic classes discovered were the lipopeptides

with daptomycin in 1986 and the oxazolidinones with linezolid in 1987.[15] However advances to natural product discovery has translated to new antibiotic discoveries.[4,7,12,14,15,17,18] One method to find compounds helpful against antibiotic-resistant strains has been to revisit forgotten molecules that were originally discovered during the "Golden Era", as some have shown remarkable activity against antibiotic-resistant bacteria.[14–16] These new discoveries are a promising start to finding an abundance of new antibiotics and may signify a new era of antibiotic discovery to help overcome challenges with antibiotic resistance. However, overcoming these complex challenges will require cooperation from many groups of people to achieve success, including scientists, healthcare workers, the pharmaceutical and agricultural industries, policy makers, legislative bodies, and the general public.[17]

1.1.3 The importance of biocatalysis in natural products

Advances in biosynthetic and biocatalytic knowledge have contributed to reinvigorated interest in natural products because of the powerful potential for this knowledge to aid in finding, producing and diversifying natural products more easily than extraction from natural producers or chemical methods alone.[6,8,20] Due to advances in genomics, knowledge and characterization of existing biosynthetic pathways can lead to the prediction and discovery of novel natural products based on known biosynthetic scaffolds.[3] New molecules with known pharmacophores can be predicted from genomic data once the biosynthetic genes responsible for building the pharmacophore are known. Investigation of species containing these biosynthetic genes would lead to more targeted searches for specific molecules with a desirable pharmocophore.[1] Additionally, expression of silent, or "cryptic", pathways has led to discoveries of new natural products that may have otherwise gone unnoticed without genomics and biosynthetic studies.[6] However, many natural products cannot be predicted from their biosynthetic pathway and each pathway may need to be identified and characterized on an individual basis to understand the biomolecules they are making, especially for the most novel of cases.[1]

Methods of producing natural product-based compounds derived from biosynthetic knowledge include heterologous expression and metabolic engineering, among others,[3,5,21,22] while methods of diversification include precursor feeding, synthetic biology, enzyme engineering and semi-synthesis.[3,4,21,23–25] A key example of a natural product-derived compound produced though biocatalytic means was the recent biocatalytic cascade created by Merck for the

production of a potential HIV treatment, islatravir.[26] Islatravir is a nucleoside analog with an alkynyl-substitution at the C4 position of 2-deoxyribose ring and a fluorine-substitution on the nucleobase. The biocatalytic cascade used to produce islatravir involved the use of nine enzymes, five of which were engineered for optimal performance, including tolerance of non-natural substrates. Compared to previous synthetic methods, this biocatalytic cascade showed improved atom economy and a reduction in the number of steps. Despite success with islatravir, major hurdles to more widespread biosynthetic production of natural product and their derivatives still exist. A few of these hurdles surround unannotated pathways and enzymes, incomplete knowledge of pathways and their enzymatic transformations and difficulty finding an enzyme to catalyze a desired reaction.[22] Additionally, attempts to engineer individual enzymes to suit a desired reaction involve an in-depth knowledge of the enzymatic chemistry and may require a crystal structure.[27] Overall, advances to biosynthetic and biocatalytic knowledge may help reinvigorate industrial interest in natural product development by aiding in the discovery, production and diversification of natural products.

1.2 Indolmycin

Indolmycin is a tryptophan-derived antibiotic originally discovered from *Streptomyces griseus* ATCC12648 in 1960[28] (**Figure 1.1**) that acts as a competitive inhibitor of prokaryotic tryptophanyl-tRNA synthetase (TrpRS) based on structural similarity to tryptophan (**Figure 1.2**).[29] It contains an oxazolinone ring connected to an indole through a chiral methine. A search for L-tryptophan antimetabolites in *S. griseus* identified indolmycin and *N*-desmethyl-*C*-desmethylindolmycin as L-tryptophan antagonists.[30] *N*-desmethyl and *C*-desmethylindolmycin, along with stereoisomers of indolmycin, all show reduced bioactivity when compared to indolmycin (**Figure 1.2**).[31] Although originally patented by Pfizer, indolmycin was not developed into a clinically used pharmaceutical because it had a narrow spectrum of activity and potential liver toxicity.[32–37] However, challenges that arose in antimicrobial resistance since the "Golden Era" of antibiotic discovery have reinvigorated interest in indolmycin (**Figure 1.1**).[32] Indolmycin was more recently shown to be active against methicillin-resistant *Staphylococcus aureus* (MRSA),[38] *Helicobacter pylori*[32] and *Plasmodium falciparum.*[39] Revived interest in indolmycin even resulted in it being patented again by Takeda Pharmaceutical Company Ltd. for the treatment of *H. pylori* infections along with a method of production.[40–43] However, due to the

lack of early interest in indolmycin, very few attempts have been made to improve the production of indolmycin and the derivatives that would potentially be needed for pharmaceutical development. Therefore, costs of purchasing the compound for research purposes have been kept very high at $185 USD per mg (Toronto Research Chemicals), as of July 2020, and no derivatives of the molecule are available for purchase.

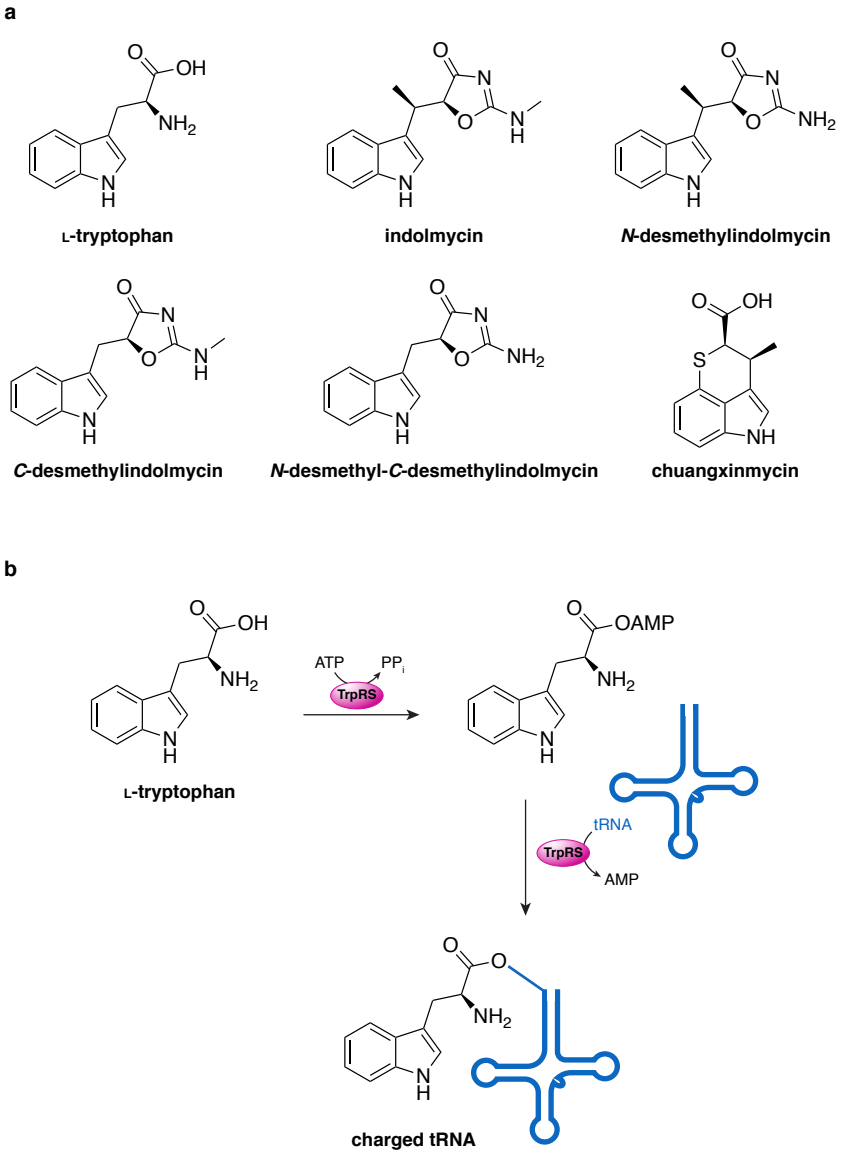


Figure 1.2: TrpRS-binding molecules and the TrpRS mechanism. (**a**) Structures of tryptophan, a TrpRS substrate, and TrpRS inhibitors, indolmycin and chuangxinmycin.[45,47] Indolmycin analogues with reduced bioactivity are also shown.[30,31] (**b**) The process of making charged tRNA using TrpRS is shown.

1.2.1 Bioactivity

TrpRS is a type of aminoacyl-tRNA synthetase (AARS) responsible for covalently attaching tryptophan to its specific tRNA molecule for further use in protein translation.[44,45] The process occurs in two steps beginning with adenylation of tryptophan and release of pyrophosphate. Then, a covalent bond is formed between the tryptophan-specific tRNA and tryptophan, releasing adenosine monophosphate (AMP) and producing a "charged tRNA" molecule (**Figure 1.2**). The essential role of AARSs in protein translation makes them attractive drug targets against pathogens, as therapies targeting protein translation are commonly used as antimicrobials.[14,44] Additionally, evolutionary differences between prokaryotic and eukaryotic AARSs allow these potential drugs to target pathogenic enzymes specifically.[45,46] Mupirocin is the only AARS inhibitor currently in clinical use, in which it is used against *S. aureus* infections by targeting the *S. aureus* isoleucine-tRNA synthetase.[44,46,47] Indolmycin and chuangxinmycin are both competitive inhibitors of TrpRS and mimic the tryptophan substrate (**Figure 1.2**).[45,47] The biosynthetic gene clusters for both indolmycin and chuangxinmycin were discovered by searching producer genomes for colocalized TrpRS encoding genes that could confer resistance in the organisms, which may be an effective strategy to identify other biosynthetic gene clusters for AARS-targeting molecules.[48,49]

Initial interest in indolmycin began after it showed activity against a variety of bacteria.[28,31] Tests on *Escherichia coli* cell lysates revealed that indolmycin targets prokaryotic TrpRS more specifically than mammalian TrpRS from calf and rat liver.[29] It was later shown that indolmycin can inhibit the *E. coli* TrpRS with an IC_{50} of 9.25 nM and the *H. pylori* TrpRS at 12.2 nM compared to 4.04 mM for TrpRS from bovine liver, making it highly selective for the prokaryotic enzyme.[32] Despite this specific inhibition of the prokaryotic TrpRSs, the binding affinity of tryptophan to both the eukaryotic and prokaryotic targets was similar (K_M = ~2 μM), which led researchers to discover that inhibition of the bacterial TrpRS was mechanism-based.[50] Through structural analysis of the *Bacillus stearothermophilus* TrpRS with indolmycin bound to the active site, it was found that indolmycin stabilized the Mg^{2+} and ATP complex in the prokaryotic TrpRS, resulting in reaction inhibition.[50] Indolmycin was also shown to be a competitive inhibitor[29] and was effective against sepsis caused by an antibiotic-resistant strain of *Staphylococcus* in mice.[31] Two types of resistance towards indolmycin have been observed: high-level resistance as the result of an H43N mutation of TrpRS and low-level resistance

potentially from mutations to the uptake system.[38] However, it also may also be possible for high-level resistance to occur due to horizontal gene transfer of the gene encoding a resistant TrpRS from a resistant strain.[36]

Despite showing inhibitory qualities against the *E. coli* TrpRS, indolmycin showed low levels of activity against gram negative bacteria and variable activity against many gram positive bacteria, making it a narrow spectrum antibiotic.[32–36] The low activity against gram negative bacteria is thought to be due to an inability to cross the hydrophilic outer lipopolysaccharide membrane to reach the amino acid transport systems, as indolmycin was calculated to be 84.2% more lipophilic than tryptophan.[33] Additionally, lower activities observed in gram positive bacteria have been attributed to differences in aromatic amino acid uptake mechanisms between different species.[33] A narrow spectrum of activity was not the only reason that development of indolmycin was originally discontinued. Indolmycin was shown to be a potent inhibitor of tryptophan catabolism enzymes, tryptophan pyrrolase and tryptophan decarboxylase in rat livers, leading to potential liver toxicity.[34,36,37] *C*-desmethylindolmycin, although weaker antimicrobially, demonstrated less of an effect on the eukaryotic system, demonstrating the importance of exploring derivative compounds in the future.[37]

More recent interest in indolmycin has come from exploration of new antibiotic uses for the compound (**Figure 1.1**). First, indolmycin was found to be a potent and selective bactericidal agent against *H. pylori* in 2001, showing stable activity even at low pH conditions that inactivated other anti-*H. pylori* compounds.[32] Additionally, an *in vivo* experiment with Mongolian gerbils showed that indolmycin was effective at clearing *H. pylori* infections and repeated exposure of indolmycin to *H. pylori, in vitro,* did not produce any resistant strains.[32] Indolmycin's narrow spectrum of activity makes it an attractive treatment against *H. pylori* because it would be less disruptive to the gut microbiome, potentially avoiding unwanted gastrointestinal side effects of oral use.[32] These attractive qualities lead to indolmycin and other oxazolinone derivatives being patented for use against *H. pylori*.[36,40,41]

Then, indolmycin's activity against *S. aureus* was revisited because of ongoing challenges treating antibiotic resistant infections.[38] Indolmycin was compared to current topical treatments for MRSA, mupirocin and fusidic acid, to which there is growing resistance threatening to disable the use of these treatments. Indolmycin was shown to have bacteriostatic activity against *S. aureus* that was similar to mupirocin and fusidic acid, but indolmycin was also

active against strains that were resistant to mupirocin and fusidic acid. In comparison of the bactericidal activity against *H. pylori*, it has been suggested that there are alternative modes of action for indolmycin in *H. pylori* that are not present in *S. aureus*. Analysis of resistance-causing mutations revealed that they arose less frequently to indolmycin than they did for mupirocin or fusidic acid. Additionally, high-level resistance to indolmycin, caused by an H43N mutation to the *S. aureus* TrpRS, disrupted the bacterial fitness of *S. aureus* with increases to doubling time and decreases to competitive fitness. Indolmycin was proposed as a topical agent against *S. aureus* so that resistance to the antibiotic would not interfere with its potential oral use in the treatment of *H. pylori*.[36,38]

Most recently, in 2016, indolmycin was shown to have bioactivity against *P. falciparum*.[39] *P. falciparum* is a malaria-causing parasite that relies on protein synthesis to facilitate rapid growth. The apicoplast is a prokaryote-resembling organelle within *P. falciparum* where some of the protein translation machinery can be found within the organism. Indolmycin was found to inhibit the apicoplast TrpRS, specifically, causing parasite death in culture, making it potentially useful as a treatment against malaria.[39] Indolmycin was also recently used to understand persistence in *Chlamydia* by mimicking tryptophan starvation.[51] Bacterial persistence complicates treatment of infections and has been implicated in the evolution of antibiotic resistant bacteria,[52] demonstrating that indolmycin could have uses in the development of antibiotics other than as a direct treatment of infections through use as a tool to study persistence and potential treatments against bacteria in this state.

1.2.2 Synthesis

Several attempts have been made to make indolmycin and some of its derivatives since its discovery. Two main strategies to make indolmycin came from either directly linking the oxazolinone ring to the indole[53,54] or by synthesizing an ester of indolmycenic acid, followed by cyclization of the oxazolinone ring (**Figure 1.3**). The latter was the more commonly applied strategy. There were three overall methods developed to make the indolmycenic acid ester. One method involved an epoxide ring opening through use of stannic chloride or with indolyl magnesium bromide.[55–58] The epoxide could be made stereoselectively through lipase-assisted enantioselective deacetylation to obtain a chiral precursor to the epoxide.[57] The second method of obtaining an indolmycenic acid ester was through installation of a chiral amine via a chiral

auxiliary-assisted Grignard reaction followed by oxygenation.[59] The final and most recent method to obtain the indolmycenic acid ester was through a palladium catalyzed coupling of *o*-iodoaniline with an alkyne,[60] analogous to Larock indole synthesis.[61] The alkyne containing the correct stereocenters was obtained by lipase-assisted enantioselective acetylation.[60] From the indolmycenic acid ester, there were three strategies to form the oxazolinone ring and obtain indolmycin. One was through reaction with *N,N′*-dimethylguanidine, but this method resulted in epimerization of the oxazolinone ring.[31,55] The second method was to use *N*-methylthiourea[59] and the third was to use guanidine with a subsequent amine exchange step.[58,60]

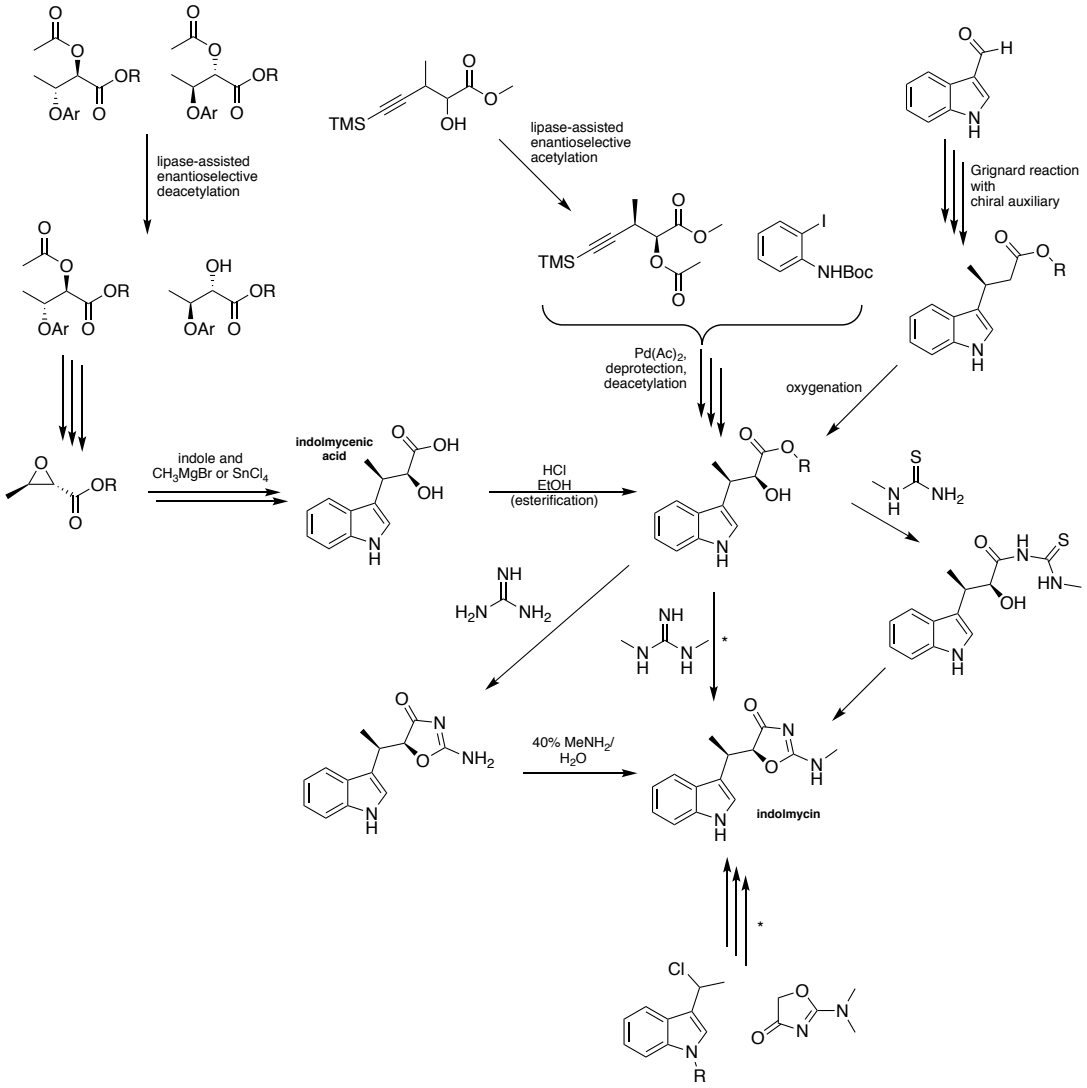


Figure 1.3: Synthetic strategies towards the production of indolmycin.[31,53–60] Strategies that resulted in diastereomers and/or epimers of the oxazolinone ring are indicated with an asterisk.

Several derivatives of indolmycin have been made by synthetic methods (**Figure 1.4**), including oxazolidinedione,[31,58] thiazolinone[62,63] and thiazolidinone[63] analogs, all of which were modifications to the oxazolinone ring. An indolmycin derivative was also made with an additional ring to restrain movement of the molecule in the hopes that it would improve bioactivity.[64] *N*- and *C*-desmethyl derivatives of indolmycin have been made[31] along with some indole substituted analogs.[58,60] Many of the syntheses of substituted indole analogs were focused on making known metabolites found from oral administration of indolmycin to rats, rather than for diversification towards pharmaceutical development.[58] The types and locations of these substituents are also relatively narrow in scope, with attachments all being oxygen-based (hydroxy, methoxy, sulfoxy, and glucuronide) and only made at the 5- and 6-positions of the indole ring.[58,60]

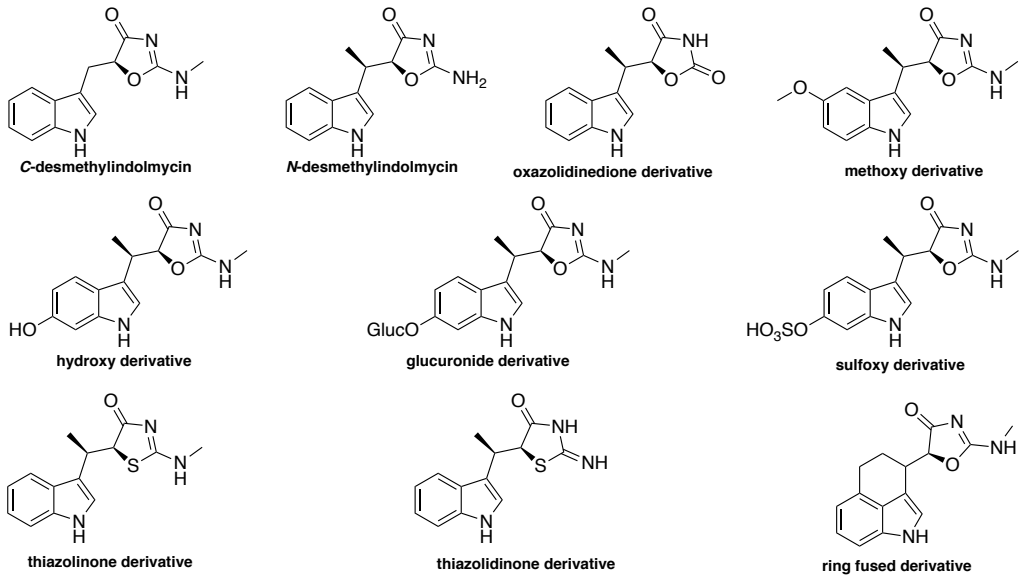


Figure 1.4: Indolmycin derivatives that have been made synthetically.

1.2.3 Biosynthesis

Early work on indolmycin biosynthesis done by Hornemann *et al.* (1971) established that indolmycin was derived from arginine, tryptophan and methionine.[65] Further work partially purifying a tryptophan transaminase and *C*-methyltransferase revealed that these enzymes catalyze some of the initial steps of indolmycin production.[66] An early precursor feeding experiment was also successful at obtaining 5-hydroxy and 5-methoxyindolmycin using the

respective tryptophan and indole derivatives with *S. griseus* ATCC 12648, resulting in improved bioactivity.[67] The indolmycin biosynthetic gene cluster from *S. griseus* ATCC 12648 was previously elucidated by the Ryan group.[48] Putative gene clusters were found by scanning the genome for genes that could potentially encode an indolmycin-resistant TrpRS and the gene cluster was confirmed through deletion mutants and *in vitro* characterization. It was found that indolmycin biosynthesis occurs in three main steps, consisting of biosynthesis of indolmycenic acid from tryptophan, dehydroarginine from arginine, and the oxazolinone ring from indolmycenic acid and dehydroarginine (**Figure 1.5**). This biosynthetic pathway will be discussed in further detail in **Chapter 4**. More recently, another indolmycin biosynthetic gene cluster was discovered from *Pseudoalteromonas luteoviolacea*, which appears to have arisen through convergent evolution.[68] These indolmycin pathways differ by their assembly of the oxazolinone ring, in which the *P. luteoviolacea* biosynthetic method uses an evolved vicinal oxygen chelate superfamily enzyme to intercept a shunt product from the *S. griseus* pathway.

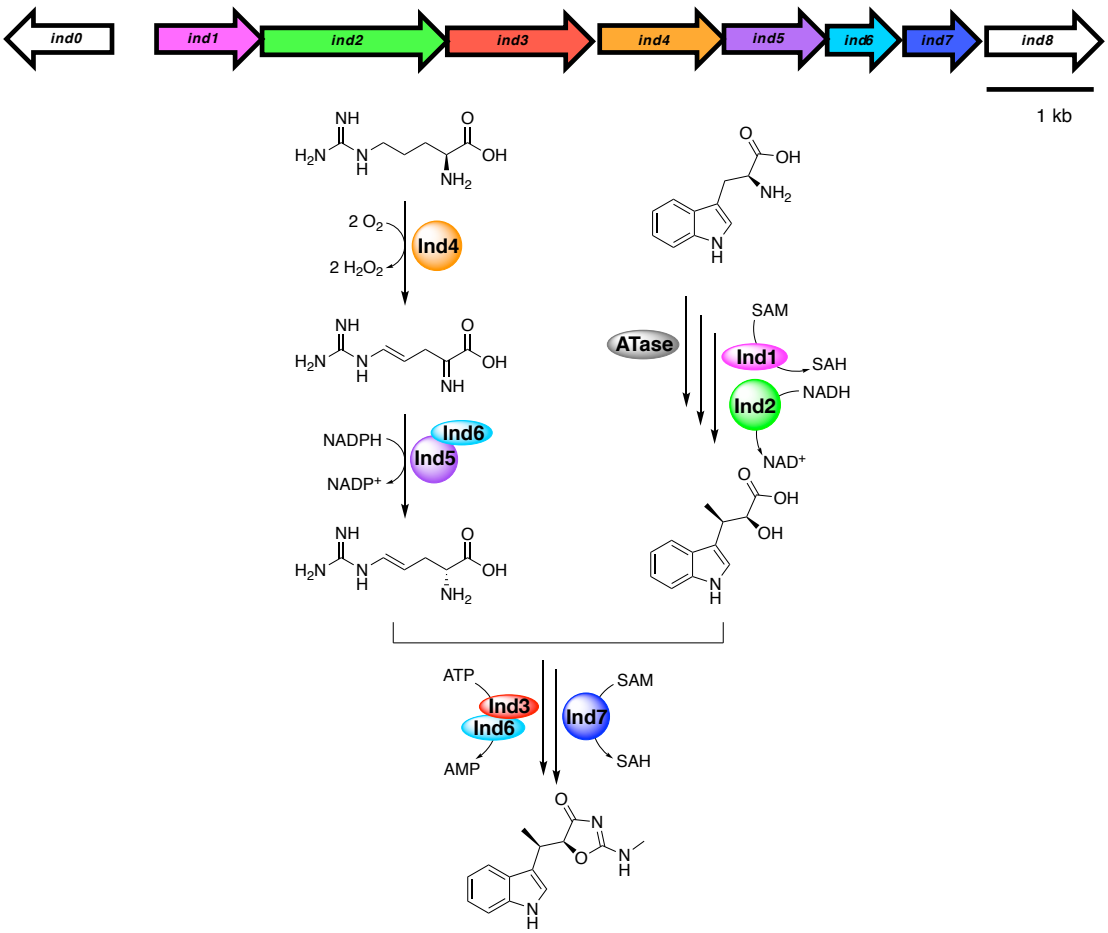


Figure 1.5: The indolmycin biosynthetic pathway. The pathway shows that indolmycin derives from L-arginine and L-tryptophan, which are each tailored by indolmycin biosynthetic enzymes before cyclization and methylation to form indolmycin. Ind4, an O_2- and PLP-dependent enzyme, converts L-arginine to 4,5-dehydro-2-iminoarginine during indolmycin biosynthesis.

1.3 PLP-dependent enzymes

1.3.1 The cofactor and mechanism

The indolmycin biosynthetic pathway includes two different PLP-dependent transformations, one of which, is an unprecedented oxygen-dependent desaturation of arginine catalyzed by Ind4, priming the arginine for reaction with adenylated indolmycenic acid. In order to better understand the importance of this unique enzyme, it is necessary to first understand PLP-dependent enzymes. PLP is a versatile organic cofactor widely employed in metabolism. The cofactor is the active form of vitamin B6, and it enables reactions with amino acid, oxoacid or amine substrates. The structure of PLP dictates several features about how it is used in enzymes. For example, its aldehyde functional group allows it to form a Schiff base with lysine

in the active site of PLP-dependent enzymes, giving the internal aldimine, which is the shared feature of most PLP-dependent enzymes and the starting point for catalysis (**Figure 1.6**). Furthermore, the pyridine ring enables PLP to stabilize carbanions through resonance. Using PLP, enzymes catalyze diverse reactions, ranging from transaminations to Claisen condensations; α-, β-, and γ-eliminations and replacements to transaldolations. The type of reaction catalyzed is dependent on the stereochemistry of the external aldimine intermediate and on which of the three α-carbon groups is aligned perpendicular with the conjugated ring system, as is described in Dunathan's hypothesis.[69] There are five main structural fold-types used to describe PLP-dependent enzymes; Fold-Type I (aspartate aminotransferase family), Fold-Type II (tryptophan synthase family), Fold-Type III (alanine racemase family), Fold-Type IV (D-amino acid aminotransferase family) and Fold-Type V (glycogen phosphorylase family).[70] Some PLP-dependent enzymes have more recently been categorized into two additional fold-types.[71] Despite their family names, each fold-type encompasses enzymes with many different reaction types and functions. PLP-dependent enzymes are estimated to account for ~4% of all enzyme activities, and recent chemo-proteomic strategies have enabled identification of unannotated proteins as PLP-dependent.[72] Recently, some of the most common reactions that use PLP as a key cofactor in natural products biosynthesis were reviewed by the Ryan group, where the diversity of catalysis is on full display.[73] Expansion of this enzymatic diversity through discovery and engineering will help build upon the PLP-dependent enzymes' already impressive utilization as a biocatalyst.

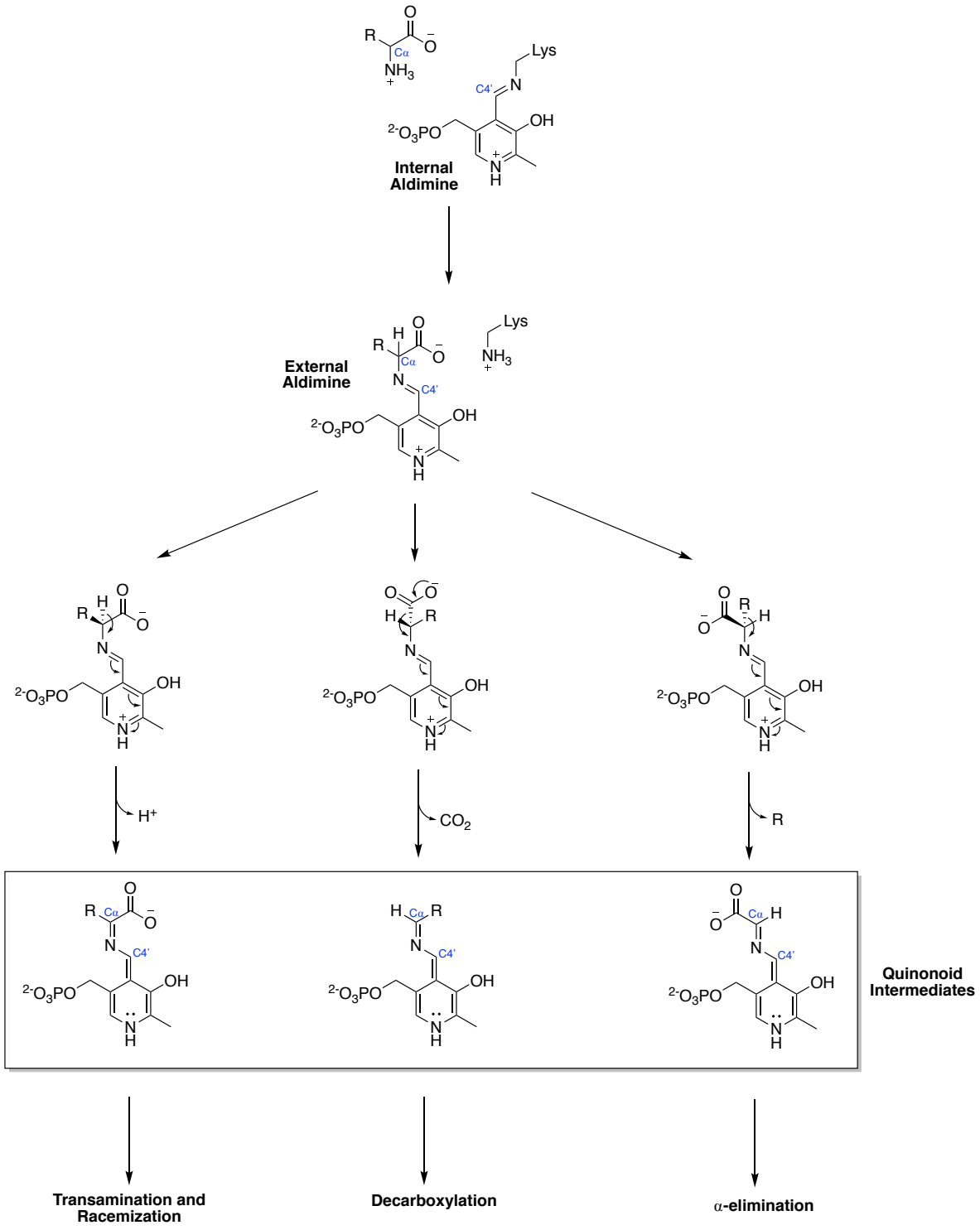


Figure 1.6: Mechanisms of forming common quinonoid intermediates in PLP-dependent enzymes based on Dunathan's hypothesis, in which the leaving group determines the type of quinonoid intermediate formed. Each quinonoid intermediate is associated with certain types of reactions, such as transamination, racemization, decarboxylation and α-elimination.

1.3.2 PLP-dependent enzymes as biocatalysts

The vast reactivity catalyzed by PLP-dependent enzymes in such stereo and regiospecific manners is an attractive quality for industrial processes of altering industrially important molecules, such as amino acids and other related amine-containing molecules.[71] There have been several areas of potential PLP-dependent enzyme use previously outlined, including the chemical, pharmaceutical and food industries. Transaminases in particular have attracted the attention of organic chemists for their ability to produce chiral amines.[74,75] Two notable examples include the use of an engineered transaminase in the production of an antidiabetic drug, sitagliptin,[76] and an anxiety drug candidate precursor, imagabalin,[77] with the former resulting in drastic improvements to yield, productivity, waste, heavy metal use and manufacturing cost.[71] Other types of PLP-enzymes that may become industrially important include lysine decarboxylases, threonine aldolase, tyrosine phenyl-lyase, amino-caprolactam racemases, cystathionine β-lyase.[71] Although these enzymes are all attractive for use in industry, more research and information is needed on these enzymes to make them amenable for industrial use. In order to be employed as a biocatalyst, the mechanistic details of each individual enzyme, even subtle differences, must be properly studied in order to fully understand the factors affecting catalysis.[71] Going forward, the use of PLP-dependent enzymes for industrial processes will require discovery of new enzymes, mechanistic studies of current enzymes and engineering efforts to adapt these enzymes to suit the needs of industry.

1.3.2.1 Tryptophan synthase (TrpS)

Although not yet used industrially, tryptophan synthase (TrpS) is a promising biocatalyst for the production of tryptophan derivatives. TrpS is a highly studied enzyme in the tryptophan biosynthesis pathway, converting indole-3-glycerol phosphate into tryptophan (**Figure 1.7**). The enzyme is heterodimeric with two dissimilar active sites, one per subunit.[78,79] The α-subunit (TrpA) catalyzes a retro-aldol cleavage of indole-3-glycerol phosphate into indole and glyceraldehyde-3-phosphate (G3P). The β-subunit (TrpB) catalyzes a PLP-dependent condensation of indole and L-serine to form L-tryptophan via a β-replacement reaction. TrpS is the first enzyme that was found to allow internal passage of an intermediate between different active sites via a tunnel. While this indicates that indole is not released after the reaction in TrpA, it is possible to produce tryptophan by feeding indole and serine, thereby bypassing the activity

of TrpA and only employing TrpB.[79] However, the overall activity of TrpB is reliant on the presence of TrpA, indicating some allosteric effects on regulation.[80] Engineering efforts have been made to TrpB to negate any allosteric effects and allow it to act as a standalone enzyme.[81–84] These efforts followed the discovery that TrpS, specifically the activity of TrpB, could not only produce tryptophan but a suite of tryptophan analogs when fed indole derivatives, making it an attractive biocatalyst.[85–89] Substantial efforts to expand the substrate scope have been made to increase the utility of TrpS as a biocatalyst, including improved acceptance of indole-substituted and non-indole analogs, as well as L-serine derivatives, such as L-theronine.[79] Recognition of the biocatalytic potential of TrpS has inspired biosynthetic researchers to include TrpS in several biocatalytic cascades to produce challenging molecules, such as D-tryptophan,[90,91] tryptamine,[92,93] psilocybin[93] and thaxtomin[94] derivatives.

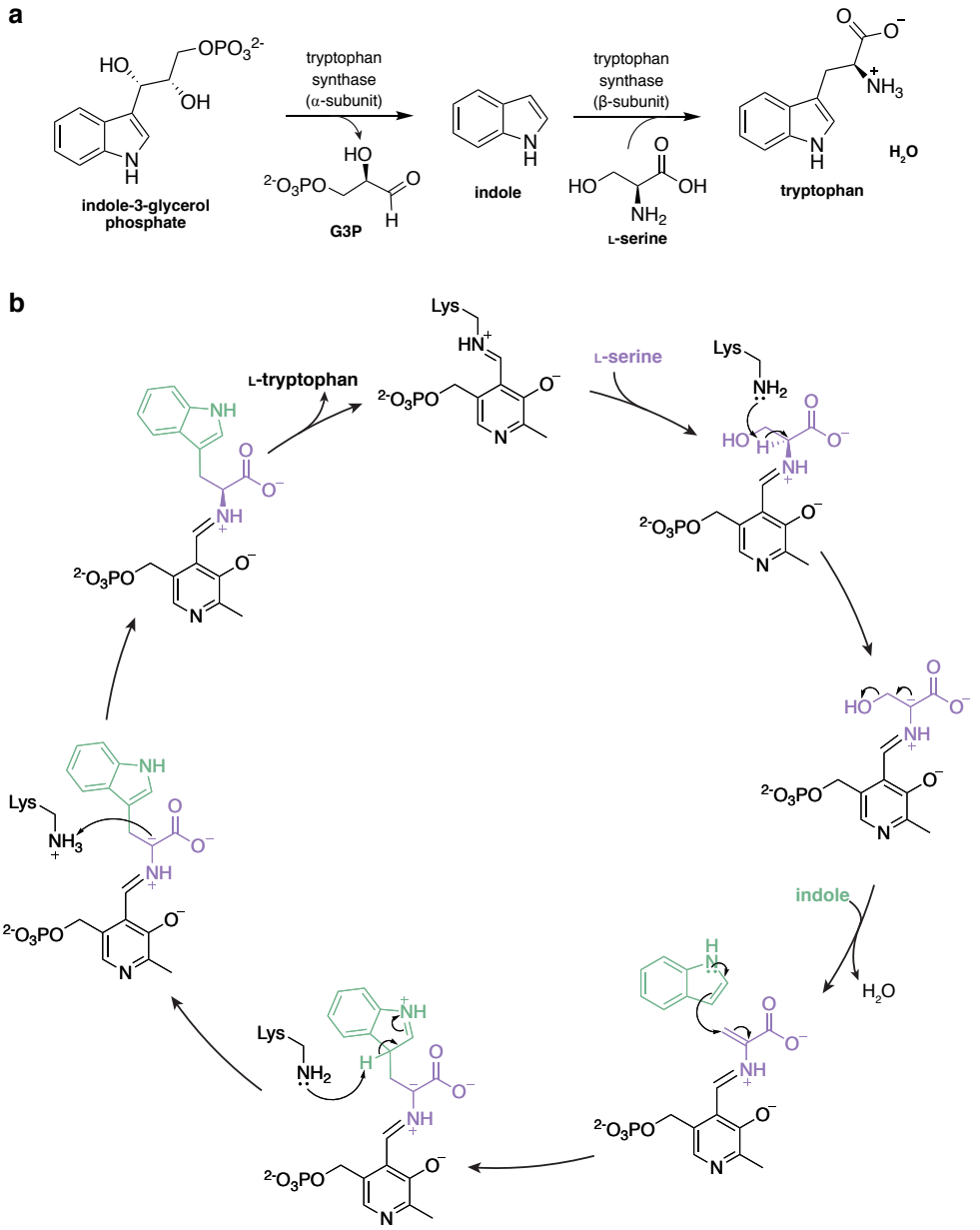


Figure 1.7: TrpS catalyzes the production of tryptophan. (**a**) Overall reaction scheme for both the α-subunit and β-subunit in TrpS. (**b**) Catalytic cycle of TrpB producing L-tryptophan from L-serine and indole.

1.4 Oxygen- and PLP-dependent enzymes

As was previously discussed, a key necessity for more widespread use of biocatalysts in industrial processes is the expansion of the known biocatalytic repertoire. While PLP-dependent enzymes are already known to catalyze extremely diverse types of reactions, one exciting new reaction type being investigated amongst PLP-dependent enzymes is oxygen utilization. Despite

the reaction diversity amongst PLP-dependent enzymes, what was not originally expected is that these enzymes would use oxygen as a co-substrate during catalysis. Oxygen is more stable in the triplet state, necessitating a cofactor that can facilitate single-electron transfers. For this reason, it is more expected that enzymes that use oxygen would employ transition metal-containing cofactors and organic cofactors, like flavin, capable of such single electron transfers. Nonetheless, in pioneering papers, Schloss and Abell anticipated that any enzymatically generated carbanionic intermediates could react with electrophiles, including oxygen;[95] a phenomenon often referred to as paracatalysis.[96–98] They explored this hypothesis to identify a number of enzymes that react paracatalytically with oxygen, beautifully demonstrating that some (but not all) enzymes that stabilize carbanions can also react with oxygen.

In recent years, the possibility that PLP-dependent enzymes can react with oxygen has been extended through discoveries in biosynthesis, highlighting enzymes that do not react paracatalytically with oxygen, but instead use oxygen as a dedicated co-substrate for the primary reaction of the enzyme. Described here, is the development of the hypothesis that PLP-dependent enzymes can react with oxygen and the experimental exploration of this hypothesis. Then recent work exploring the emergence of enzymes that use O_2 as a co-substrate is described and the biological contexts of these enzymes are highlighted. It is interesting to note that O_2-, PLP-dependent reactions have now been described in diverse protein families (**Table 1.1**), suggesting that the evolution of reactions with O_2 may have emerged multiple times. This observation suggests that PLP-dependent enzymes that do not yet react with oxygen might also be evolved in the laboratory to use oxygen as a co-substrate.

Table 1.1: PLP-enzymes with oxygen-dependent or paracatalytic reactions. Table was adapted from Hoffarth *et al*. (2020).

Enzyme	Protein Family*	Predicted Function*	Reaction Catalyzed	References
DDC (paracatalytic)	PF00282	DOPA decarboxylase	Decarboxylation/Oxidative Deamination/Abortive Transamination	Bertoldi *et al.* (1998)[99]
GAD (paracatalytic)	PF00282	Glutamate Decarboxylase	Decarboxylation/Oxidative Deamination/Abortive Transamination	Bertoldi *et al.* (1999)[100]
AAS Plants and Fungi	PF00282	Tyrosine/DOPA/Serine Decarboxylase	Oxidative Deamination	Kaminaga *et al.* (2006),[101] Sakai *et al.* (2007),[102] Torrens-Spence *et al.* (2012),[103] Torrens-Spence *et al.* (2018),[104–106] Gutensohn *et al.* (2011)[107] and Torrens-Spence *et al.* (2014)[108]
AAS Insects	PF00282	Tyrosine/DOPA Decarboxylase	Oxidative Deamination	Vavricka *et al.* (2011)[109]
ODC (paracatalytic)	PF02784	Ornithine Decarboxylase	Decarboxylation/Oxidative Deamination/Abortive Transamination	Bertoldi *et al.* (1999),[100] Sakai *et al.* (1997)[110] and Pontrelli *et al.* (2018)[111]
CcbF	PF00155	Aspartate Amino Transferase	Oxidative Deamination	Wang *et al.* (2016)[112]
Ind4	PF00155	Aspartate Amino Transferase	Arginine Desaturation	Du *et al.* (2016)[113]
MppP	PF00155	Aspartate Amino Transferase	Arginine Hydroxylation	Han *et al.* (2015)[114] and Han *et al.* (2018)[115]
RohP	PF00155	Aspartate Amino Transferase	Arginine Hydroxylation	Hedges *et al.* (2018)[116]
CuaB	PF00155	8-amino-7-oxononanoate synthase	Decarboxylating Oxygenation	Dai *et al.* (2020)[117]
PvdN	PF00266	Cysteine Desulferase, CsdA	Decarboxylating Oxygenation	Ringel *et al.* (2016)[118]
Cap15	PF03841	Selenocysteine Synthase, SelA	Decarboxylating Oxygenation	Huang *et al.* (2018)[119]

*Protein Family and Predicted Function were determined from a combination of the cited literature and BLAST search predictions of each enzyme.

1.4.1 Understanding reactivity of PLP-dependent enzymes with O_2

The chemical reasons that PLP-dependent enzymes use oxygen in reactions are not yet fully understood. Oxygen, which normally exists in the triplet state, is fairly unreactive towards organic cofactors, typically found in a singlet state, as this is a spin-forbidden process. Therefore, enzymes catalyzing such reactions must be able to overcome this challenge. Flavin-dependent enzymes are a ubiquitous group of enzymes that commonly react with oxygen and are often used as a model to understand enzymatic oxygen reactivity in the absence of a metal cofactor. In these reactions, the need for intermediate spin inversion is overcome by an initial single electron transfer from reduced flavin to the triplet state oxygen, generating superoxide and a flavin semi-quinone as a caged radical pair.[120,121] This process can be understood through Marcus theory, in which electron transfers are governed by the redox potentials of the reactants and the reorganization energy,[122] the latter of which is thought to be the more important contributor to flavin-oxygen reactivity in flavo-proteins.[120] Active site polarity, confined oxygen binding sites and flavin stereochemistry are specific contributors to lowering the reorganization energy in flavo-proteins resulting in enhanced oxygen reactivity.[120,121] Studies of cofactor-independent oxygenases, which employ carbanions, used flavo-proteins as a model to understand oxygen reactivity. Experimental observation of radical intermediates by electron paramagnetic resonance (EPR) in a cofactor-independent dioxygenase suggested that the chemistry of cofactor-independent oxygenases likely traverses through carbanionic donation of a single electron to dioxygen.[123]

It is conceivable that something similar to the flavo-protein mechanism of oxygen activation could be occurring within some PLP-dependent decarboxylases (**Figure 1.8**). Most mechanistic proposals[96,124–127] describe a streamlined process in which the quinonoid intermediate donates a single electron to oxygen, forming superoxide and a semiquinone as a caged radical pair. Then, the radical pair can collapse to form a hydroperoxyl-intermediate localized at either the C4′ of PLP (**Hydroperoxy I**) or the α-carbon of the substrate (**Hydroperoxy II**). Eliminating the hydroperoxyl-intermediate forms hydrogen peroxide and an imine product, which hydrolyzes in water to release ammonia and a carbonyl-containing molecule. Flavoprotein investigations also suggest that formation of the hydroperoxyl-intermediate may vary between enzymes, which could also be the case for the PLP-dependent enzymes.[120,121]

Although it is predicted that the PLP-dependent oxygen reactivity will likely proceed through a single electron transfer, it is theoretically possible to activate a carbanionic intermediate by excitation to the triplet state. In the triplet state, the intermediate reaction with oxygen would no longer be spin forbidden. Those studying oxygen activation by cofactor-independent enzymes also discussed a similar possibility prior to obtaining experimental support for a single electron transfer process.[123,128] At this time neither hypotheses can be disregarded for the PLP-dependent oxygen reactivity due to a lack of evidence. It is hoped that this discussion will inspire others to further investigate the oxygen reactivity both experimentally and computationally.

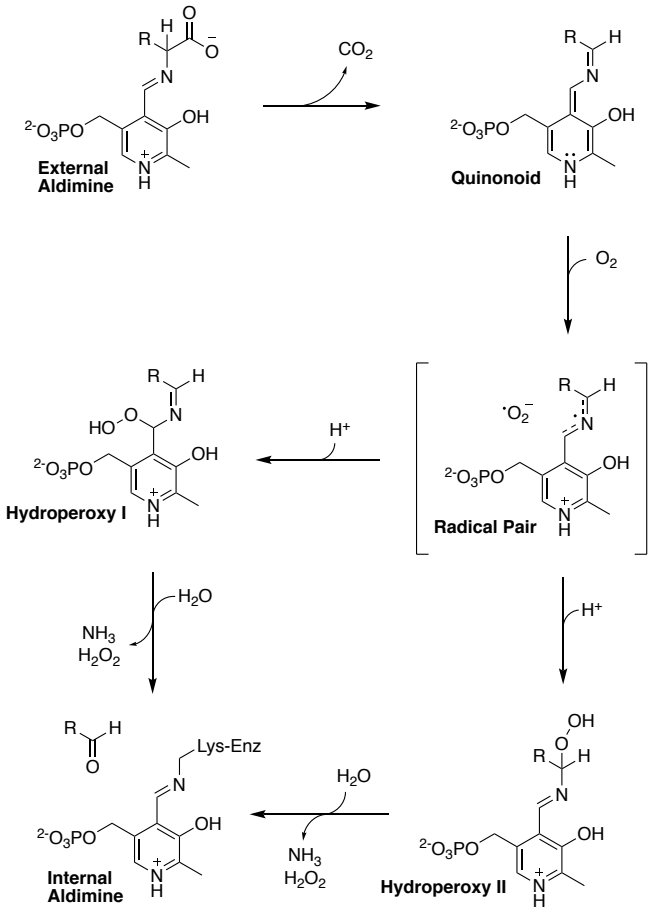


Figure 1.8: Proposed mechanism for paracatalytic oxidative deamination from PLP-dependent decarboxylases.

1.4.2 Paracatalytic reactions of PLP-dependent enzymes with O_2

The term 'paracatalytic' describes the use of a non-physiological oxidizing reagent in an enzymatic reaction,[96–98] a term that also encompasses oxidative side-reactions with O_2.[129] The first enzyme found to undergo a paracatalytic reaction with O_2 itself was Rubisco, which produces two molecules of 3-phosphoglycerate from ribulose-1,5-bisphosphate, but will also use O_2, paracatalytically, to make one molecule of 3-phosphoglycerate and one molecule of 2-phosphoglycerate.[130,131] Since the discovery of Rubisco's paracatalytic activity, several other enzymes were also found to have paracatalytic reactions with oxygen, including glutamate decarboxylase, which is the first PLP-dependent enzyme discovered to react with oxygen.[95] Until recently, knowledge of PLP-dependent reactions with oxygen were limited to paracatalytic activities. Paracatalytic reactions with oxygen, including and extending beyond the PLP-dependent examples, have been reviewed previously.[96]

There are three types of known paracatalytic PLP-dependent enzymes that react with oxygen: 3,4-dihydroxyphenylalanine (DOPA) decarboxylase, glutamate decarboxylase (GAD) and ornithine decarboxylase (ODC). DOPA decarboxylase (DDC) catalyzes the conversion of aromatic L-amino acids L-DOPA and L-hydroxytryptophan (L-HTP) into the corresponding amines via decarboxylation (**Figure 1.9A** and **Figure 1.10A**). Additionally, DDC has two additional reactions: abortive transamination and oxidative deamination (**Figure 1.9A** and **B**).[99] Abortive transamination proceeds through protonation of the quinonoid at the C4′ position, giving the ketimine, which is subject to hydrolysis, releasing the corresponding aldehydes or ketones and pyridoxamine 5′-phosphate (PMP) (**Figure 1.9B**). Oxidative deamination by DDC also gives the corresponding aldehydes or ketones but through paracatalytic reactions with O_2, regenerating PLP instead of forming PMP (**Figure 1.9C**).[132] Similarly, GAD from *E. coli* and ODC from *Hafnia alvei* and Lactobacillus 30a also exhibit abortive transamination and oxidative deamination in addition to their typical decarboxylation reactions (**Figure 1.10B** and **C**).[95,100,110]

Interestingly, it has been shown that the reactivity of some of these decarboxylases can be modulated to favour oxidative deamination by mutating certain residues. The pig kidney DDC Y332F variant exclusively catalyzes oxidative deamination under aerobic conditions and abortive deamination under anaerobic conditions, with abolished decarboxylation activity.[126] The predicted function of Tyr332 is as a proton donor for the substrate α-carbon following decarboxylation. Such proton donation which would no longer be possible in the Y332F variant,

thus leading to oxidative deamination.[126] This claim is supported by a crystal structure for human GAD67 which showed that Tyr434 (equivalent to DDC Tyr332) is positioned to protonate the α-carbon with the help of a nearby histidine in Chain A, supporting that the tyrosine residue could be responsible for modulating oxygen reactivity through α-carbon protonation (**Figure 1.11A**).[133] However, Chain B shows Tyr434 interacting with the backbone amine of Tyr292, while His291 is closer to the α-carbon of the substrate, suggesting that His291 could act as the proton donor (**Figure 1.11B**).[133] More recently, ODC from *E. coli*, SpeC, was found to accumulate mutations, G655A and G655S, that converted the decarboxylase into a dedicated oxidative transaminase in response to an L-alanine auxotrophic mutation, ΔpanD.[111] Wild-type SpeC also demonstrated some oxidative deamination activity, additional to the expected decarboxylase activity, but an explanation for why these mutations might cause the change in reactivity has not yet been proposed.

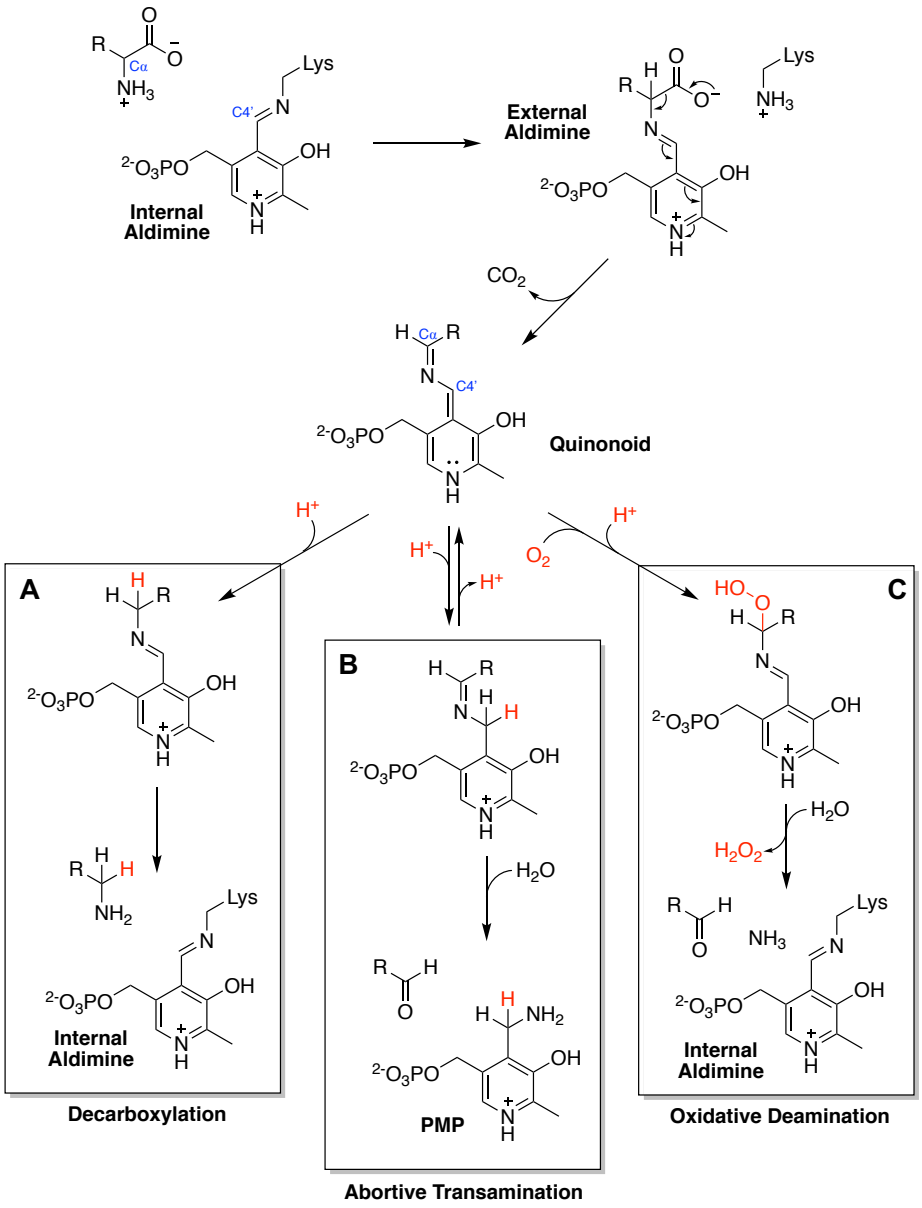


Figure 1.9: Mechanisms of decarboxylation, abortive transamination and oxidative deamination in the paracatalytic decarboxylases. All mechanisms traverse through a decarboxylated quinonoid intermediate and are differentiated by the site of reprotonation or by reaction with oxygen.

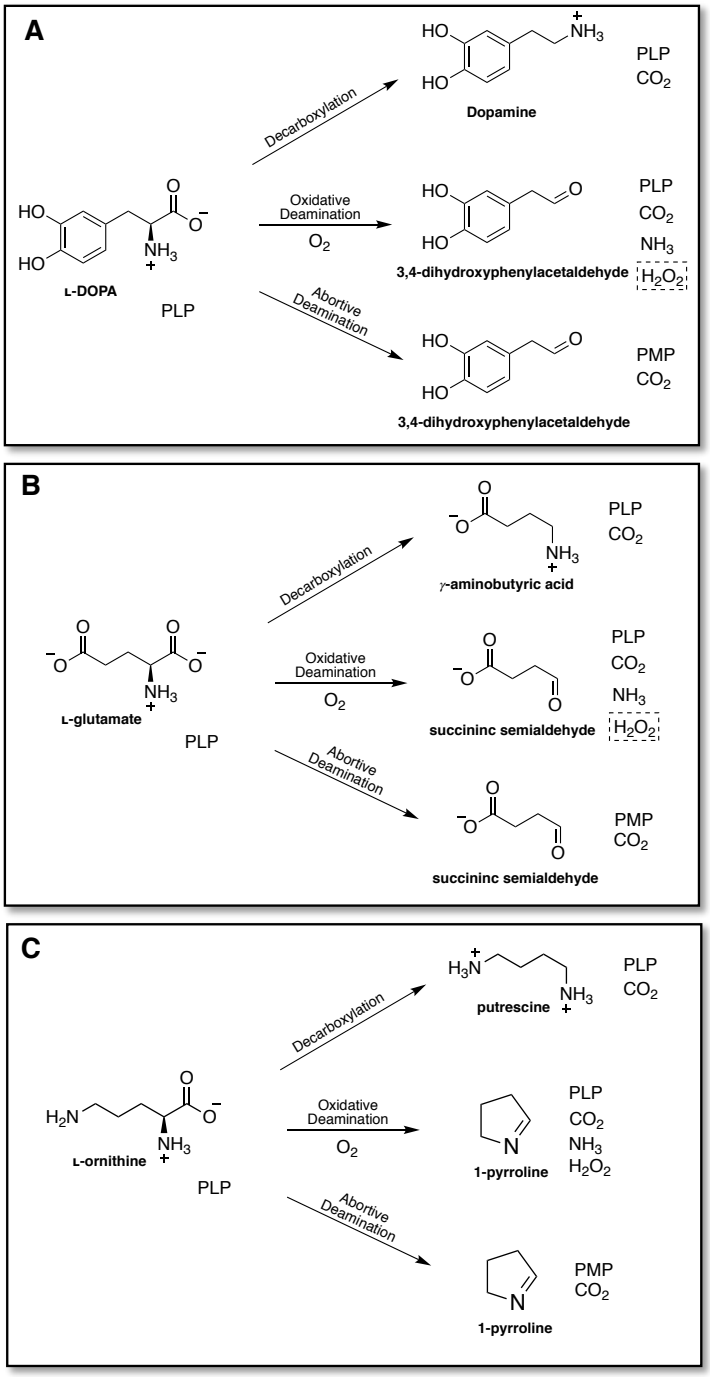


Figure 1.10: Overall schemes for the three types of PLP-dependent decarboxylases with paracatalytic reactivity. (**A**) DOPA decarboxylase (DDC). (**B**) Glutamate decarboxylase (GAD). (**C**) Ornithine decarboxylase (ODC). Each scheme shows the substrate and products for decarboxylation, oxidative deamination, and abortive deamination. Predicted, but unconfirmed species are indicated by a dashed box.

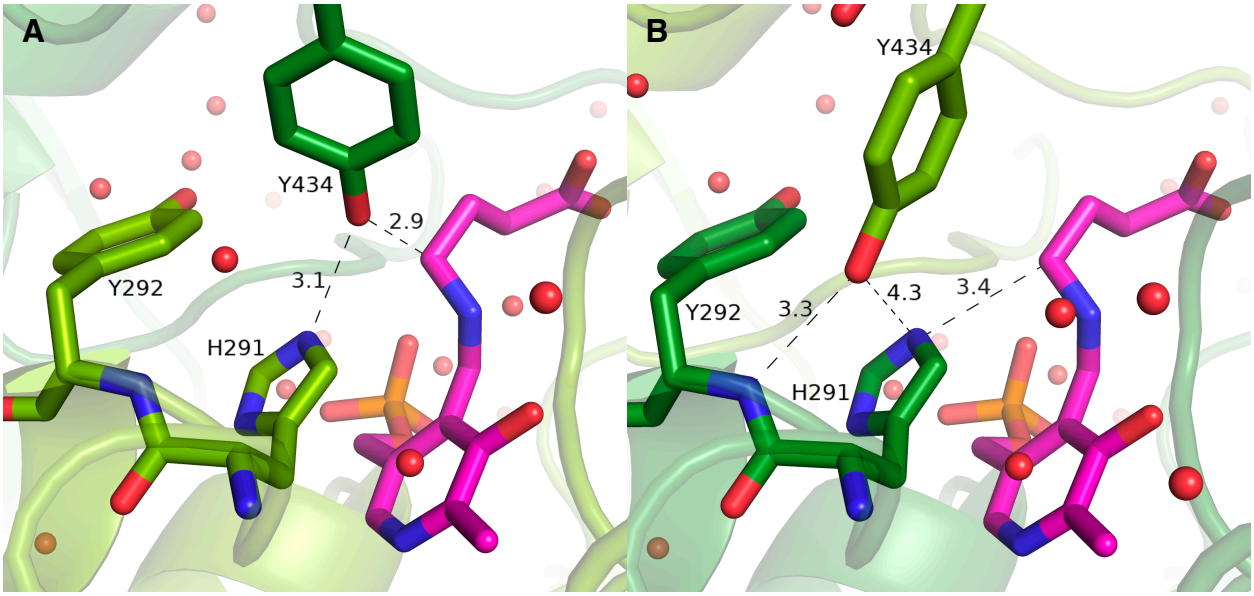


Figure 1.11: X-ray crystal structure of human GAD67 (PDB code: 2OKJ). Each active site, the A-chain (**A**) or the B-chain (**B**), from the dimer is shown. The A-chain is shown in light green while the B-chain is shown in dark green. A trapped intermediate (PLP-GABA) from the B-chain active site is coloured magenta. PLP-GABA from Chain B was superposed into Chain A for easier comparison of the structures. Distances between atoms are shown by dashed lines and are given in Å. Water molecules are shown as red spheres.

1.4.3 Oxidative deamination

1.4.3.1 Plant amino acid aldehyde synthases

The first PLP-dependent enzyme discovered to have a dedicated oxygen reactivity was the phenylacetaldehyde synthase (PAAS) from Petunia hybrida cv. Mitchell.[101] Phenylacetaldehyde (PHA) is an important scent compound, giving fragrance to flowers, fruits and some processed foods, such as bread and cheese.[134,135] Kaminaga *et al.* found that the petunia PAAS produces stoichiometric amounts of carbon dioxide, ammonia, hydrogen peroxide and PHA, paired with oxygen consumption (**Figure 1.12A**) and that decarboxylation precedes deamination in the mechanism.[101] Similarly, the rose PAAS produces PHA, ammonia and hydrogen peroxide in nearly stoichiometric amounts under aerobic conditions.[102] Since the discovery of the rose and petunia PAASs, more examples of amino acid aldehyde synthases (AASs) with activity on various substrates have been discovered from plants, including from parsley (*Petroselinum crispum*),[103] *Rhodiola rose*,[104] *Arabidopsis thaliana,*[107] *Cicer arietinum* (chickpeas),[108] *Medicago truncatula*[108] and *Eucalyptus grandis*[106] (**Table 1.1**). The first known fungal AAS example, 3,4-dihydroxyphenylacetaldehyde (DHPAA) synthase, has also been discovered from *Psilocybe cubensis*.[105]

A study comparing *Arabidopsis thaliana* and parsley AASs with similar amino acid decarboxylases (AADCs) in *Catharanthus roseus* and *Papaver somniferum*, which do not catalyze oxidative deaminations, aimed to understand the distinguishing features between the two groups of enzymes.[136] An active site tyrosine residue is conserved in the decarboxylases, while the corresponding residue is a phenylalanine or valine in the AASs (**Figure 1.13A**). Mutating the active site phenylalanine to a tyrosine in the AAS resulted in AADC activity, and vice versa, demonstrating the importance of this residue on distinguishing the reactivity. This result is consistent with work on the paracatalytic DDC Y344F variant.[126] The results also offered a new hypothesis for the role of the tyrosine, suggesting that it might stabilize ionization of a neighbouring, conserved histidine residue while the histidine itself protonates the α-carbon.[136] The AAS from *Eucalyptus grandis* contains a histidine to asparagine substitution, which supports the importance of this conserved histidine, but molecular dynamics simulations in this study suggested the tyrosine as the proton donor.[106]

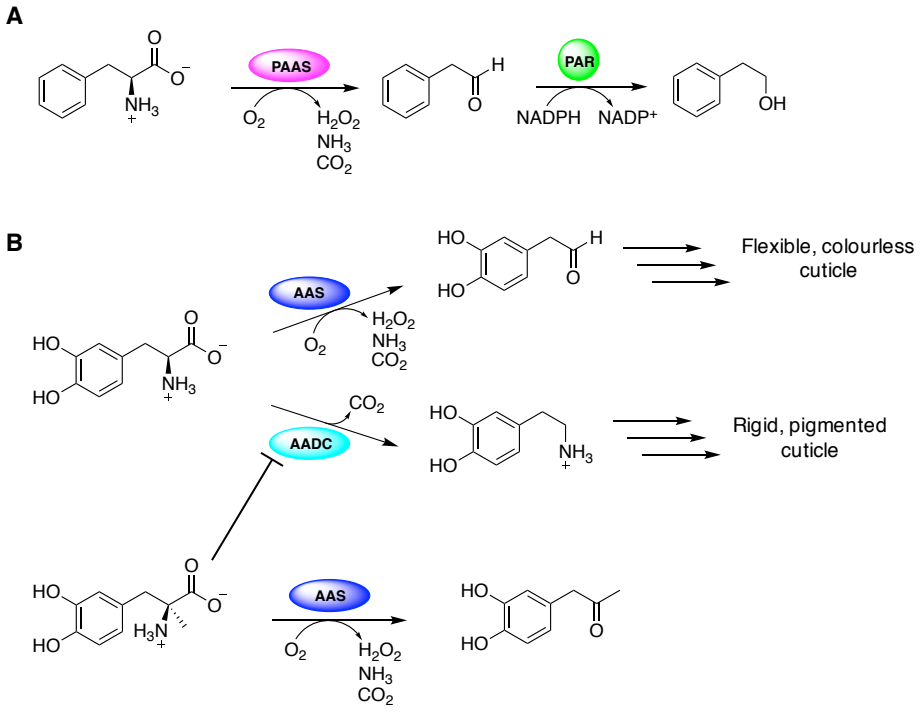


Figure 1.12: Schemes for physiological roles of major AAS enzymes are shown. (**A**) Plant PAAS involvement in producing 2-phenylethanol from L-phenylalanine. (**B**) Insect AAS involvement proposed in the formation and flexible, colourless cuticle from L-DOPA. The physiological connection between effects of α-methylDOPA feeding and observed AAS activity is also described.

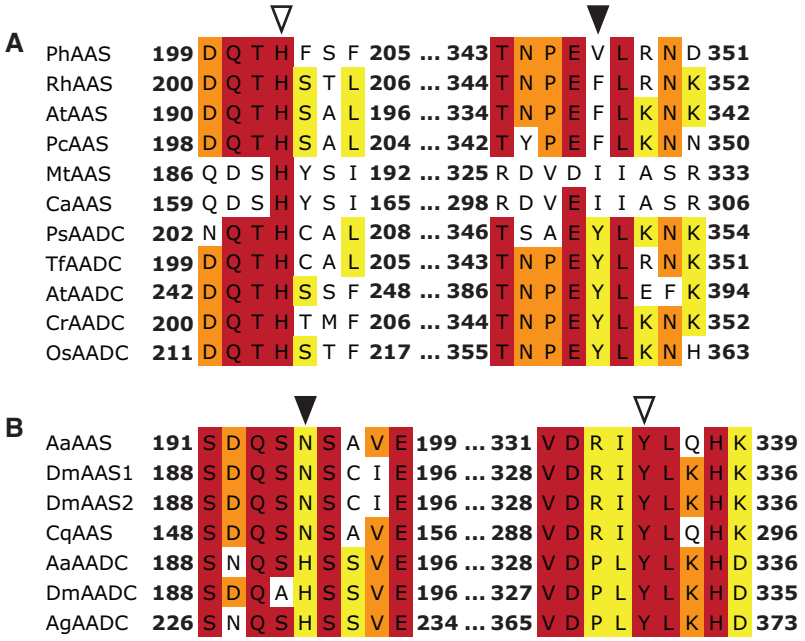


Figure 1.13: Partial sequence alignments comparisons of AASs and AADCs in (**A**) plants and (**B**) insects. Alignments were made with Clustal Omega in UGENE.[175] Arrows indicate the proposed catalytic residues responsible for protonating the α-carbon in the decarboxylases in both the plants and insects. A filled in arrow indicates the key residue responsible for modulating oxidative deaminase activity. (**A**) Plant sequences [*PhAAS*: *Petunia hybrida* cv. Mitchell (ABB72475.1), *RhAAS*: *Rosa hybrid cultivar* (ABB04522.1), *AtAAS*: *Arabidopsis thaliana* (NP_849999.1), *PcAAS*: *Petroselinum crispum* (Q06086.1), *MtAAS*: *Medicago truncatula* (XP_003592128.2), *CaAAS*: *Cicer arietinum* (XP_004496485.1), *PsAADC: Papaver somniferum* (AAC61842.1), *TfAADC*: *Thalictrum flavum* subsp. Glaucum (AAG60665.1), *AtAAS*: *Arabidopsis thaliana* (NP_001078461.1), *CrAADC*: *Catharanthus roseus* (P17770.1), *OsAADC*: *Oryza sativa* (BAG91223.1)] are aligned to show catalytic residues that distinguish oxidative from non-oxidative reactivity. (**B**) Insects sequences [*AaAAS: Aedes aegypti* (EAT37247.1), *DmAAS1*: *Drosophila melanogaster* (NP_476592.1), *DmAAS2*: *Drosophila melanogaster* (NP_724162.1), *AgAAS: Anopheles gambiae* (XP_319838.3), *CgAAS*: *Culex quinquefasciatus* (EDS39158.1), *AaAADC*: *Aedes aegypti* (XP_001648264.1), *DmAADC*: *Drosophila melanogaster* (NP_724164.1), *AgAADC*: *Anopheles gambiae* (AAC16249.1)] are aligned to show catalytic residues that distinguish oxidative from non-oxidative reactivity. Figure was adapted from Hoffarth *et al*. (2020).

1.4.3.2 Insect amino acid aldehyde synthases

Insects have several annotated AADCs in their genomes, however one enzyme, DHPAA synthase (DHPAAS), catalyzes an oxidative deamination of L-DOPA, as demonstrated in *Aedes aegypti*, *Drosophila melanogaster*, *Anopheles gambiae*, and *Culex quinquefasciatus*.[109] The product of this enzyme, DHPAA, is hypothesized to be important in uncoloured, flexible cuticle

formation in insects, perhaps through aiding crosslinking of cuticular proteins (**Figure 1.12B**).[109,137] The gene responsible for producing DHPAAS in *Drosophila* had been previously designated the *α-methyldopa hypersensitive* (*amd*) gene, based on observed sensitivity to α-methylDOPA (AMD) when the gene was mutated.[138–141] The mutants have defects in cuticle development, but the function of the *amd* gene was unclear at the time. The insect DHPAAS was able to convert AMD to 3,4-dihydroxyphenylacetone, demonstrating a link between the observed phenotype and the enzymatic function (**Figure 1.12B**).[109]

Liang *et al*. produced homology models to compare an AADC with an AAS from *Drosophila* and observed that active site His192 was associated with decarboxylation, whereas Asn192 in the corresponding position resulted in oxidative deamination (**Figure 1.13B**).[142] The importance of this residue was confirmed by mutating the His in AADC to Asn, and the Asn in AAS to His, resulting in loss of oxidative deamination in the Asn to His mutant and gain of oxidative deamination in the His to Asn mutant. The authors propose that the His192 is able to protonate the α-carbon to facilitate decarboxylation, while the asparagine residue cannot. Another study aimed at optimizing the decarboxylation and oxidative deamination reactivity of insect AAS for improved biosynthesis of benzylisoquinonline alkaloids supported the key role of Asn192, but also found that Phe79 and Tyr80 are important for modulating this activity.[143] Although, with the exception of *Eucalyptus grandis*,[106] the key residue implicated with distinguishing AADCs from AASs in insects is not the same as in plants, but these residues are not unrelated because the catalytic His192 residue in the insect AADC corresponds to the same active site histidine residue described in **Section 1.4.3.1** for plants (**Figure 1.13**). Therefore, it seems that each species took a different evolutionary pathway by mutating a different residue to converge on the same outcome. Any disruption to the catalytic histidine or tyrosine residues seems to result in oxidative deamination, exclusively, or bifunctionally with decarboxylation.[106]

1.4.3.3 Lincosamide biosynthetic enzyme CcbF

Lincosamide antibiotics inhibit protein synthesis in gram-negative cocci and in gram-positive cocci and bacilli.[144] Structurally, they are characterized by an eight-carbon amino sugar with a sulfur appendage linked to an amino acid moiety by an amide bond. Lincomycin and celesticetin are lincosamide antibiotics from *Streptomyces lincolnensis* and *Streptomyces caelestis*, respectively, that differ in regard to their S-substitution: where lincomycin contains a

methylsulfhydyl moiety, and celestietin contains a salicylate moiety attached to sulfur via a 2-carbon chain (**Figure 1.14**). The biosynthetic gene clusters differ by a pair of homologous PLP-dependent enzymes, LmbF and CcbF (**Figure 1.14A**).[112,145] The lincomycin biosynthetic enzyme, LmbF, is responsible for C-S bond cleavage, removing most of the S-cysteine residue through a PLP-dependent β-elimination reaction, while the celesticetin biosynthesis enzyme CcbF was predicted to process the S-cysteine residue, through decarboxylation-dependent transamination.[145] In an effort to confirm the proposed decarboxylation-dependent transamination reaction, Ushimaru and colleagues further investigated CcbF, and confirmed formation of the aldehyde product, but could not detect formation of PMP or use of a keto-acid co-substrate to regenerate PLP, indicating that catalysis proceeds through another mechanism.[146] Wang and colleagues found that oxygen is essential for CcbF enzymatic activity, and that hydrogen peroxide, ammonia and carbon dioxide are generated along with the product.[112] These findings led the authors to propose a mechanism in which decarboxylation of the external aldimine intermediate results in a quinonoid intermediate that is susceptible to react with oxygen, producing hydrogen peroxide followed by imine hydrolysis of the product (**Figure 1.14B**).

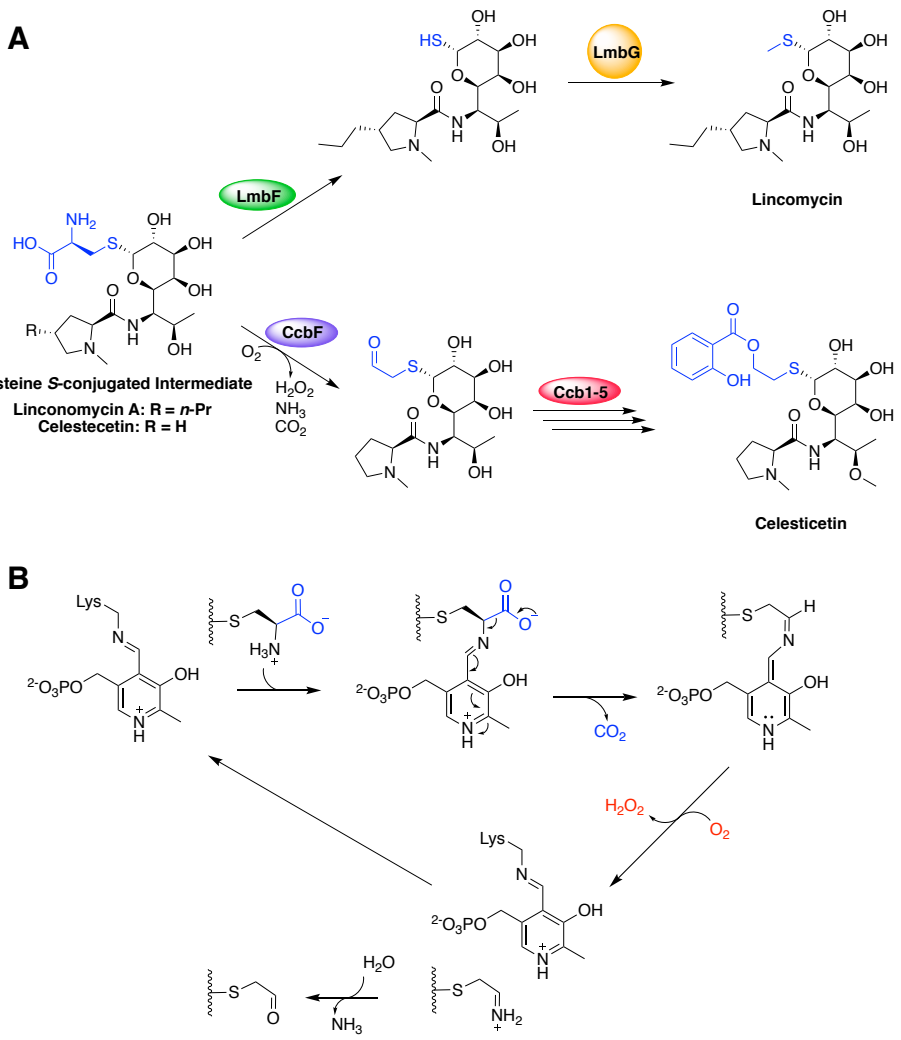


Figure 1.14: Biosynthesis of lincomycins and celesticetins. (**A**) Branching point catalyzed by LmbF and CcbF in lincomycin and celesticetin biosynthesis, respectively. (**B**) Proposed mechanism for CcbF.

1.4.4 Decarboxylating oxygenation

1.4.4.1 PvdN

Pyoverdines are molecules composed of a 2,3-diamino-6,7-dihydroxyquinoline linked to a peptide moiety which play a role in virulence and toxicity in pathogenic *Pseudomonas* species.[147–149] Of the five enzymes involved in the biosynthetic pathway to pyoverdine (PvdM, PvdN, PvdO, PvdP, and PvdQ), only the roles of PvdP and PvdQ were known until recently.[150–152] Ringel and colleagues investigated the catalytic role of PvdN, demonstrating that it converts a glutamic acid moiety at the 3-amino group of the pyoverdine chromophore to a succinamide

group via a *ΔpvdN* mutation and *pvdN* complementation (**Figure 1.15A** and **Figure 1.15B**).[118] The authors postulated that the conversion to succinamide could occur via an O_2-, PLP-dependent oxygenation-decarboxylation cascade, in which PvdN forms a Schiff base between the PLP-cofactor and the α-amino nitrogen of the glutamic acid moiety (**Figure 1.15C**). Decarboxylation forms a quinonoid intermediate, which reacts with a molecule of oxygen, forming a peroxy-adduct at the glutamic acid α-carbon. As the peroxy-adduct intermediate degrades through breakage of the oxygen-oxygen bond, the amide group of the succinamide moiety observed in pyoverdine is formed. The structure of PvdN supports this proposal through the existence of two tunnels connecting the enzyme's surface to active site, one of which is large enough for the substrate, while the other is narrower to accommodate the exchange of CO_2 and O_2.[118,153] Future work is needed to assess the validity of this proposal through *in vitro* studies, particularly to demonstrate the enzyme's reliance on oxygen.

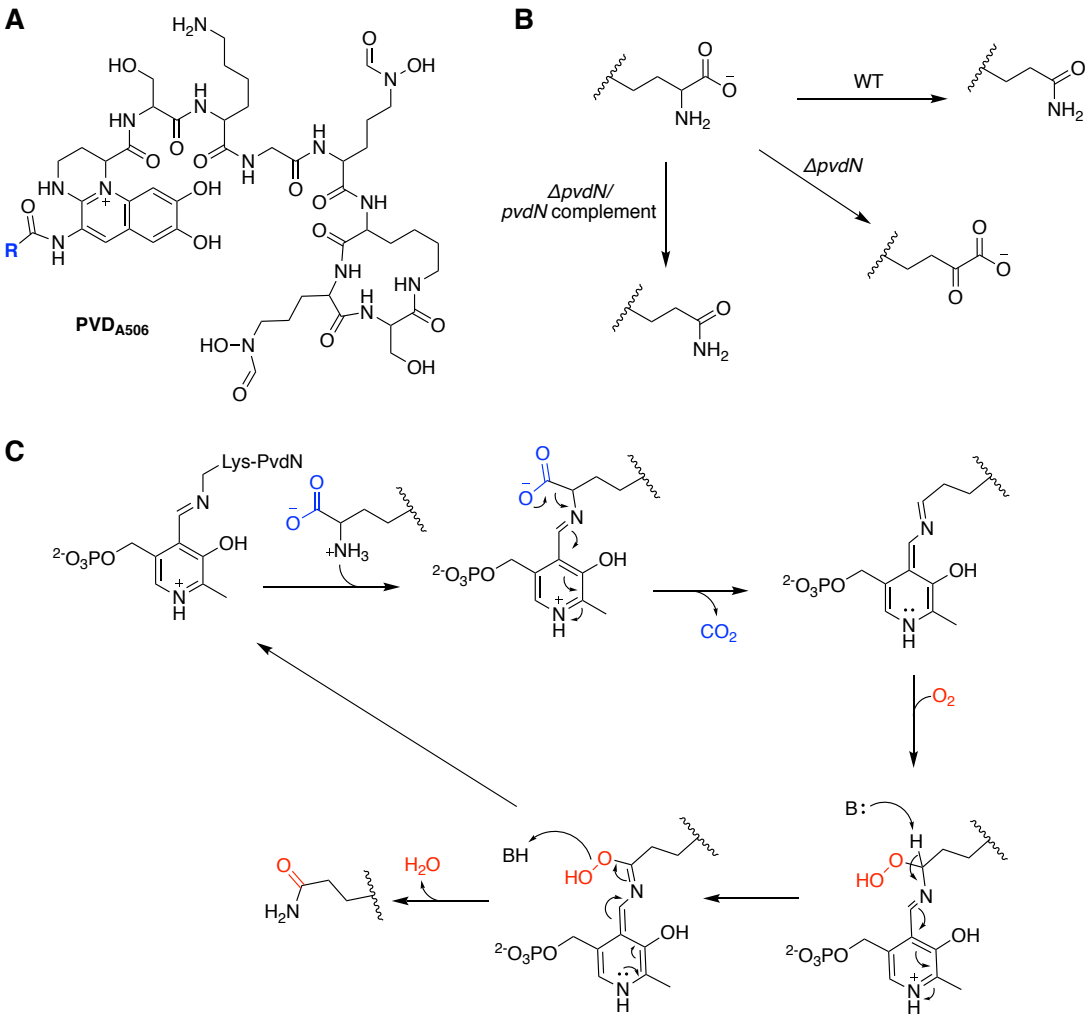


Figure 1.15: Pyoverdine and proposed mechanism of PvdN. (**A**) Structure of pyoverdine, PVD$_{A506}$ with the location of the PvdN-modified substituent indicated by an R-group. (**B**) Biosynthesis of succinamide and α-ketoglutarate substituents from glutamic acid in wild-type *P. fluorescens* strain A506, *ΔpvdN* mutant strain, and *ΔpvdN* mutant strain complemented with *pvdN*. (**C**) Proposed reaction mechanism for PvdN.

1.4.4.2 Cap15

Capuramycins are a class of nucleoside antibiotics originally isolated from *Streptomyces griseus.*[154] Structurally, this class of compounds is characterized by a uridine 5′-carboxamide core appended to an unsaturated hexuronic acid through a glyosidic linkage (**Figure 1.16A**). The biosynthetic gene clusters for capuramycin natural products have been identified.[155,156] The final step in the biosynthesis of the uridine 5′-carboxoamide (CarU) core involves the conversion of epimerized 5′-glycyluridine (GlyU) catalyzed by Cap15 (**Figure 1.16B**). *In vitro* assays showed

that Cap15 is oxygen-dependent because the reaction did not proceed in anaerobic conditions.[119] An isotopic labelling experiment using either $H_3P^{18}O_4$, $H_2{}^{18}O$ or $^{18}O_2$ established Cap15 as a monooxygenase, since isotopic enrichment was only observed in reactions with $^{18}O_2$. CO_2 was also detected in the reaction, which further classifies Cap15 as a monooxygenase-decarboxylase. The authors proposed a mechanism for Cap15 based on ultraviolet-visible spectroscopy (**Figure 1.17**), in which removal of the α-carbon forms a quinonoid intermediate capable or reacting with oxygen. Decarboxylation then occurs, eliminating the distal hydroperoxide oxygen in the form of water, and yielding a CarU-PLP aldimine that then hydrolyzes to release CarU.

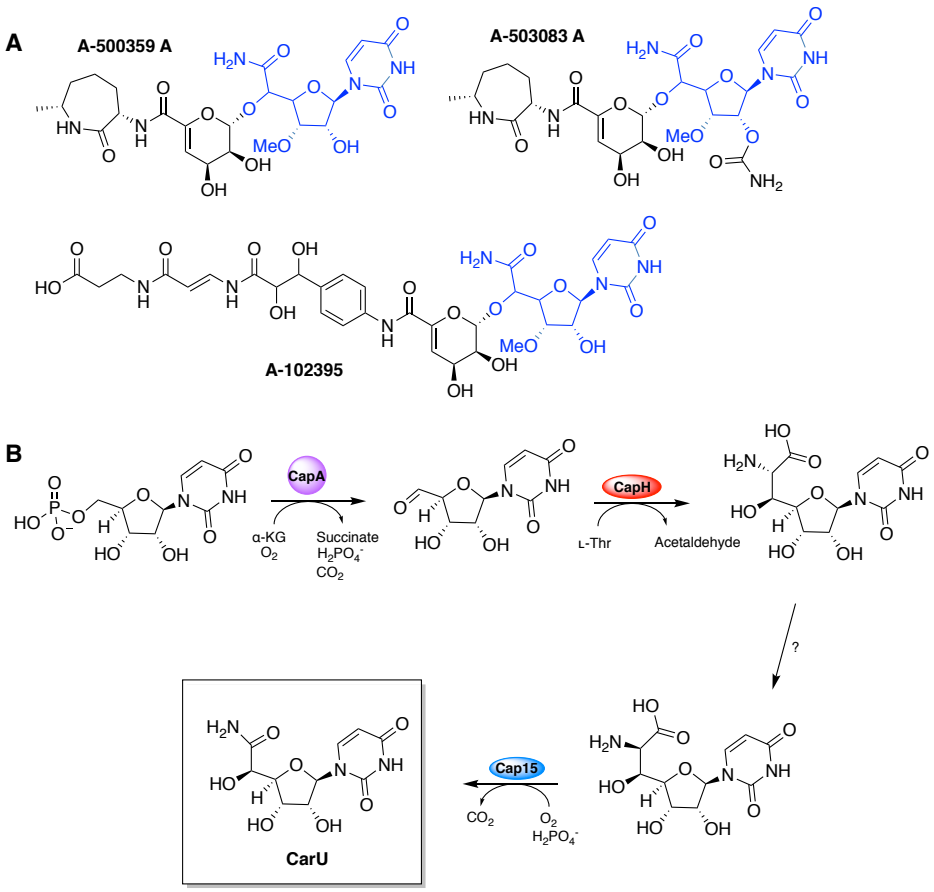


Figure 1.16: Structures and biosynthesis of capuramycin-type molecules. (**A**) Example structures of capuramycin-type metabolites with shared CarU core shown in blue. (**B**) Biosynthesis of CarU core.

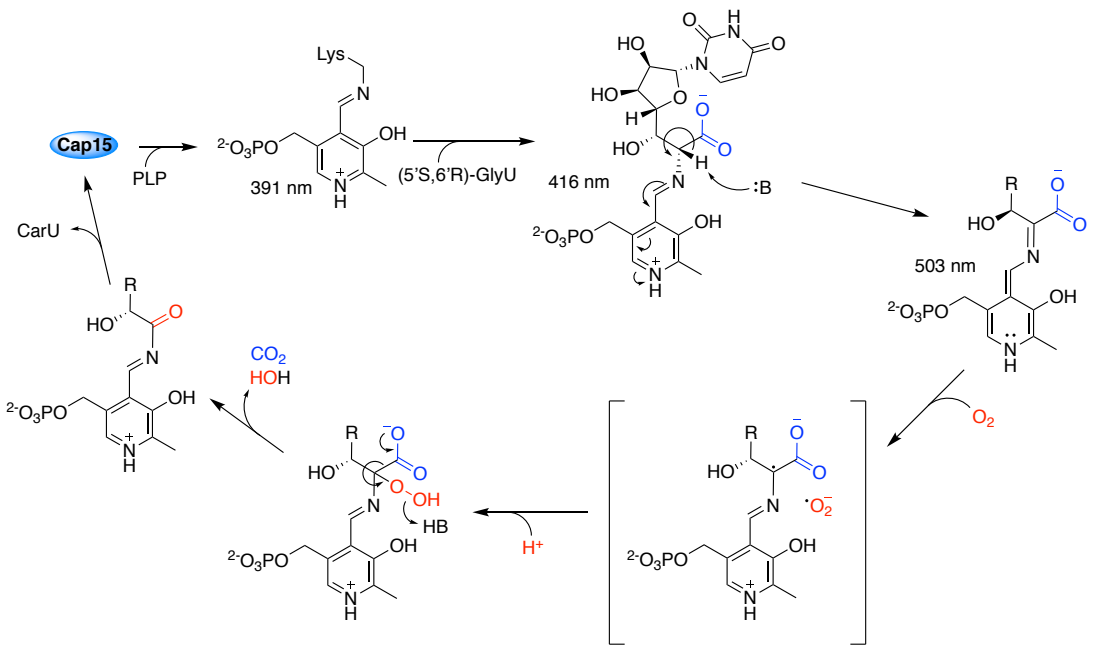


Figure 1.17: Proposed mechanism of Cap15, including observed UV spectra assigned to predicted intermediates.

1.4.4.3 CuaB

Indolizidine alkaloids are a bicyclic-containing natural product group of molecules with antibacterial, antiinflamitory and antitumor bioactivities.[157–159] Curvulamine is an antibacterial indolizidine alkaloid from *Curvularia* sp. IFB-Z10, a fungus associated with the white croaker.[160] During analysis of curvulamine biosynthesis, a novel bifunctional PLP-dependent enzyme, CuaB, was found to be essential to the formation of the bicyclic indolozidine skeleton.[117] First, CuaB catalyzed a Claisen condensation reaction between an acyl carrier protein-bound tetraketide and alanine, similar to the previously known 8-amino-7-oxononanoate synthase (AONS). Then, CuaB catalyzes an oxygen-dependent hydroxylation and cyclization, which was confirmed by isotopic labelling experiments using either $H_2{}^{18}O$ or $^{18}O_2$, the latter of which showed isotopic enhancement in the products. The oxygen-derived hydroxylation appears at the α-carbon of the alanine residue of the Claisen condensation product, suggesting that a hydroperoxyl intermediate forms at this location.

1.4.5 Arginine oxidation

1.4.5.1 MppP

Enduracididine is an arginine-derived, non-proteinogenic, guanidine-containing, heterocyclic amino acid and is a biosynthetic precursor to teixobactin,[161] enduracidin,[162] and mannopeptimycin,[163] which are antibiotics against a variety of bacteria (**Figure 1.18**). Genes for making this non-canonical amino acid were proposed during the discovery of the enduracidin gene cluster[164] through comparison with the mannopeptimycin gene cluster;[165] however, the function of these genes was established by Silvaggi through *in vitro* and crystallographic characterization of enzymes in the pathway[114,115,166] (**Figure 1.18**). First, a PLP-dependent oxidase, MppP/EndP, converts L-arginine into 4-hydroxy-2-ketoarginine,[114,115] which is dehydrated and cyclized by MppR/EndR,[166] and proposed to be aminated by MppQ/EndQ, another PLP-dependent enzyme. Enduracididine is sometimes further hydroxylated at the α-position by MppO, as is seen in mannopeptimycin,[165,167] before being incorporated into a natural product by a non-ribosomal peptide synthetase. Further studies of MppP[114,115] found that it was an oxygen- and PLP-dependent arginine oxidase that catalyzes a two- and four-electron oxidation to produce 2-ketoarginine and (S)-4-hydroxy-2-ketoarginine, respectively, with the hydroxyl group on the (S)-4-hydroxy-2-ketoarginine product coming from water (**Figure 1.19**). The stereochemistry of this product was established using an X-ray crystal structure of MppP soaked with a mixture of reaction products, allowing the 4-hydroxy-2-ketoarginine to bind to the active site.[115] The enzyme uses one or two oxygen molecules, which are stoichiometrically converted to hydrogen peroxide.[115]

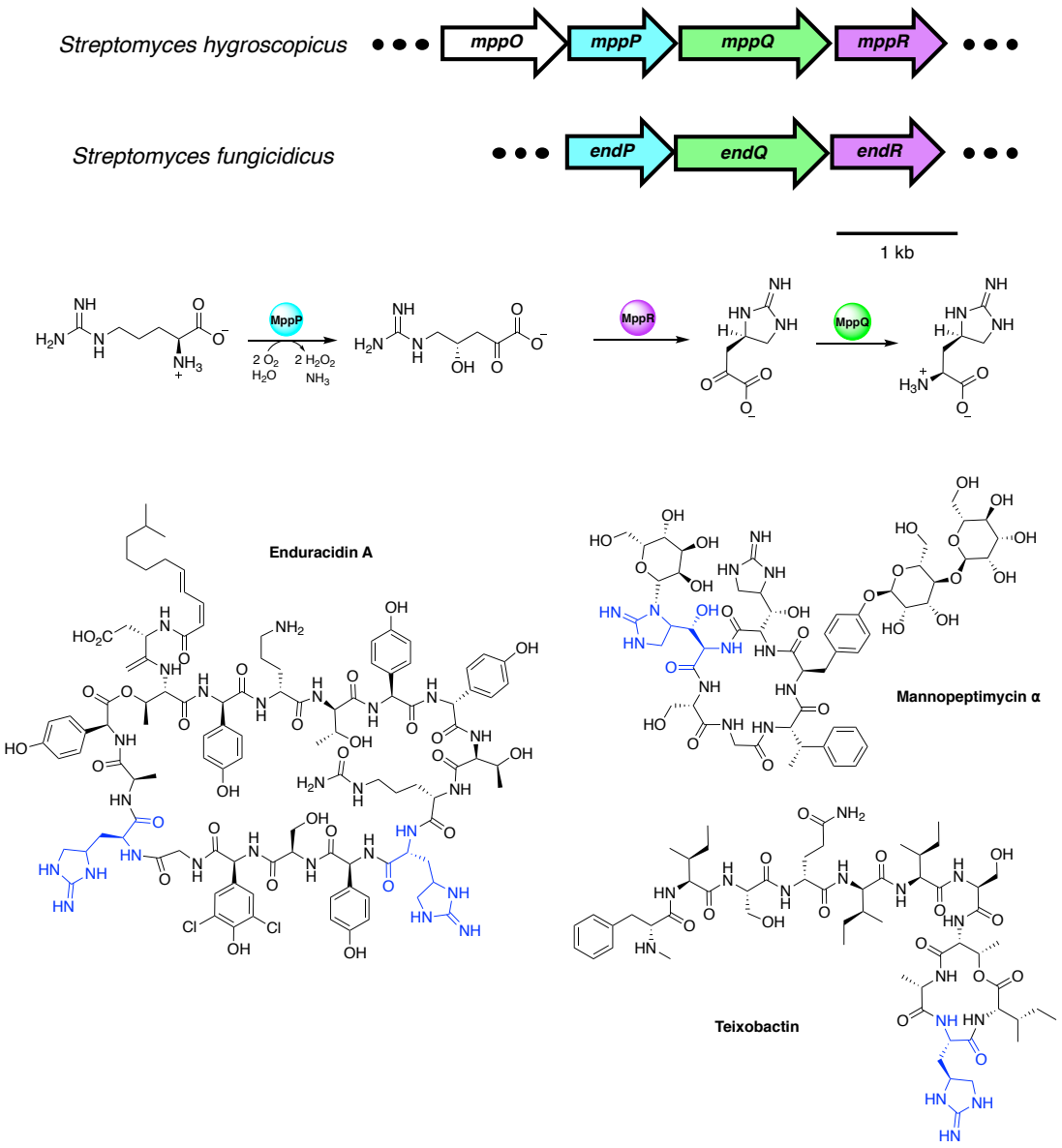


Figure 1.18: The enduracididine biosynthetic pathway and gene cluster. MppP, an O_2- and PLP-dependent enzyme, converts L-arginine to (*S*)-4-hydroxy-2-ketoarginine during enduracididine biosynthesis. Enduracididine-containing natural products are also shown with the enduracididine moiety highlighted in blue. Figure adapted from Hoffarth *et al.* (2020).

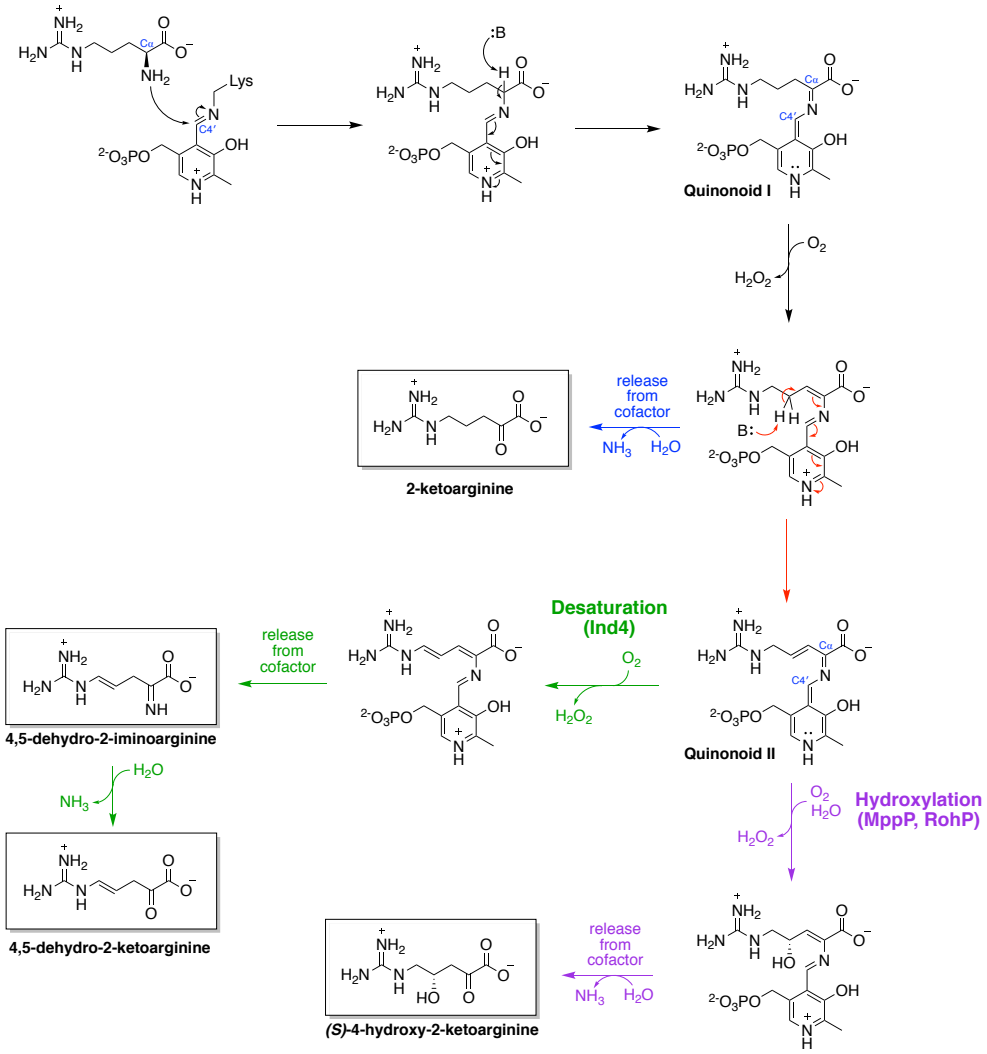


Figure 1.19: Proposed mechanism for known arginine oxidase products. Observed products are indicated with a box. The pathway leading to 2-ketoarginine is indicated with a blue reaction arrow, while the pathway leading to the second oxidation step through quinonoid II is indicated with red reaction arrows. The branching point between desaturation and hydroxylation is shown with green and purple reaction arrows, respectively.

1.4.5.2 RohP

Azomycin, or 2-nitroimidazole, inspired a suite of synthetic nitroimidazole drugs for antimicrobial use.[168] Biosynthetic studies of this molecule revealed that 2-aminoimidazole is a precursor to azomycin.[169–171] In 1977, Eguchi demonstrated that 4-hydroxy-2-ketoarginine was a key precursor to 2-aminoimidazole by feeding radiolabeled arginine to crude enzyme fractions of *Streptomyces eurocidicus*.[172] The authors suggested that an unknown enzyme could catalyze a retro-aldol type reaction on 4-hydroxy-2-ketoarginine to produce pyruvate and

guanidinoacetaldehyde, which would lead to a guanidine cyclization and formation of 2-aminoimidazole. They also showed that the production of 4-hydroxy-2-ketoarginine was dependent on oxygen and PLP. This proposal was recently linked to a group of biosynthetic enzymes by the Ryan group[173]. The biosynthetic pathway begins with arginine being converted to 4-hydroxy-2-ketoarginine by RohP.[116] Then RohR catalyzes a retro-aldol reaction on 4-hydroxy-2-ketoarginine to produce pyruvate and guanidinoacetaldehyde, which is cyclized by RohQ to 2-aminoimidazole. RohS then oxidizes the 2′-amino group to form azomycin (**Figure 1.20**).

RohP was characterized as an oxygen- and PLP-dependent arginine oxidase through *in vitro* characterization, which revealed that oxygen is consumed while 2-ketoarginine, 4-hydroxy-2-ketoarginine and H_2O_2 are produced, similar to what was observed in MppP.[116] H_2O_2 was produced stoichiometrically to oxygen consumption. Additionally, reactions in 50% $H_2{}^{18}O$ demonstrated that the hydroxyl group is derived from water rather than hydrogen peroxide (**Figure 1.19**). Crystal structures of this enzyme have been solved for the holoenzyme, two intermediates (quinonoid I and quinonoid II), and the product-bound enzyme. The configuration of the hydroxyl group was established as (*S*)-4-hydroxy-2-ketoarginine using a product-bound crystal structure. This enzyme plays a key role in the biosynthesis of the azomycin heterocycle because it allows for a retro-aldol reaction to occur in the next step of biosynthesis, after which guanidinoacetaldehyde is capable of cyclizing into 2-aminoimidazole either spontaneously, or by enzymatic catalysis from RohQ.

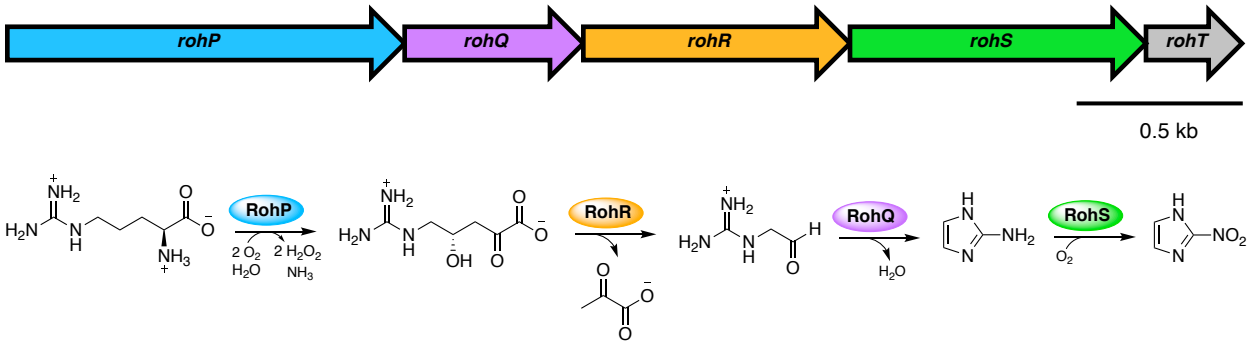


Figure 1.20: The azomycin gene cluster and biosynthetic pathway. RohP, an O_2- and PLP-dependent enzyme, converts L-arginine to (*S*)-4-hydroxy-2-ketoarginine during azomycin biosynthesis. Figure adapted from Hoffarth *et al.* (2020).

1.4.5.3 Ind4

Indolmycin, described earlier in **Section 1.2**, is an antibiotic that was originally discovered in 1960[28] and inhibits prokaryotic TrpRS.[29] It consists of an indole moiety linked to an oxazolinone heterocycle, which is formed by the ATP-dependent condensation of indolmycenic acid and D-4,5-dehydroarginine (**Figure 1.5**). The production of D-4,5-dehydroarginine is dependent on Ind4, an oxygen- and PLP-dependent oxidase, and Ind5, a stereospecific nicotinamide adenine dinucleotide phosphate (NADPH)-dependent dehydrogenase, respectively[113,174]. The double bond formed adjacent to the guanidine group is essential to the formation of the oxazolinone ring, as arginine cannot be used in this transformation, and the adjacent double bond probably plays a role in increasing the electron density toward the guanidine group, allowing one of the nitrogens to act as a nucleophile towards the adenylated carboxylic acid.

Ind4 utilizes one or two molecules of oxygen to catalyze a two- or four-electron oxidation of arginine, producing 2-ketoarginine and 4,5-dehydro-2-iminoarginine, respectively (**Figure 1.19**). Stoichiometric hydrogen peroxide is produced for every molecule of oxygen consumed. The reaction does not proceed in the absence of oxygen and does not require the addition of external co-factor or α-ketoglutarate to regenerate PLP from PMP. Through spectroscopic stopped-flow analysis, a three-step model was proposed to best fit the observed kinetic data under aerobic conditions. In this model, catalysis involves formation of a quinonoid intermediate, stable in the absence of oxygen; formation of a second more oxidized quinonoid intermediate; and conversion of the more oxidized quinonoid into a highly conjugated product-PLP complex. Stopped flow analysis under anaerobic conditions demonstrated accumulation of the first quinonoid intermediate, suggesting that the quinonoid species is reactive with oxygen. In the biosynthetic pathway, the Ind56 complex then stereospecifically reduces the imino acid products to D-configured amino acids, which are not substrates of Ind4. In addition to the indolmycin biosynthetic gene cluster, a number of other non-indolmycin gene clusters contain paired ind45 gene homologs,[113] suggesting that the functions of Ind4- and Ind5-like enzymes are also linked in other biosynthetic pathways.

1.5 Aims of study

Studies in natural product biosynthesis show the importance of exploring natural systems in order to discover new and useful tools that can be added to the biocatalytic repertoire. Previous discovery of the indolmycin biosynthetic machinery reveals that it may be a rich source of biocatalytic potential in two ways: expansion of known reactions catalyzed by enzymes and the ability to manipulate production of indolmycin. This work focuses on using the indolmycin biosynthetic pathway to expand the knowledge of newly discovered enzymatic reactions and to build new indolmycin analogs, which will demonstrate the biocatalytic potential of this pathway and the importance of studying natural product biosynthesis.

In **Chapters 2** and **3**, it was hypothesized that structural, biochemical and phylogenetic studies of the arginine desaturases would facilitate a better understanding of the arginine oxidases and their mechanisms. There are two elusive aspects of the mechanism that this study aims to solve: the mechanism of oxygen usage and the differentiation of hydroxylation and desaturation. X-ray crystallography will be used to obtain the first crystal structures of an arginine desaturase and mutagenesis based on these crystal structures will be used to explore the catalytic residues of the desaturases. Additionally, the sequence similarity and phylogenetic relationship between the desaturases and hydroxylases will be analyzed to better understand the evolutionary history of these enzymes. Several new homologs of Ind4 with desaturase activity will be identified and their products will be characterized in further detail. Finally, a mechanism is proposed that unifies the experimental and calculated data for both the hydroxylases and desaturases. This information will help guide potential efforts to discover and engineer arginine oxidases and other oxygen- and PLP-dependent enzymes. Advances in oxygen- and PLP-dependent enzyme catalysis will expand the biocatalytic repertoire and aid the development of more PLP-dependent enzymes for use as biocatalysts.

In **Chapter 4**, it was hypothesized that heterologous expression of the indolmycin biosynthetic genes in *E. coli* would provide a useful method of making indolmycin. Additionally, advances in the biocatalytic use of TrpS to make various tryptophan derivatives from substituted indoles and L-serine suggest that coupling this enzyme with the indolmycin biosynthetic genes, *in vivo*, may be a simple way to diversify indolmycin analogs. Therefore, heterologous expression of indolmycin genes, paired with a promiscuous TrpS, will be explored here for use in the production of diverse indolmycin analogs. The production of more diverse indolmycin

analogs may benefit the fight against antimicrobial resistance by providing more molecules that can be tested for desirable qualities. Additionally, use of this heterologous system, or parts of it, will demonstrate the importance of understanding how biosynthetic enzymes can be used effectively as biocatalysts to make new molecules that may be challenging for solely synthetic chemical methods.

Chapter 2: Crystallization and mutagenesis of arginine oxidases[2]

2.1 Introduction

As described in **Chapter 1**, the arginine oxidases are a group of oxygen- and PLP-dependent enzymes that catalyze two types of reactions: arginine hydroxylation and arginine desaturation. The hydroxylases, RohP and MppP, use two molecules of molecular oxygen to produce one molecule of (*S*)-4-hydroxy-2-ketoarginine (**3**) and two molecules of hydrogen peroxide from L-arginine (**1**) or one molecule of molecular oxygen to produce one molecule of 2-ketoarginine (**4**) and one molecule of hydrogen peroxide from **1** (**Figure 2.1**).[114–116] Ind4, a desaturase, uses two molecules of molecular oxygen to produce one molecule of 4,5-dehydro-2-iminoarginine (**2**), converting to 4,5-dehydro-2-ketoarginine (**5**) under longer reaction times, and two molecules of hydrogen peroxide from **1**. Ind4 also consumes one molecule of molecular oxygen to produce one molecule of **4** and one molecule of hydrogen peroxide from **1** (**Figure 2.1**).

Despite catalyzing different reactions, the hydroxylase and desaturase mechanisms are expected to be related, based on their similar usage of a PLP cofactor, stoichiometric conversion of oxygen to hydrogen peroxide and production of **4** when only one molecule of oxygen is used. Additionally, the hydroxylases share between 33%-42% sequence identity with Ind4, which will be explored with further detail in **Chapter 3**. The nature of the (potentially) shared mechanism within the arginine oxidases has not been adequately studied, primarily due to a lack of structural information from the desaturases. Crystal structures have been previously solved for both SwMppP and RohP, both of which showed that N-terminal ordering occurred as a result of substrate binding and was important for catalysis.[114–116] Additionally, it was shown that an active site His residue was important in the production of **3**.[116] There were two methods used to trap intermediates of catalysis in the arginine hydroxylases. In SwMppP, an external aldimine intermediate was formed when **1** was soaked into the crystal, but a product-bound structure was obtained by soaking the crystal with reaction products.[115] By contrast in RohP, intermediate-bound and product-bound structures were obtained by forming the intermediate and product ligands *in crystallo* by soaking **1** into the crystals and freezing each at different time points.[116] Each time point and trapped intermediate corresponded to a different point in catalysis and

[2] Adapted from a manuscript that is in revision for publication.

displayed a different colour and ultraviolet-visible (UV-Vis) signal. No oxygen atoms from molecular oxygen are incorporated into the final products of the arginine oxidase products and crystallographic studies of the arginine hydroxylases did not provide other evidence of peroxy-intermediate formation during the reaction, meaning the mechanism of oxygen utilization also remains elusive.

In this chapter, the first crystal structure of an arginine desaturase is described and compared with known structures of arginine hydroxylases. This structural comparison, along with sequence alignment information, was then used to generate several desaturase variants to better understand the similarities and differences between the two types of enzymes. This new information helped to identify some of the catalytic residues of the desaturases. When compared to the hydroxylases, this new structural information helps gain a deeper understanding of the mechanism and provides insight to these unique enzymes, which is invaluable in efforts to further expand the known biocatalytic repertoire.

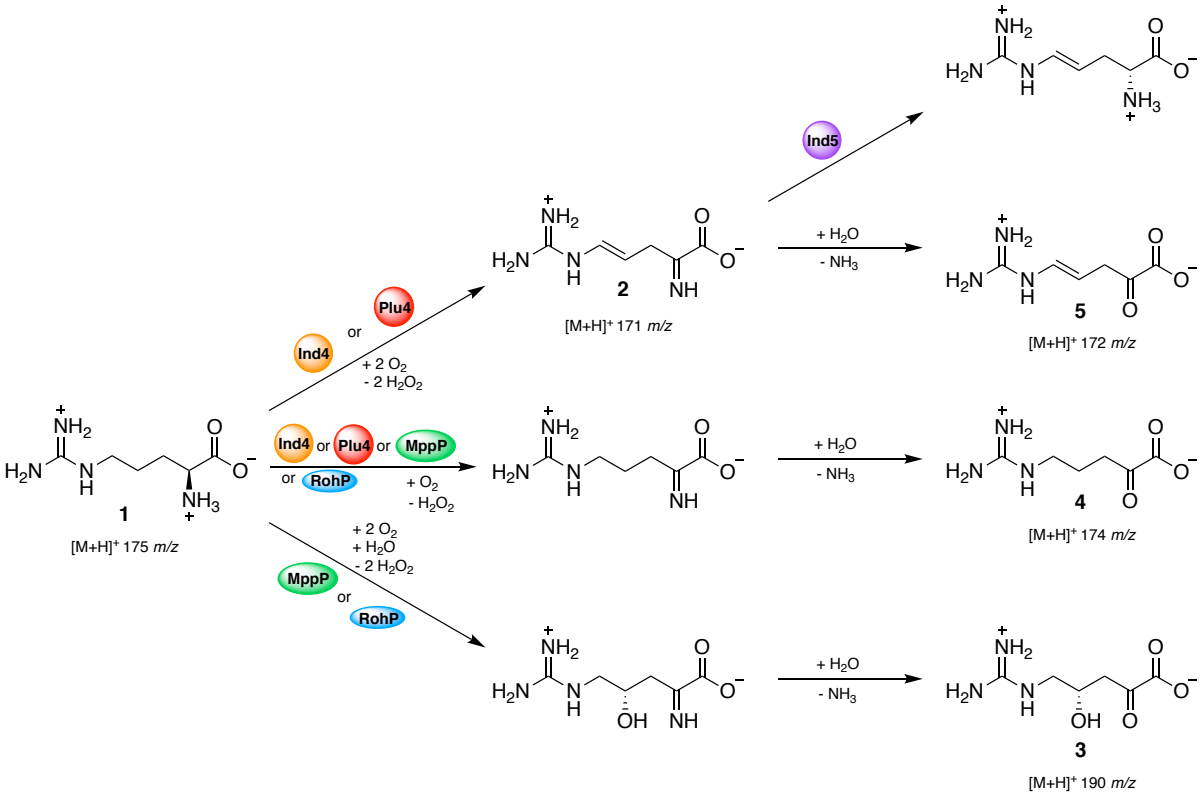


Figure 2.1: Scheme of O_2- and PLP-dependent arginine oxidase reactivities. $[M+H]^+$ signals are given below observed products. The Ind4-coupled reaction with Ind5 from the indolmycin biosynthetic pathway is also shown.

2.2 Materials and methods

2.2.1 General methods, materials and strains

Reagents were purchased from Sigma-Aldrich, Hampton Research, Thermo Fisher Scientific Canada, New England BioLabs (NEB), Bio-Rad, Bio Basic Inc., Gold Biotechnology, and VWR International. Columns and resins for protein purification were obtained from GE Healthcare Life Sciences. Primers were purchased from Integrated DNA Technologies (IDT). Sequencing was done through the NAPS Unit in the Sequencing and Bioinformatics Consortium at the University of British Columbia or in the DNA Sequencing Core Facility at the Center for Molecular Medicine and Therapeutics at the University of British Columbia. The strains and plasmids used in this study are listed in **Table 2.1**.

Table 2.1: Bacterial strains and plasmids used in this study.

Strain/Vector	**Description**	**Source**
Strain		
E. coli DH5*α*	General cloning host	Laboratory stock
E. coli BL21 (DE3)	Host for protein expression	Laboratory stock
Plasmid		
pET22b	Vector for protein expression in *E. coli*	Laboratory stock
pET28a	Vector for protein expression in *E. coli*	Laboratory stock
pET22b-ind4	Vector for Ind4 expression, cloning sites NdeI/XhoI	Du *et al.*[48]
pET22b-plu4	Vector for Plu4 expression, cloning sites NdeI/XhoI	This study
pET22b-plu4H27A	Vector for Plu4 variant H27A expression	This study
pET22b-plu4D225A	Vector for Plu4 variant D225A expression	This study
pET22b-plu4D225N	Vector for Plu4 variant D225N expression	This study
pET22b-plu4D25A	Vector for Plu4 variant D25A expression	This study
pET22b-plu4D25N	Vector for Plu4 variant D25N expression	This study
pET22b-plu4Y29H	Vector for Plu4 variant Y29H expression	This study
pET22b-plu4Y29F	Vector for Plu4 variant Y29F expression	This study
pET22b-plu4Y29R	Vector for Plu4 variant Y29R expression	This study

Strain/Vector	Description	Source
pET22b-plu4C251A	Vector for Plu4 variant C251A expression	This study
pET22b-plu4C251G	Vector for Plu4 variant C251G expression	This study
pET22b-plu4C251N	Vector for Plu4 variant C251N expression	This study
pET22b-plu4T189C	Vector for Plu4 variant T189C expression	This study
pET22b-plu4T189S	Vector for Plu4 variant T189S expression	This study
pET22b-plu4G26A	Vector for Plu4 variant G26A expression	This study
pET22b-plu4A28T	Vector for Plu4 variant A28T expression	This study
pET22b-plu4L117D	Vector for Plu4 variant L117D expression	This study
pET22b-ind4G24A	Vector for Ind4 variant G24A expression	This study
pET22b-ind4A26T	Vector for Ind4 variant A26T expression	This study

2.2.2 Cloning and mutagenesis

The Plu4-containing plasmid (pET22b-plu4) was previously synthesized by Bio Basic Inc. and was sub-cloned into pET22b to contain a C-terminal histidine tag (unpublished). Plasmids from successful colonies were purified using the QIAprep Spin Miniprep Kit from Qiagen and transformed into chemically competent *E. coli* BL21 (DE3) by heat shocking for protein expression. pET22b-ind4 was made in a previous study.[48] A description of the identification, cloning and characterization of other Ind4 homologs is described in **Chapter 3**.

Plu4 and Ind4 residues targeted for mutagenesis were identified by analyzing crystal structures and sequence alignments made from known and predicted arginine hydroxylases and desaturases. Sequence alignments were made using Muscle in UGENE.[175] The full sequence alignment was visualized with secondary structure information using ESPript 3.0.[176] Plu4 and Ind4 variants were made using the primers listed in **Table 2.2**. Q5 polymerase (NEB) was used to amplify the template plasmid (pET22b-plu4 or pET22b-ind4) and introduce mutations by polymerase chain reaction (PCR). The PCR product was digested with DpnI to remove template DNA and transformed into *E. coli* DH5α. Colonies were miniprepped and sequenced using T7 primers (**Table 2.2**) to confirm that the gene had the intended mutation and was free of errors. Successfully mutated plasmids were transformed into chemically competent *E. coli* BL21 (DE3) by heat shocking for protein expression and purification.

Table 2.2: Oligonucleotide primers used in this study.

Primer	Sequence (5′ to 3′)	Description
T7	TAATACGACTCACTATAGGG	T7 promoter for sequencing
T7-term	GCTAGTTATTGCTCAGCGG	T7 terminator for sequencing
Plu4-H27A-F2	CTTCGCAGACGGTGCGGCATACCATGACATCAACGAG	H27A variant of Plu4
Plu4-H27A-R2	GATGTCATGGTATGCCGCACCGTCTGCGAAGTTGTACAGG	H27A variant of Plu4
Plu4-D225A-F3	CTGGCCGACGCAAGCGCTGAAAATCTCCCTGATGGTG	D225A variant of Plu4
Plu4-D225A-R3	GGAGATTTTCAGCGCTTGCGTCGGCCAGGTTTTACC	D225A variant of Plu4
Plu4-D225N-F2	CTGGCCGACGCAAAACCTGAAAATCTCCCTGATGG	D225N variant of Plu4
Plu4-D225N-R2	GGAGATTTTCAGGTTTTGCGTCGGCCAGG	D225N variant of Plu4
Plu4-D25A-F	CAACTTCGCAGCCGGTCATGCATACC	D25A variant of Plu4
Plu4-D25A-R	GGTATGCATGACCGGCTGCGAAGTTG	D25A variant of Plu4
Plu4-D25N-F	CAACTTCGCAAACGGTCATGCATACC	D25N variant of Plu4
Plu4-D25N-R	GGTATGCATGACCGTTTGCGAAGTTG	D25N variant of Plu4
Plu4-Y29H-F	CAGACGGTCATGCACATCATGACATCAACGAG	Y29H variant of Plu4
Plu4-Y29H-R	CTCGTTGATGTCATGATGTGCATGACCGTCTG	Y29H variant of Plu4
Plu4-Y29F-F	CAGACGGTCATGCATTCCATGACATCAAC	Y29F variant of Plu4
Plu4-Y29F-R	GTTGATGTCATGGAATGCATGACCGTCTG	Y29F variant of Plu4
Plu4-Y29R-F	CAGACGGTCATGCACGCCATGACATCAAC	Y29R variant of Plu4
Plu4-Y29R-R	GTTGATGTCATGGCGTGCATGACCGTCTG	Y29R variant of Plu4
Plu4-C251A-F	AAGAAATTTTCCTGGCGAGCAGCAACTTCGCC	C251A variant of Plu4
Plu4-C251A-R	GGCGAAGTTGCTGCTCGCCAGGAAAATTTCTT	C251A variant of Plu4
Plu4-C251G-F	AAGAAATTTTCCTGGGCAGCAGCAACTTC	C251G variant of Plu4
Plu4-C251G-R	GAAGTTGCTGCTGCCCAGGAAAATTTCTTC	C251G variant of Plu4
Plu4-C251N-F	AAGAAATTTTCCTGAACAGCAGCAACTTCGCC	C251N variant of Plu4
Plu4-C251N-R	GGCGAAGTTGCTGCTGTTCAGGAAAATTTCTT	C251N variant of Plu4
Plu4-T189C-F	CCTGCTGGACCGTTGCTTCCGTATCTACGG	T189C variant of Plu4
Plu4-T189C-R	CCGTAGATACGGAAGCAACGGTCCAGCAGG	T189C variant of Plu4
Plu4-T189S-F	CCTGCTGGACCGTAGCTTCCGTATCTACGG	T189S variant of Plu4
Plu4-T189S-R	CCGTAGATACGGAAGCTACGGTCCAGCAGG	T189S variant of Plu4
Plu4-G26A-F	CAACTTCGCAGACGCGCATGCATACCATGAC	G26A variant of Plu4
Plu4-G26A-R	GTCATGGTATGCATGCGCGTCTGCGAAGTTG	G26A variant of Plu4
Ind4-G24A-F	CTCGCCGATGCGCACGCCCATCAGGG	G24A variant of Ind4
Ind4-G24A-R	CCTGATGGGCGTGCGCATCGGCGAGATTG	G24A variant of Ind4
Plu4-A28T-F	CTTCGCAGACGGTCATACCTACCATGACATCAACG	A28T variant of Plu4
Plu4-A28T-R	CGTTGATGTCATGGTAGGTATGACCGTCTGCGAAG	A28T variant of Plu4
Ind4-A26T-F	GATGGCCACACCCATCAGGGCCAG	A26T variant of Ind4
Ind4-A26T-R	CTGATGGGTGTGGCCATCGGC	A26T variant of Ind4
Plu4-L117D-F	GATAACCTGTACGATCTGCTGAAACGTCGTGGTGTTG	L117D variant of Plu4
Plu4-L117D-R	GACGTTTCAGCAGATCGTACAGGTTATCAAAAGCCGGCTCG	L117D variant of Plu4

2.2.3 Protein purification and Plu4/Ind4 variant characterization

E. coli BL21 (DE3) cultures were inoculated from a 5 mL overnight culture into 1 L of Luria-Bertani (LB) media containing 100 μg/mL ampicillin at 37°C and 200 rpm to an OD_{600} of 0.4-0.6. The temperature and shaking were then lowered to 16°C and 150 rpm, respectively, for 30 min before addition of 0.5 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG). The cultures were incubated for 16 h at 16°C and 150 rpm before the cells were pelleted by centrifugation. Pellets were resuspended in a Buffer A (50 mM phosphate pH 8.0, 300 mM NaCl, 10 mM imidazole) and lysed by sonication. Cell lysates were clarified by centrifugation (40,500 x g for 45 min) to remove cell debris and diluted five-fold with Buffer A. The diluted supernatant was loaded onto Ni-charged Chelating Sepharose Fast-flow resin by gravity filtration. The resin was washed with five volumes of Buffer A, then with five volumes of Buffer B (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM imidazole). The protein was eluted from the resin with Buffer C (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 500 mM imidazole). Fractions containing the protein were yellow and their purity was assessed using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

The protein-containing fractions were pooled and concentrated to ~6-8 mg/mL using an Amicon Ultra-15 Centrifugal Spin Filter (10 kDa molecular weight cut-off) from Millipore Sigma before being further purified on a HiLoad™ 16/600 Superdex™ 200 pg gel filtration column. The column was pre-equilibrated with Buffer D (20 mM HEPES pH 7.5, 50 mM NaCl) and was run at 1 mL/min to elute the protein. Fractions were pooled and all PLP and bound substrates were removed from Plu4 by dialyzing in Buffer D containing 10% glycerol and 50 mM of hydroxylamine for 2 h, then overnight into Buffer D containing 10% glycerol to remove hydroxylamine. Plu4 was reconstituted with PLP by dialysis overnight in Buffer D containing 10% glycerol PLP in five times molar excess of the protein concentration. The dialyzed samples were concentrated to ~40 mg/mL, as described above. The protein was aliquoted into 100 μL samples, frozen in liquid nitrogen and stored at −70°C. Protein concentrations were determined throughout purification by absorbance at 280 nm.

Variants were purified on a smaller scale using the above method with small changes. Cells were lysed using BugBuster® 10X Protein Extraction Reagent (Novagen), diluted to working concentration in Buffer A. After centrifuging, the supernatant was diluted two-fold with Buffer A and purified using HisPur™ Ni-NTA Spin Column (Thermo Scientific), instead of the

Ni-charged Chelating Sepharose Fast-flow resin. The variants were not further purified by gel filtration and were dialyzed overnight in Buffer D containing 10% glycerol and PLP in five times molar excess of the protein concentration to reconstitute the PLP without first removing all bound PLP and substrates. Each Plu4 or Ind4 variant was assayed by electrospray ionization-mass spectrometry (ESI-MS). 100 µL reactions were set up overnight at 30°C with 1 mM L-arginine, 25 µg/mL catalase, 50 mM Tris-HCl (pH 7.5) and 5 µM of enzyme. Reactions were stopped with 100 µL methanol after 16 h and particulate matter was removed by centrifugation. A 20 µL aliquot of the supernatant was analyzed using a 6120 Quadrupole liquid chromatography-mass spectrometry (LC-MS) system (Agilent) operated in positive ion mode using a Poroshell 120 EC-C18, 2.7 µm, 4.6 x 50 mm column (Agilent).

2.2.4 Crystallization

Initial crystals were obtained from 96-well screens, set up manually, using 5 and 15 mg/mL of protein. Diffraction-quality crystals were obtained from 2 µL of PLP-reconstituted Plu4 at 12-14 mg/mL mixed in equal amounts with the crystallization solution (0.1 M Bis-Tris:HCl, pH 6.5; 27% PEG 3350; 0.2 M ammonium acetate) in using hanging drop vapour diffusion methods. Spontaneous crystals were unable to grow at pH 6.5, so crystals were grown in the same condition at pH 6.0 and were used for seeding drops at pH 6.5. The seed stock was made by crushing a large, thin plate in a 7 µL drop of the mother liquor. Seeding was done by passing a seeding tool (Hampton Research) through the seed stock drop and transferring microseeds to a protein drop, which had been allowed to equilibrate over the well solution for 30 minutes. Crystals grew to diffraction quality after 3-7 days.

Crystals were soaked by adding 1 µL of 10-15 mM of L-arginine and saturated PLP (~10 mM) in mother liquor or 1 µL of saturated PLP in mother liquor directly to the 4 µL crystallization drop, producing a final concentration of 2-3 mM L-arginine and/or ~2 mM PLP. Crystals were frozen when they were either yellow (2 h with PLP only) for the holo-enzyme structure, orange (8 minutes with PLP and L-arginine) for the intermediate-bound structure, or yellow (16 h with PLP and L-arginine) for the product-bound structure. Crystals were cryoprotected with 0.1 M Bis-Tris:HCl (pH 6.5), 28% PEG 3350, and 15% ethylene glycol prior to freezing in liquid nitrogen and collecting X-ray diffraction data.

2.2.5 Data collection, structure determination and computational methods

Data were collected from the Stanford Synchrotron Radiation Lightsource (Menlo Park, California) at beamline 9-2 using a Dectris Pilatus 6M detector and a wavelength of 0.97946 Å. UV-Vis spectra were measured on each single crystal while mounted on the X-ray detector using a microspectrophotometer prior to collecting X-ray diffraction data sets. The microspectrophotometer uses UV solarization-resistant optical fibres, two reflective Newport Schwardchild objectives, Ocean Optics QE65000 Spectrum Analyzer and a Hamamatsu light source with deuterium and halogen lamps. The data set was scaled and integrated using autoxds. Plu4 crystallized as a homodimer in the space group $P2_1$, containing two or four monomers in the asymmetric unit. The data set was phased using Phaser-MR[177] in the Phenix software package using Chain A of a RohP crystal structure (PDB: 6C3D)[116] as a search model, which had 30% sequence identity to Plu4. The search model was generated using CHAINSAW[178] in CCP4 to generate a poly-alanine model of the structure, without the N-terminus (residues 1-26 were removed). The initial model was built using Phenix Autobuild[179]. The structure was iteratively modified in COOT[180] and refined with Phenix refine[181] using translation-liberation-screw (TLS) refinement. Simulated annealing was also used during initial refinement iterations to remove model bias. Residues with side chains that did not fit the electron density were deleted. Missing residues and side chains were built in gradually according to the positive F_o-F_c density contoured at 3.0 σ. Active site molecules, including cofactor, substrate or other molecules from crystallization conditions, were also built in according to the positive F_o-F_c density contoured to 3.0 σ. Restraints for the non-standard molecules were generated in Phenix eLBOW.[178] Ligands were built into the intermediate structure when refinement was nearly complete, including placement of water molecules. Although there is likely a mixture of intermediate ligands present in the active site of this structure, only Quinonoid I was built into the active site with full occupancy based on the dominant UV-Vis signal observed at 510 nm with the microspectrophotometer. Data collection information and refinement statistics are given in **Table 2.3**.

Density functional theory (DFT) calculations were done to determine that the imine was the most stable tautomer form of the 4,5-dehydro-2-iminoarginine product independent of influence from the enzyme active site. All quantum mechanical calculations were performed with Gaussian 09.[182] Geometry optimizations were calculated with the B3LYP[183–186] density

functional with the IEFPCM model (diethyl ether),[187] and the 6-31G(d) basis set for the imine and enamine system (**Table 2.4**). Free energy corrections were calculated using Truhlar's quasiharmonic approximation.[188,189] Single point energies were calculated using M06-2X[190] with IEPPCM (diethyl ether),[187] and the 6-311++G(d,p) basis set for the imine and enamine system (**Table 2.5**). Monte Carlo conformational searches were performed to identify lowest energy conformations with the OPLS3 force field[191] in Maestro/Macromodel.[192] Structure graphics were generated using CYLview.[193] Docking of the more stable imine products were done through Autodock Vina[194] to determine the likelihood of the observed product-binding mode in the product-bound structure.

Table 2.3: Data collection and refinement statistics.

	Holo	Intermediate	Product
Data Collection[a]			
Wavelength	0.9795	0.9795	0.9795
Space Group	$P2_1$	$P2_1$	$P2_1$
Dimensions			
a	75.02	74.92	76.37
b	72.19	70.28	72.19
c	76.27	138.57	139.63
beta	112.09	98.11	98.44
Resolution (Å)	1.98	1.93	2.5
High Resolution Range (Å)	2.05 – 1.98	2.00 – 1.93	2.59 – 2.5
$CC_{1/2}$[b]	0.997 (0.514)	0.996 (0.852)	0.995 (0.486)
Completeness	98.6 (94.2)	98.8 (90.7)	99.1 (98.0)
Unique Reflections	52155 (3510)	106646 (4801)	51795 (4444)
Multiplicity	6.9 (6.6)	13.6 (12.3)	13.6 (13.1)
$I/\sigma I$	15.2 (1.4)	11.1 (1.9)	7.3 (1.5)
Refinement[a]			
R_{work}[c]	0.1663 (0.2753)	0.1512 (0.2334)	0.2339 (0.3804)
R_{free}[d]	0.2159 (0.3241)	0.1944 (0.2900)	0.2711 (0.4526)
No. non-hydrogen atoms	6411	13549	11779
Protein	6029	12130	11503
Solvent	330	1130	180
Ligands	52	289	96
RMSD bonds (Å)	0.010	0.012	0.007
RMSD angles (deg)	1.31	1.51	0.83
Ramachadran favoured (%)	97.45	97.94	96.84
Ramchadran allowed (%)	2.55	2.06	3.16
Ramachadran outliers (%)	0	0	0
Average B-factor ($Å^2$)	52.1	30.6	67.3
Protein	51.8	29.6	67.4
Solvent	52.1	39.3	54.7
Ligands	80.1	41	82.4
No. TLS groups	23	38	27

[a]Values in outermost resolution shell are given in parentheses

[b]$CC_{1/2} = \sum_i (x_i - \bar{x})(y_i - \bar{y}) / \sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}$ for a random half of unique reflection measurements

[c]$R_{work} = \sum_{hkl} \left| |F_o(hkl)| - |F_c(hkl)| \right| / \sum_{hkl} |F_o(hkl)|$ for the 95% of reflection data used in refinement

[d]R_{free} is calculated identically to R_{work}, but only for the 5% of reflection data not used in refinement

Table 2.4: Cartesian coordinates of optimized structures for DFT calculations of imine and enamine tautomer stabilities.

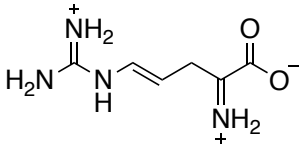
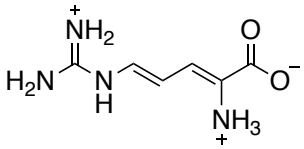
	N -0.843153 -2.576308 -0.254737 C -0.538314 -1.511212 0.397897 C -1.041715 -1.226064 1.776727 C 0.440129 -0.545243 -0.378946 C -1.768424 0.098559 1.812864 O 0.710692 -0.966771 -1.522319 C -3.046291 0.21358 2.178174 N -3.683775 1.469455 2.291029 H -3.089864 2.291223 2.291936 C -5.010863 1.675031 2.359907 N -5.864144 0.691593 2.053299 N -5.46707 2.864678 2.760479 H -5.587259 -0.060723 1.437834 H -6.854172 0.796412 2.228491 H -4.860876 3.535247 3.212811 H -6.434988 3.125991 2.632134 H -1.44156 -3.324479 0.087095 O 0.783508 0.459293 0.267258 H -1.668014 -2.042464 2.150647 H -1.173268 0.960867 1.526783 H -3.66011 -0.633075 2.467842 H -0.150708 -1.157087 2.417263 H -0.408746 -2.620687 -1.188919
	N -0.000203 -2.599957 0.314223 H 0.394569 -2.250783 -0.657646 C -0.872337 -1.430574 0.545575 C -1.649166 -1.328627 1.637326 C -0.686144 -0.485783 -0.654518 C -2.509257 -0.192796 1.892036 O 0.141375 -0.970294 -1.497975 C -3.298118 -0.120622 2.978247 N -4.168412 0.962533 3.189361 H -4.280566 1.616697 2.421712 C -4.848678 1.231394 4.32028 N -4.520724 0.629356 5.468384 N -5.867909 2.093464 4.275392 H -3.586774 0.273491 5.619096 H -5.117055 0.709272 6.281154 H -6.295045 2.360646 3.398721 H -6.257455 2.488264 5.12047 H 0.748815 -2.712195 1.003066 O -1.313579 0.582243 -0.690791 H -1.641298 -2.13068 2.375381 H -2.48591 0.595824 1.144365 H -3.349444 -0.905335 3.725405 H -0.49822 -3.489732 0.221317

Table 2.5: Energies of imine and enamine tautomers. ZPVE = zero-point vibrational energy; TCE = thermal correction to energy; TCH = thermal correction to enthalpy; TCG = thermal correction to Gibbs free energy. Energies are given in kcal/mol.

	ZPVE	**TCE**	**TCH**	**ECG**	**E**	**H (E+TCH)**	**G (E+TCG)**
imine	0.188227	0.201426	0.202370	0.147147	-604.539284	-604.336914	-604.392137
enamine	0.187753	0.200762	0.201706	0.148001	-604.334089	-604.333145	-604.386850

2.3 Results

2.3.1 Plu4 crystal structures

To elucidate the determinants of catalysis in arginine desaturases, attempts were made to solve the X-ray crystal structure of Ind4. However, crystals could not be obtained for Ind4, so Ind4 homologs were cloned and purified to obtain crystal structures of various points in the enzyme's catalytic cycle. Plu4, an Ind4 homolog (32% sequence identity) from a convergent indolmycin gene cluster in *P. luteoviolacea*,[68] was the only homolog that produced crystal structures with trapped intermediates and products. By freezing the crystals soaked under aerobic conditions for different periods of time coinciding with colour changes in the crystal (**Figure 2.2**), three crystal structures were obtained: holo-enzyme (yellow), intermediate-bound enzyme (orange) and product-bound enzyme (yellow after previously being orange) (**Figure 2.3**). Each crystal structure represents a different stage of catalysis and can be compared to the hydroxylase crystal structures of RohP and SwMppP.

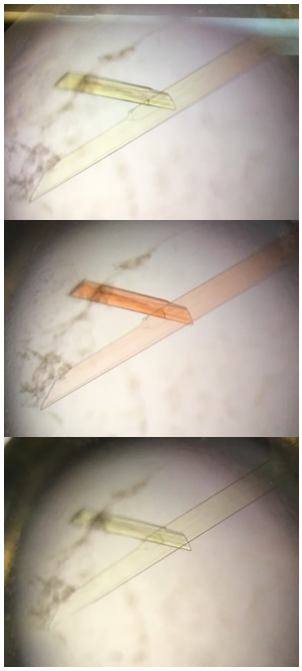


Figure 2.2: Plu4 crystals after soaking PLP or PLP and arginine at different time points.

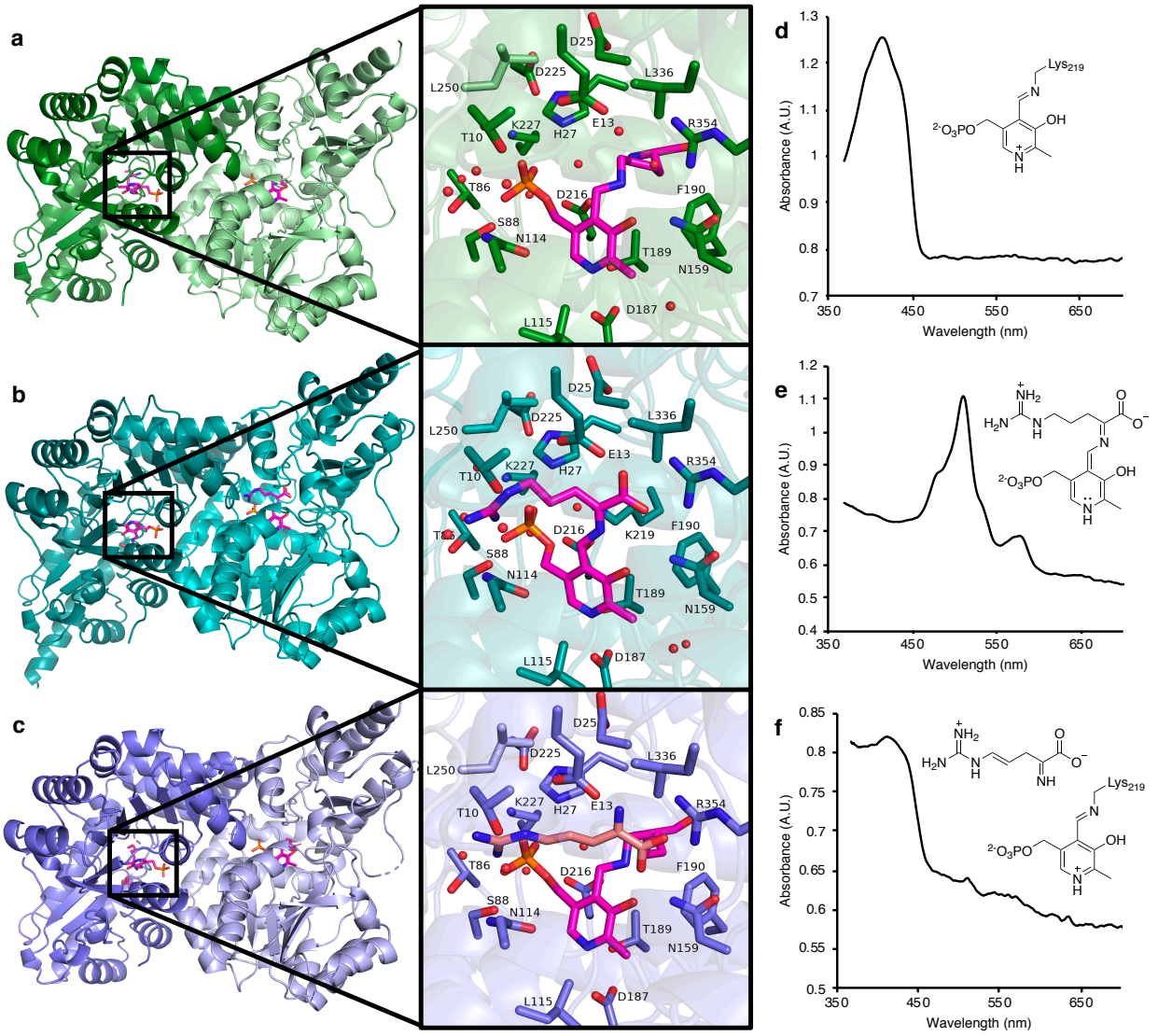


Figure 2.3: X-ray crystal structures of Plu4. (**a**) Holo-enzyme. (**b**) Quinonoid I structure. (**c**) Product **2**-bound structure. The left image shows the overall fold, the middle panel shows the active-site view, and the right panel shows microspectrophotometer UV-Vis spectra obtained from the crystal used for X-ray crystallographic data collection and the corresponding molecules built into the active site. The PLP cofactor is shown with magenta carbons.

2.3.1.1 Holo-enzyme structure

The holo-enzyme structure was obtained by soaking with 2 mM PLP for 2 h and the structure was solved to 2.0 Å resolution (**Figure 2.3a**). This structure was a homo-dimer with two molecules in the asymmetric unit. The overall fold of this structure matches the hydroxylases and other fold-type I PLP-enzymes[195,196] with root-mean-square deviations (RMSDs) of 1.184 Å (523 atoms) when compared to MppP-5DJ1[114] and 1.474 Å (622 atoms)

when compared to RohP-6C3B[116] using Cα alignments (**Table 2.6**). Each monomer in the dimer contained three domains: a large domain (residues Asp31-Phe270), a small domain (residues Gly271-Lys381) and the N-terminal domain (residues Met1-His30). The large domain consists of 6 parallel β-strands and 1 antiparallel β-strand and 10 α-helixes arranged into an α-β-α sandwich. The small domain contains two antiparallel β-strands and three α-helices. The N-terminal domain consists of one α-helix and is located at the dimer interface, helping to enclose the active site. Unlike the RohP and MppP holo-enzyme structures, where the N-terminal regions are disordered or pointed completely away from the active site, the N-terminal region of the Plu4 holo-enzyme structure is fully ordered and covers the active site (**Figure 2.4**). The active sites are found between the large and small domains within the same monomer with additional contributions from some residues of the large domain of the other protomer in the dimer.

Table 2.6: RMSD values for Plu4 holo-enzyme, intermediate-bound, and product-bound structures compared with RohP and SwMppP structures. The RohP Quinonoid I value is represented as QI, while Quinonoid II is represented as QII. The SwMppP intermediate was an external aldimine. RMSDs are from Cα alignments and are given in Å with the number of atoms included in each alignment in brackets.

	RohP	**SwMppP**
Holo	1.474 (622 atoms)	1.184 (523 atoms)
Intermediate	QI: 1.456 (621 atoms); QII: 1.428 (657 atoms)	1.12 (529 atoms)
Product	1.476 (667 atoms)	1.153 (571 atoms)

The UV-Vis data collected from the holo-enzyme crystal showed a characteristic 415 nm peak for the Schiff base (**Figure 2.3a**).[197] Therefore, an internal aldimine was built into the holo-enzyme active site attached to Lys219 via a Schiff base. The PLP cofactor is also held in place by other residues characteristic of the fold-type I PLP-enzymes. Asp187 interacts with the nitrogen group of the PLP-pyridine ring, modulating the electrophilic properties of the PLP cofactor (**Figure 2.5b**).[196] Asn159, found at the C-terminal end of the 4th β-strand, interacts with the phenolic oxygen and is common among PLP-dependent aminotransferases, but is not invariant.[198] Additionally, the phosphate group of PLP interacts with the Ala87 backbone nitrogen and the side chains of Ser88, Lys227, and His27 (**Figure 2.5b**). While the identity and conformation of most of the active site residues are identical between the RohP, SwMppP, and Plu4 holo-enzyme structures, some residues are in slightly different conformations. His27 in

Plu4 matches the conformation of His34 in RohP, but His29 in SwMppP has two observed conformations (**Figure 2.5c**). Phe112 in Plu4 appears in a similar conformation to Phe115 in MppP, but the corresponding residue (Phe119) has rotated ~120° in RohP (**Figure 2.5d**). As a result of this rotation, the conformations of Asn121 and Leu122 in RohP have also shifted in comparison to the corresponding Asn114 and Leu115 in Plu4. Finally, Phe190 in Plu4 is rotated ~90° compared to Phe191 and Phe201 in MppP and RohP, respectively (**Figure 2.5e**).

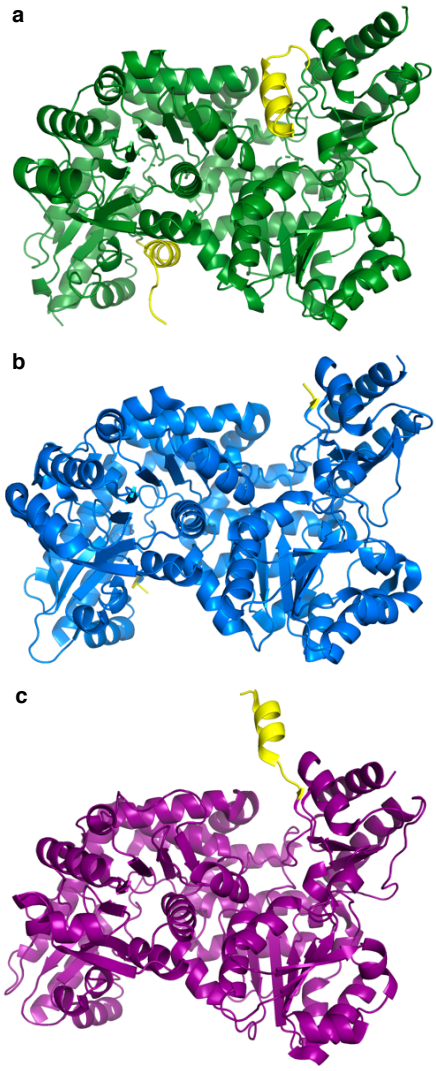


Figure 2.4: Comparison of holo-enzyme overall structures for (a) Plu4, (b) RohP (PDB: 6C3B), and (c) SwMppP (PDB: 5DJ1). The N-terminal regions, if present, are shown in yellow.

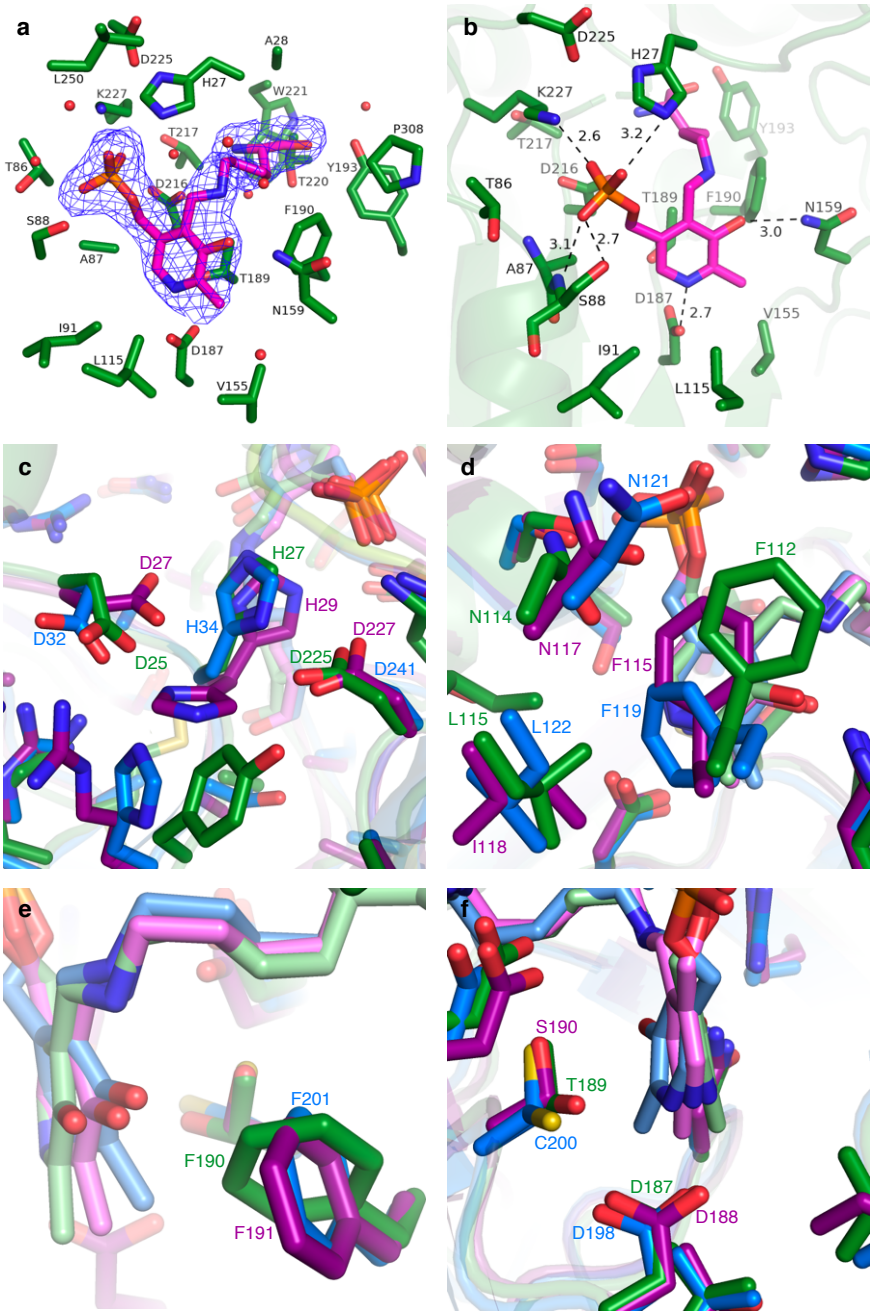


Figure 2.5: Plu4 Holo-enzyme active site analysis. (**a**) F_o-F_c omit map is shown for the PLP cofactor contoured to 3σ. The PLP internal aldimine cofactor is shown in magenta. (**b**) Polar interactions between the enzyme and the cofactor are indicated with dashed lines and the corresponding distance in Å. The cofactor is shown in magenta. (**c**), (**d**), (**e**) and (**f**) shows a comparison of Plu4, RohP (PDB: 6C3B; blue) and SwMppP (PDB: 5DJ1; plum) active site residues, with cofactors shown in light green for Plu4, light blue for RohP and pink for SwMppP. Residue identities and numbers are given in the colour of the corresponding structure. Structures were aligned using the superpose function in PyMOL using Cα pairs.

2.3.1.2 Intermediate-bound structure

An intermediate-bound structure of Plu4 was obtained to 1.9 Å resolution by soaking a crystal with 3 mM of **1** and 2 mM of PLP for 8 minutes, during which the crystal changed from yellow to orange. This structure, similar to the holo-enzyme structure, was a homodimer but it consisted of four molecules in the asymmetric unit, rather than two. The UV-Vis data obtained from the microspectrophotometer measurement of the orange crystal showed a major peak absorbance at λ_{max} 510 nm and a minor peak at λ_{max} 578 nm (**Figure 2.3b**). These peaks were similar to UV-Vis data obtained for Ind4, where a peak at λ_{max} 511 nm was assigned as Quinonoid I and a peak at λ_{max} 567 nm was assigned as Quinonoid II during stopped-flow experiments.[113] When refinement was nearly complete, including placement of water molecules, five possible intermediates were built into the active site density (**Figure 2.6**). Although the presence of major and minor peaks in the UV-Vis microspectrophotometer data indicates that there is likely a mixture of intermediates present in the intermediate-bound structure, Quinonoid I was chosen as the ligand for the final structure based on the dominant UV-Vis signal observed at 510 nm.

While fold-type I PLP-dependent enzymes can undergo large structural changes upon ligand binding,[199,200] no major changes to the domains were observed between the holo-enzyme structure and the intermediate-bound structure. Unlike RohP and SwMppP, whose N-terminal helices become ordered upon intermediate formation, the N-terminus of Plu4 did not change conformation. Despite the N-termini appearing in similar, more ordered states, for the RohP, SwMppP and Plu4 intermediate-bound structures, Plu4 remains distinguished from the hydroxylases by having lower B-factor values assigned to the atoms in its N-terminus (**Figure 2.7**). The lower B-factor values indicate that there could be less N-terminal movement in Plu4 than in RohP or SwMppP. However, differences in the N-terminal helix could potentially be the result of crystallographic contacts between asymmetric units because the area surrounding the RohP and SwMppP N-terminal helices is more open than for Plu4, likely contributing to a more ordered N-terminal helix in Plu4 (**Figure 2.7**).

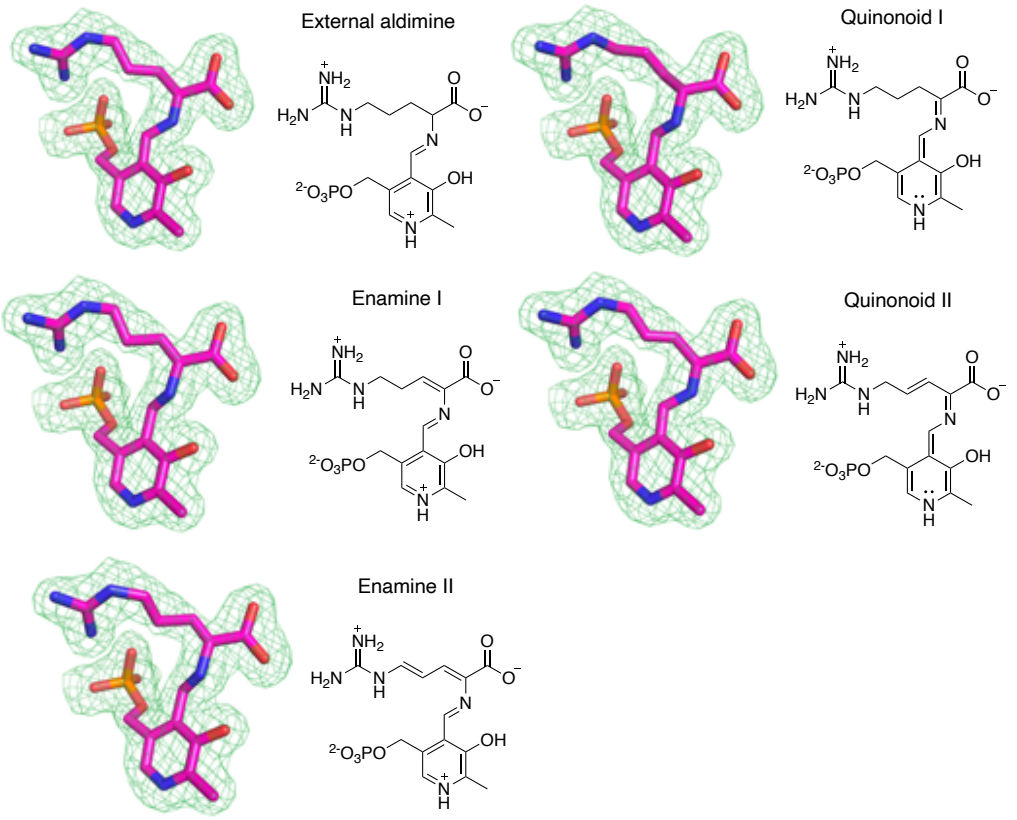


Figure 2.6: Comparison of possible intermediates modeled into the active site density for the intermediate-bound structure. F_o-F_c electron density maps are shown in green and contoured to 3σ.

Many of the contacts between the internal aldimine of the holo-enzyme structure and the Quinonoid I ligand of the intermediate-bound structure remain similar; however, the PLP cofactor rotates away from Lys219 by 23.8° (**Figure 2.8a** and **b**), anchored by the heterocyclic nitrogen and phosphate group. Additionally, the bond between carbon and oxygen from the phosphate group rotates to accommodate this movement (**Figure 2.8a**). As the Quinonoid I intermediate is formed, Lys219 is released from the imine bond and forms a polar interaction with His27, Asp216 and the PLP phosphate group, facilitated by a water molecule (**Figure 2.8c**). With the addition of substrate, several new polar contacts are formed between the Quinonoid I intermediate and the enzyme. The anionic carboxylate group from Quinonoid I forms an ionic interaction with the positively charged Arg354 guanidinium group, which is a conserved residue in the PLP aminotransferases[198]. The carboxylate group from Quinonoid I also forms a complex hydrogen bonding network between Asn159 and Glu13 (**Figure 2.8d**). The guanidinium portion

of the Quinonoid I ligand interacts with a phosphate oxygen atom from PLP and the Ser88 side chain as well as with Thr10, found in the N-terminal helix and backbone oxygens from Glu246 and Glu247, found in the other protomer of the dimer (**Figure 2.8e**). Additionally, both His27 and Glu13 are both within 4 Å of C4 and C5 of the Quinonoid I intermediate and could serve as the catalytic residues responsible for deprotonating these carbons during catalysis (**Figure 2.8f**). His27 also has several nearby residues, including Asp225, Tyr29, and Cys251 that could be aiding His27 in any potential catalytic role (**Figure 2.8f** and **Figure 2.9b**). All other residues surrounding the active site appear to be unchanged in response to formation of Quinonoid I (**Figure 2.8a**).

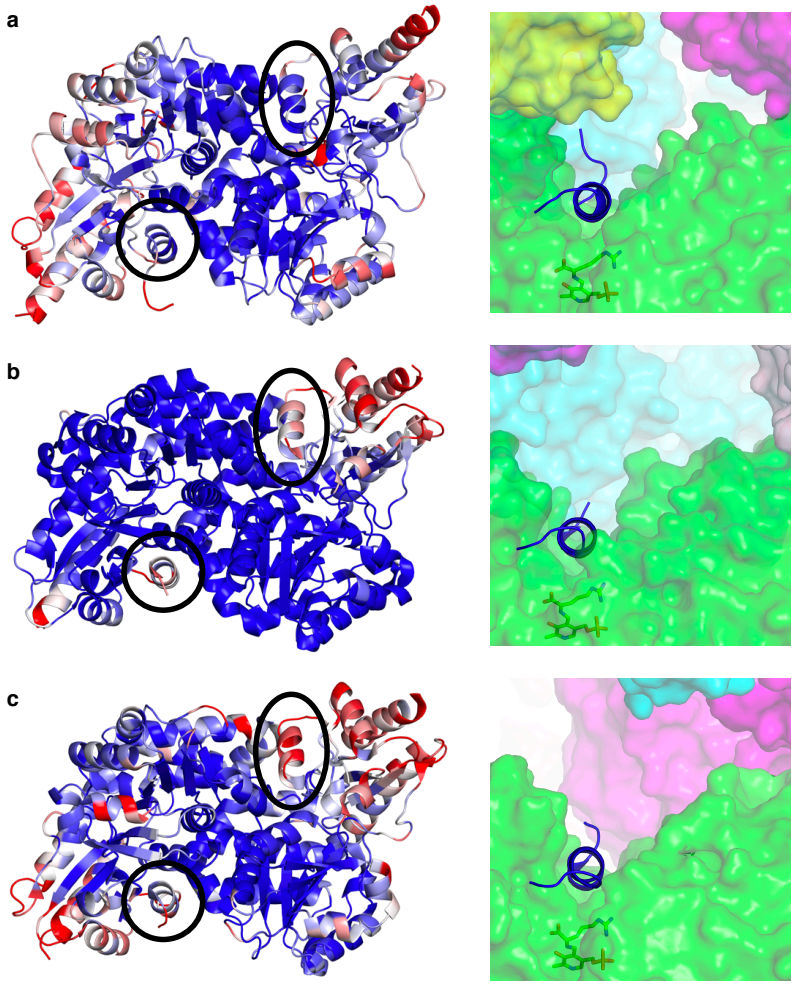


Figure 2.7: Comparison of N-termini in intermediate-bound structures, (**a**) Plu4, (**b**) RohP (PDB: 6C3D), and (**c**) SwMppP (PDB: 6C8T). The left panel shows B-factors for the respective structures. B-factors are indicated by a spectrum of blue, white and red with blue having the lowest B-factors (below 20 Å^2) and red having the highest B-factors (above 50 Å^2). The N-terminal helices of both monomers are indicated with a black circle. The right panel shows N-termini (blue) for the respective structure (green) with surrounding asymmetric units (cyan/magenta/yellow/lightpink). Differences to the B-factors of the N-terminal helices could be due to crystal contacts (shown in the right panel) and would require a further comparison of the molecular dynamics of each protein to elucidate the role of the N-termini.

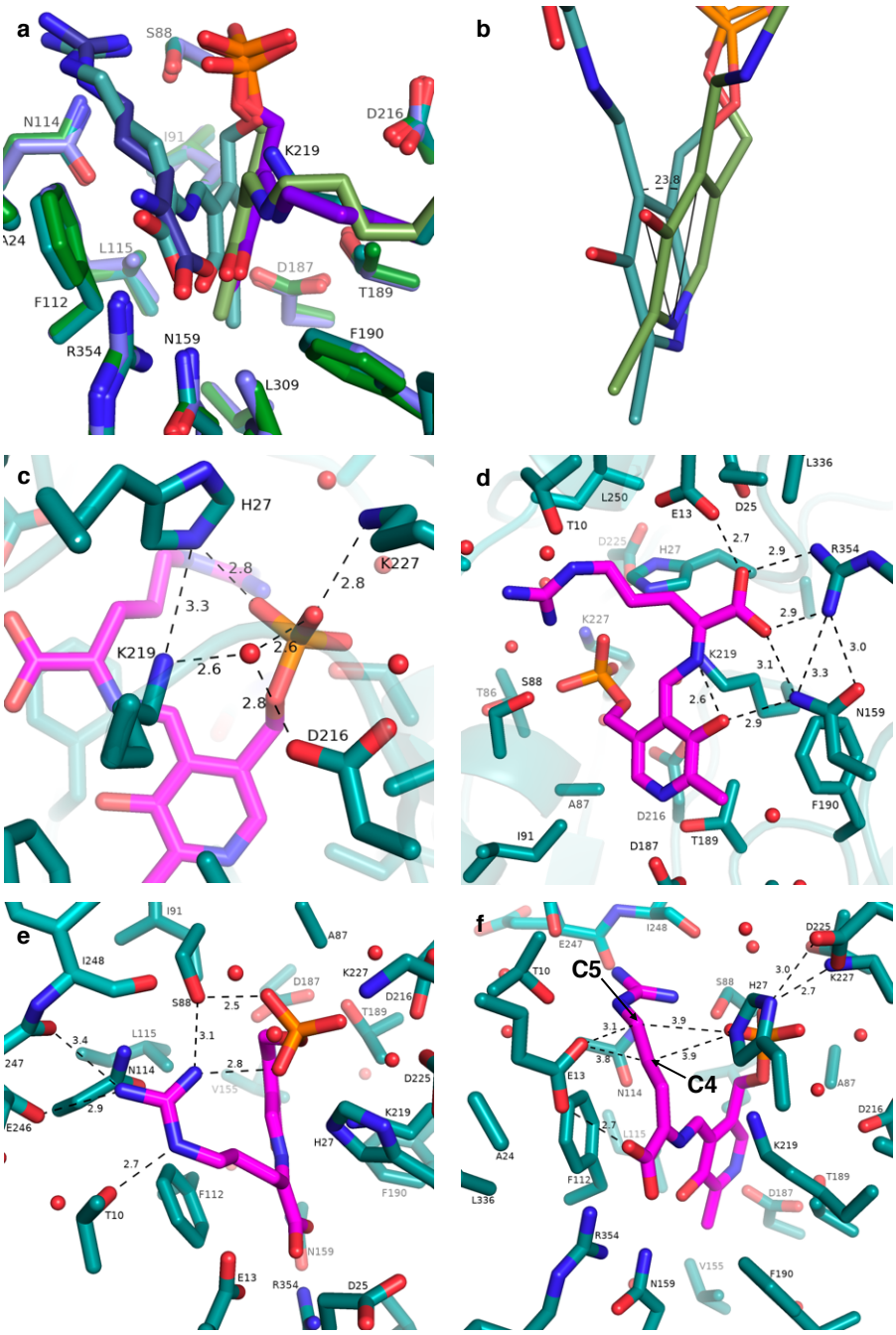


Figure 2.8: Plu4 intermediate ligand interactions with protein in the intermediate-bound structure (**a**) Comparison of holo (green), intermediate (teal) and product (purple) structures for Plu4 with their respective ligands/cofactors shown in light green, light teal, and deep purple. (**b**) Comparison of holo internal aldimine (light green) and intermediate quinonoid I (light teal) structures with angle of movement given in degrees. (**c**) Interactions between Lys219 and surrounding His27, water, Asp216, Lys227 and the PLP phosphate group. (**d**) and (**e**) show polar contacts between the intermediate and the protein in Å. (**f**) Distances between carbons 4 and 5 and nearby His27 and Glu13 residues. Polar distances between His27 and Asp225 in the catalytic dyad are also shown.

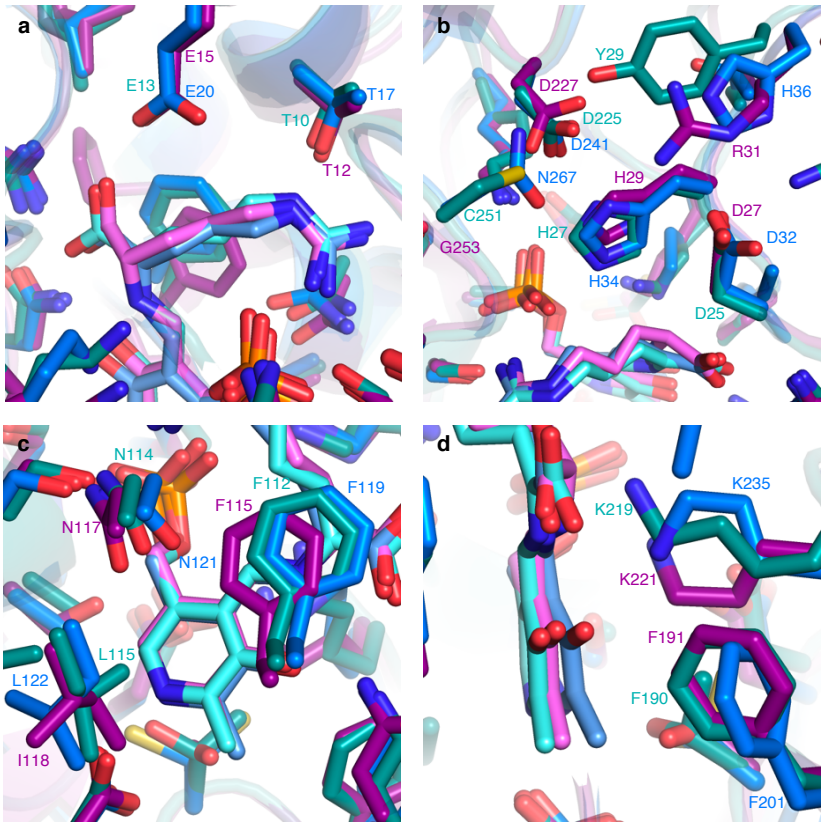


Figure 2.9: Comparison of Plu4, RohP (blue; PDB: 6C3D) and SwMppP (plum; 6C8T) intermediate-bound active site structures. The intermediate ligands are show in cyan, light blue and pink, respectively for Plu4, RohP and SwMppP. Residue identities and numbers are given in the colour of the corresponding structure. Structures were aligned using the superpose function in PyMOL using Cα pairs from Chain A.

2.3.1.3 Product-bound structure[3]

The product-bound structure of Plu4 was determined to 2.5 Å resolution by soaking a Plu4 crystal with 2 mM PLP and 2 mM **1** for 16 h, during which the crystal changed from yellow to orange and back to yellow. There were four molecules in the asymmetric unit, each forming half of a homodimer. The UV-Vis spectra obtained from this crystal had a peak at 413 nm, indicating the presence of an internal aldimine, which was subsequently built into the structure (**Figure 2.3c**). There were no large structural changes observed in this structure compared to the holo-structure or the intermediate structure. However, one or two segments of an external loop (Asp302-Glu303 in Chain A and Gly293-Thr294 and Asp302-Glu303 in Chain B) were

[3] DFT calculations in this section were completed by M.G.-B.

disordered in the product-bound structure but were ordered in the holo-enzyme and intermediate-bound structures.

There were two active sites that displayed electron density corresponding to a product, one in each dimer in the asymmetric unit. However, before modelling any products in this density, DFT model calculations were performed to determine which tautomeric form of the four-electron oxidized product was the most stable: the enamine or the imine structure. Surprisingly, the imine form was calculated to be the most stable, despite the enamine form being more conjugated (**Figure 2.10**). The imine form is likely to be more stable than the enamine because the distal ends of the conjugated system have highly electrophilic functional groups (guanidine and carboxylate), which may pull electron density away from the middle of the molecule, allowing the conjugation to break and the imine to form. Additionally, when bound to the enzyme, residues surrounding the guanidine and carboxylate may be acting to increase the electrophilicity of these electrophilic functional groups. Therefore, the imine form of **2** (**Figure 2.1**) was modelled into Chain A, where it fit well into the density, but **4** (**Figure 2.1**) was modelled into the active site of Chain C because it fit the electron density better (**Figure 2.11**). Although 2-iminoarginine is expected to be a product of Plu4 (**Figure 2.1**), its production has not been confirmed, so **4** was modelled instead.

kcal/mol	ΔE	ΔG
imine	0.0	0.0
enamine	3.0	3.6

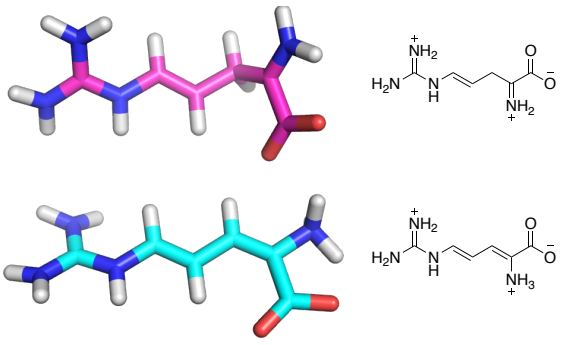


Figure 2.10: DFT calculations for four electron oxidized product in imine and enamine forms.

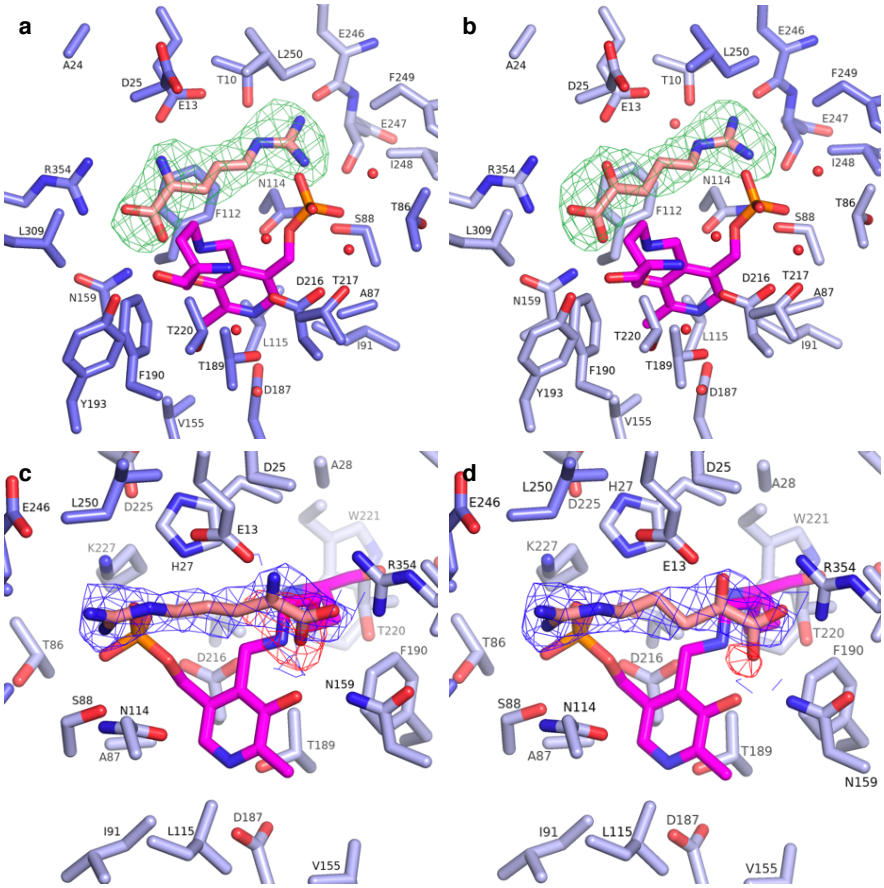


Figure 2.11: Electron density for products **2** and **4** in the product-bound structure. (**a**) F_o-F_c omit map density for **2** (salmon) in Chain A of the product bound structure. (**b**) F_o-F_c omit map density for **4** (salmon) in Chain C of the product bound structure. (**c**) $2F_o$-F_c and F_o-F_c of **2** when fit into the Chain C active site. (**d**) $2F_o$-F_c and F_o-F_c of **4** when fit into Chain C active site. F_o-F_c maps are contoured to 3σ and $2F_o$-F_c maps are contoured to 1σ.

Each product was modelled into the active site with the guanidine group adjacent to the phosphate group of PLP, while the carboxylate group formed a salt bridge with Arg354 (**Figure 2.11** and **Figure 2.12**), similar to what was observed in the RohP and SwMppP product-bound structures. However, the imine of **5** (**Figure 2.1**) and the ketone of **4** (**Figure 2.1**) were pointed toward the carboxylate group of Glu13 rather than towards the 4′-carbon of PLP. The latter was the expected conformation of the products considering that the intermediate ligand requires that the imine nitrogen be bonded to the 4′-carbon of PLP prior to reformation of the internal aldimine. The RohP and SwMppP product-bound structures both showed the ketone group of **3** pointed towards the 4′-carbon of PLP in the expected conformation (**Figure 2.12c** and **Figure 2.13a**). To determine if this unexpected binding conformation observed in the Plu4 product-

bound structure is reasonable, docking experiments were done to determine energetic feasibility of the observed binding conformations for both modelled products, **4** and **5**. The docking studies determined that both binding conformations, with the imine/ketone pointing either toward Glu13 or the 4′-carbon of PLP are energetically feasible (**Figure 2.12**).

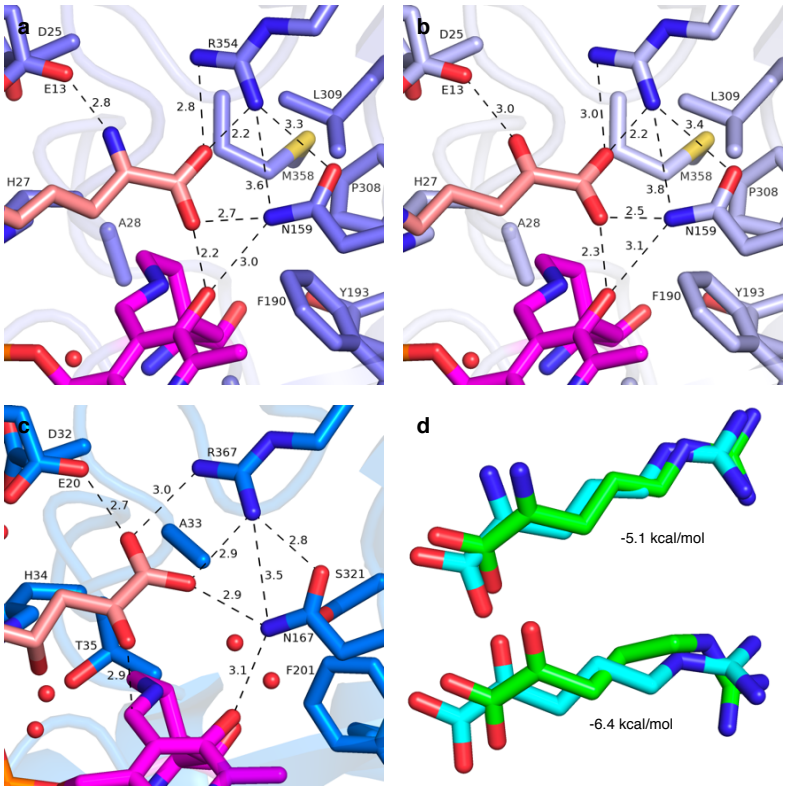


Figure 2.12: Analysis of product binding conformations. (**a**) Polar interactions of the carboxylate and imino groups of **2** (salmon) in Chain A. (**b**) Polar interactions of the carboxylate and ketone groups of **4** (salmon) in Chain C. (**c**) Polar interactions between the carboxylate group of **3** (salmon) and active site residues in RohP (PDB: 6C3A). The distance between the ketone of **3** and 4′-carbon of the PLP cofactor is also shown. (**d**) Docking conformations (green) closely matching observed conformations in the Plu4 product-bound crystal structure (cyan) for **2** (top) and **4** (bottom) with binding energy indicated below.

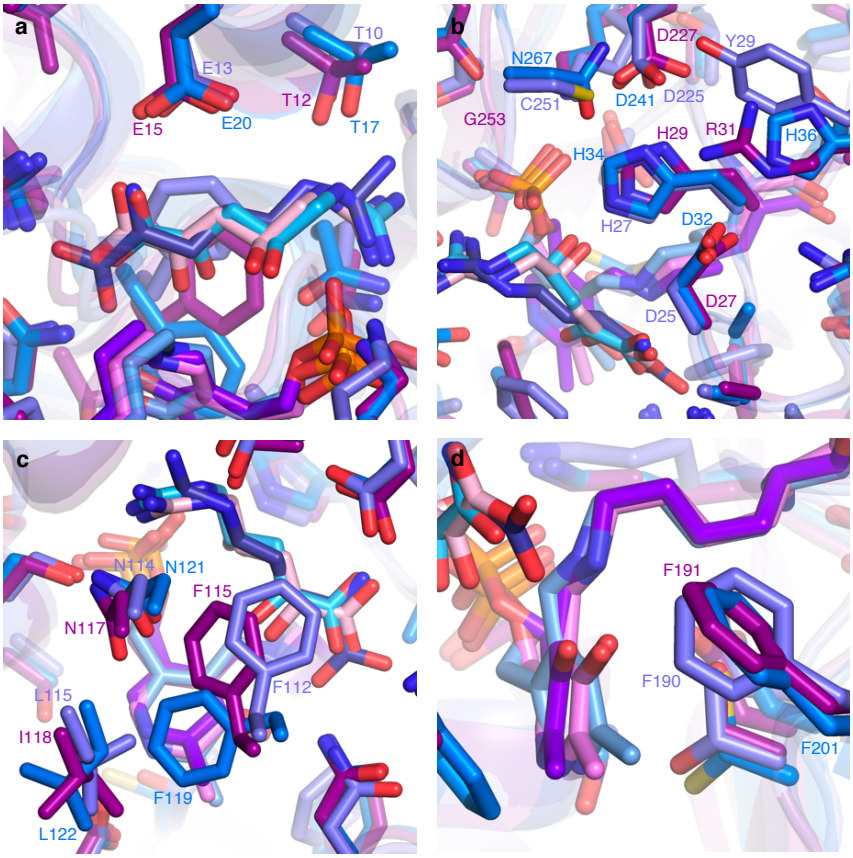


Figure 2.13: Comparison of Plu4, RohP (blue; PDB: 6C3A) and SwMppP (plum; 6C9B) product-bound active site structures. The product ligands and PLP are shown in deep purple, light blue and pink, respectively for Plu4, RohP and SwMppP. Residue identities and numbers are given in the colour of the corresponding structure. Structures were aligned using the superpose function in PyMOL using Cα pairs.

2.3.2 Site-directed mutagenesis

Guided by crystallographic studies of Plu4 and previous mutagenesis work on the PLP-dependent arginine oxidases, it was hypothesized that His27 and Asp225 in the active site could play a role in the second oxidative step. Both the His27 and Asp225 residues are highly conserved in the PLP-dependent arginine oxidases (**Figure 2.14)** and the His residue has been implicated previously, through mutagenesis, as a catalytic residue for installing the 4-hydroxyl group of **3** (**Figure 2.1**) in RohP.[116] To test the effect that conserved His and Asp residues may have on catalysis in Plu4, His27 and nearby conserved residues Asp225 and Asp25 were mutated. In the H27A, D225A, D225N, and D25A variants, the first oxidation product **4 (Figure 2.1)** is still made, but the second oxidation product **5** is not (**Table 2.7**). D25N had the same

product profile as the wild-type enzyme (**Table 2.7**), suggesting that Asp25 does not have an essential role in catalysis.

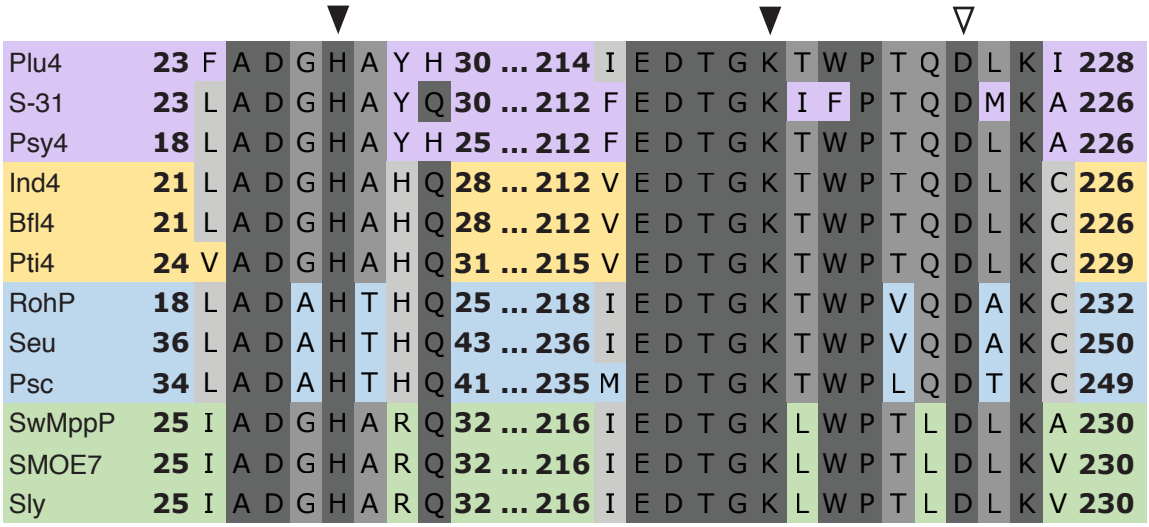


Figure 2.14: Alignment of sequences from characterized SSN groups (I-III). Group Ia is shown in blue, Group Ib is shown in yellow, Group II is shown in green, and Group III is shown in purple. Known catalytic residues are indicated with a filled black arrow, while Asp225, which participates in a catalytic dyad, is shown with a hollow black arrow. Sequences shown are abbreviated as follows: Plu4 represents WP_039608190.1 from *Pseudoalteromonas luteoviolacea*, S-31 represents WP_030739174.1 from *Streptomyces* sp. NRRL S-31, Psy4 represents SDZ12441.1 from *Pseudomonas syringae* strain BS2900, Ind4 represents AJT38685.1 from *Streptomyces griseus* subsp. *griseus*, Bfl4 represents WP_061786414.1 from *Bacillus flexus*, Pti4 represents WP_143006797.1 from *Paenibacillus tianmuensis*, RohP represents WP_041824998.1 from *Streptomyces cattleya*, Seu represents WP_102919045.1 from *Streptomyces eurocidicus*, Psc represents WP_046237671.1 from *Pseudomonas syringae* pv. *Coryli*, SwMppP represents the sequence from PDB ID: 5DJ3_A from *Streptomyces wadayamensis*, SMOE7 represents ARH92424.1 from *Streptomyces* sp. MOE7, and Sly represents AOP50599.1 from *Streptomyces lydicus*.

Table 2.7: Product profiles of enzyme variants based on LC-MS data.

Variant	Enzyme	Products Made
WT	Plu4	**4** and **5**
H27A	Plu4	**4**
D225A	Plu4	**4**
D225N	Plu4	**4**
D25A	Plu4	**4**
D25N	Plu4	**4** and **5**
Y29H	Plu4	**4** and **5**
Y29F	Plu4	none

Variant	Enzyme	Products Made
Y29R	Plu4	**4** and **5**
C251A	Plu4	**4** and **5**
C251G	Plu4	none
C251N	Plu4	**4** and **5**
T189C	Plu4	**4** and **5**
T189S	Plu4	**4** and **5**
T186C	Ind4	none
G26A	Plu4	none
G24A	Ind4	none
A28T	Plu4	none
A26T	Ind4	none
L117D	Plu4	none

Next, residues within 5 Å of the active site in Plu4 that differ between the hydroxylases and the desaturases were targeted to try to find the cause of the different reactivities (**Figure 2.15**). These residues include Thr189, Gly26, and Ala28, each of which has a different identity to Plu4 in both RohP and SwMppP. The corresponding mutations were also made in Ind4 to match the RohP identity. Substitution of G26A and A28T in Plu4 showed complete loss of both products, but T189C and T189S in Plu4 gave no change in the product profile (**Table 2.7**). However, all mutations at these positions in Ind4 resulted in loss of both products (**Table 2.7**). Two non-conserved residues, Tyr29 and Cys251, near Asp225 were also targeted to see if they changed the product profile, but the resulting variants either abolished production of both **4** and **5** or did not affection production (**Table 2.7**). These residues seem likely involved in maintaining structure around the catalytic residues, rather than determining hydroxylase and desaturase reactivity. Leu117, which is in a hydrophobic tunnel leading to the active site in the Plu4 structure, was also targeted. This tunnel is far less hydrophobic in the RohP structure, seemingly as a result of an aspartate residue at the position of Leu117. However, no products were observed from this variant (**Table 2.7**).

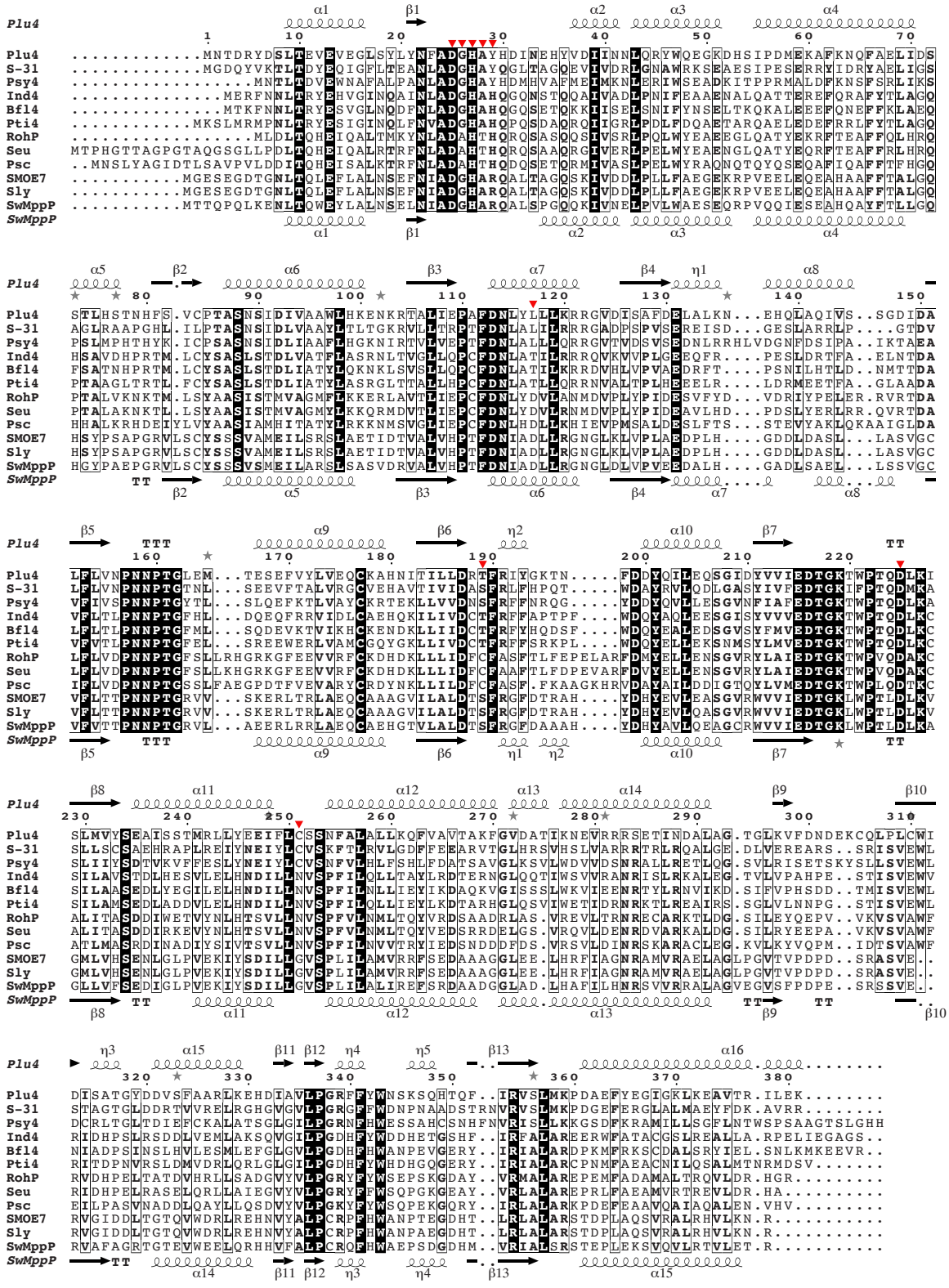


Figure 2.15: Full sequence alignment of representative arginine oxidases. Secondary structural elements are shown above and below for Plu4 and SwMppP, respectively. Residues that were targeted for mutagenesis are highlighted with red arrows. Descriptions of each abbreviated sequence name is given in **Table 2.7**.

2.4 Discussion

These results show, for the first time, the X-ray crystal structure of an arginine desaturase. Through these crystal structures, catalytic residues His27 and Asp225 were identified as a catalytic dyad and were demonstrated to have a role in forming **5**. While the corresponding His residue has been previously implicated to have a role in forming **3** in the hydroxylases,[115,116] the Asp residue, which is also conserved across the arginine oxidases, has not previously been reported for its role in catalysis. The highly conserved nature of the hydroxylase and desaturase active sites shows that the key to distinguishing the two reactivities may be more complex than a mutation of a single active site residue. Additionally, the mutagenesis work described here did not identify any residues that were responsible for distinguishing hydroxylase and desaturase reactivity. Therefore, the key to distinguishing these activities may lie within other aspects of the protein's structure.

There were several residues surrounding the active site of Plu4 that were observed in various conformations in both MppP and RohP structures, but do not appear in different conformations for Plu4. The N-terminal region of the hydroxylases appears more dynamic than in Plu4, demonstrated by lower B-factors in Plu4 and lack of ordering in the hydroxylases without substrate bound (**Figure 2.4** and **Figure 2.7**). Residues Phe119, Asn121, Leu122 appear to change conformations in the different RohP crystal structures, but the corresponding residues do not appear to move in Plu4 (**Figure 2.5d**, **Figure 2.9c** and **Figure 2.13c**). Phe191 rotates 90° in MppP between the internal aldimine and external aldimine but remains in a single conformation in Plu4 (**Figure 2.5e**, **Figure 2.9d** and **Figure 2.13d**). Reduced movement of these residues suggests increased rigidity of the desaturase active sites, which could differentiate the reactions. The conformations of the hydrophobic residues in particular could prevent water from reaching the active site in the desaturases, making the movement and/or position of water in the active site a potential way of distinguishing reactivity.

The oxygen reactivity mechanism within the arginine oxidases remains highly elusive as the crystal structures reported here have not shown any direct evidence of a peroxy-intermediate. However, in flavin-dependent enzymes and cofactor-independent enzymes, positively charged species in the active site are thought to contribute to oxygen activation by stabilizing the negatively-charged peroxy-intermediates formed during the reactions.[120,121] In Plu4 and perhaps other arginine oxidases, Lys219, Lys227, or the arginine substrate itself could be responsible for

providing the positive environment needed to stabilize the formation of a negatively-charged peroxy-intermediate. There is some evidence that such an environment exists in Plu4 by analysis of the electrostatic map of the active site surrounding the quinonoid intermediate (**Figure 2.16**).

Overall, this work helps to decipher the mechanism of arginine oxidases by providing structural insights for the arginine desaturases and identifying catalytic residues for the first time. A thorough knowledge of these enzymes and their mechanism helps to expand the known biocatalytic repertoire by exploring a new type of enzymatic reaction. Should this new reaction become an attractive tool to synthetic chemists, this work will aid future efforts in enzyme engineering towards a useful biocatalyst by providing a more in depth understanding of the natural enzyme. Complementary to the work described here, product analyses of Plu4 and other Ind4 homologs will be described in **Chapter 3**. Further work to explore the mechanism and other non-structural characteristics of the arginine oxidases will also be analyzed in **Chapter 3**.

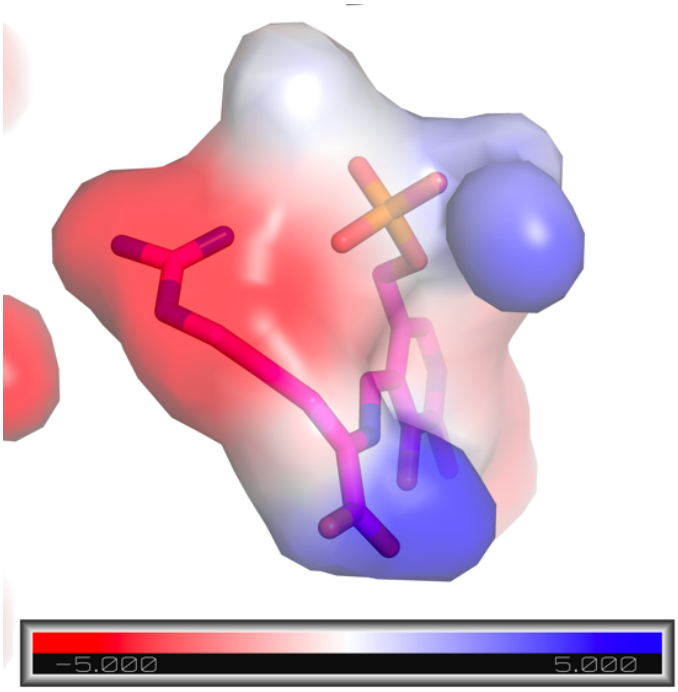


Figure 2.16: Electrostatic map of the Plu4 active site pocket surrounding Quinonoid I in the intermediate-bound structure, coloured according to surface potential as indicated in the figure legend with values given in kT/e. The positively charged pocket near the phosphate group is formed by Lys219 and Lys227. Map was created using the APBS Electrostatics Plugin in PyMOL.[236]

Chapter 3: Understanding the evolutionary relationship and bioactivity of arginine desaturases[4]

3.1 Introduction

In **Chapter 2**, a deeper understanding of the arginine oxidases was accomplished by obtaining the first crystal structure of an oxygen- and PLP-dependent arginine desaturase, Plu4, after attempting to crystallize a number of predicted homologs of the first known arginine desaturase, Ind4.[113] Despite being a predicted Ind4 homolog, Plu4 appears from a convergent indolmycin pathway from *P. luteoviolacea*.[68] Plu4's sequence identity to Ind4 (32%) is also much lower than that of other homologs to Ind4 (>54%) and is more similar to the sequence identity between the hydroxylases and Ind4 (37-41%). Therefore, it is important to further characterize Plu4 and ensure that it behaves similarly to Ind4 through biochemical characterization. Also, despite this new structural information, no active site residues were implicated to distinguish the hydroxylation and desaturation activities in the arginine oxidases, making the mechanism of differentiation still highly elusive. Therefore, other characteristics of these enzymes must be explored to gain a better insight into their mechanism.

An aspect of all arginine oxidases that remains to be understood is their oxygen reactivity. In the hydroxylases, the hydroxyl group is derived from water, which makes the role of molecular oxygen elusive as no atoms from the co-substrate are incorporated into the final product for either the hydroxylases or desaturases. The current understanding of arginine oxidase reactivity with oxygen comes from a comparison with flavo-proteins. The nature of oxygen reactivity in the flavo-proteins has been debated and widely investigated, but they are thought to catalyze a single electron transfer to oxygen, forming superoxide and a semi-quinone as a caged radical pair, which can have a variable fate depending on the enzyme.[120,121] While superoxide peroxy-intermediates have been proposed as reaction intermediates in the arginine oxidases, their existence and fate during catalysis have not been explored in these enzymes.

In this chapter, the sequence similarity relationships amongst the arginine oxidases are analyzed to better understand potential evolutionary origins and be able to predict hydroxylase or desaturase activity in uncharacterized enzymes. This knowledge will be helpful to identify new

[4] A version of this chapter, in combination with Chapter 2, has been submitted for publication. Some of the experiments and calculations described in this chapter was completed by collaborators.

potential biocatalyst candidates and aid in any future efforts to understand how these reactivities evolved. The biochemical activities of the desaturase homologs used in crystallographic studies are also characterized to confirm their identities as desaturases. Plu4 was investigated further with a more in-depth look at its kinetics, side reactions and oxygen reactivity. Then, armed with the crystallographic and mutagenesis data described in **Chapter 2**, a unified mechanism for the arginine hydroxylases and desaturases is proposed that accounts for the new information that will be described.

3.2 Materials and methods

3.2.1 General methods, materials and strains

Reagents, columns and primers were purchased from the same sources as listed in **Chapter 2**, except for the use of materials from Hampton Research. Sequencing was carried out as described in **Chapter 2.** A list of strains and plasmids used in this study is given in **Table 3.1**.

Table 3.1: Bacterial strains and plasmids used in this study.

Strain/Vector	**Description**	**Source**
Strain		
E. coli DH5*α*	General cloning host	Laboratory stock
E. coli BL21 (DE3)	Host for protein expression	Laboratory stock
Plasmid		
pET22b	Vector for protein expression in *E. coli*	Laboratory stock
pET28a	Vector for protein expression in *E. coli*	Laboratory stock
pET22b-ind4	Vector for Ind4 expression, cloning sites NdeI/XhoI	Du *et al.*[48]
pET22b-pel4	Vector for Pel4 expression, cloning sites NdeI/XhoI	Du *et al.*[48]
pET28a-swmppP	Vector for SwMppP expression, cloning sites NdeI/XhoI	This study
pET22b-plu4	Vector for Plu4 expression, cloning sites NdeI/XhoI	This study
pET22b-bfl4	Vector for Bfl4 expression, cloning sites NdeI/XhoI	This study
pET22b-pehi4	Vector for Pehi4 expression, cloning sites NdeI/XhoI	This study
pET22b-punc4	Vector for PUNC4 expression, cloning sites NdeI/XhoI	This study

Strain/Vector	Description	Source
pET22b-bsp4	Vector for Bsp4 expression, cloning sites NdeI/XhoI	This study
pET22b-pti4	Vector for Pti4 expression, cloning sites NdeI/XhoI	This study
pET22b-f5135	Vector for F5135 expression, cloning sites NdeI/XhoI	This study
pET22b-mst1	Vector for P_Mst1 expression, cloning sites NdeI/XhoI	This study
pET22b-m63	Vector for P_M63 expression, cloning sites NdeI/XhoI	This study

3.2.2 Network analysis, sequence alignment, phylogenetic trees and homolog identification

Sequence similarity networks (SSNs) were made using Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST)[201–204] from the top 1000 hits in an Ind4 search of the UniprotKB database with an initial search threshold of 10^{-30}. The threshold was increased to 10^{-65} and 10^{-85} to observe separation of nodes with different reactivities. Genomic neighbourhood analysis was done using Enzyme Function Initiative-Genomic Neighbourhood Tool (EFI-GNT).[201–204] Networks were analyzed using Cytoscape.[205] A sequence alignment was made using Muscle in UGENE.[175] Phylogenetic trees were made from the alignment using IQtree[206–208] using the top 1000 BLASTp hits from Ind4. LG+F+I+G4 was used as a substitution model and the resulting tree was visualized using Interactive Tree of Life.[209] Homologs were identified from a BLASTp search of Ind4 by having sequence similarity above 50% or were from known indolmycin producers. These homologs were cross referenced with the UniprotKB database to determine their location in the SSN. Homologs chosen for characterization had between 50%-95% sequence identity to Ind4 and had less than 96% sequence identity to other chosen homologs. Ten homologs of Ind4 (**Table 3.2**) fit this criteria and attempts were made to further clone, purify and characterize them.

Table 3.2: Enzymes used in this study, including arginine desaturase library and hydroxylase control. Homologs were chosen from a BLASTp search of Ind4 (March 2017). Enzymes with unknown activity were not tested for activity.

Homolog	Species	Activity	Accession Number	Size (aa)	Identity (%)	Extinction coefficient ($M^{-1}cm^{-1}$)
Ind4	*Streptomyces griseus* ATCC 12648	Desaturase	AJT38685.1	383	100	43,430
Pel4	*Paenibacillus elgii* B69	Desaturase	WP_139074851.1	382	59	46,410
Plu4	*Pseudoalteromonas luteoviolacea* strain HI1	Desaturase	WP_039608190.1	381	32	51,340
Bfl4	*Bacillus flexus* NBRC 15715	Desaturase	WP_061786414.1	383	54	45,380
Bsp4	*Bacillus sp*. 5mfcol3.1	Desaturase	WP_090977143.1	384	56	47,330
Pti4	*Paenibacillus tianmuensis* strain CGMCC 1.8946	Desaturase	WP_143006797.1	382	59	47,900
F5135	*Streptomyces* sp. NRRL F-5135	Desaturase	WP_051765546.1	386	95	44,920
Pehi4	*Paenibacillus ehimensis* A2	Desaturase	WP_152547756.1	382	59	71,390
P_Mst1	*Paenibacillus tyrfis* strain MSt1	Unknown	WP_036690068.1	382	59	46,410
P_M63	*Paenibacillus elgii* strain M63	Unknown	KZE74284.1	378	59	40,910
Punc4	*Paenibacillus* sp. UNC496MF	Desaturase	SFI40191.1	382	58	49,390
SwMppP	*Streptomyces wadayamensis* strain A23	Hydroxylase	KDR62041.1	376	37	43,430

3.2.3 Cloning

Genes F5135, Pehi4, P_Mst1, P_M63, Pti4, Punc4, Bfl4, and Bsp4 were synthesized by Bio Basic Inc. DNA coding for the chosen homologs were synthesized in the pUC57-Kan vector. The homolog genes were sub-cloned into pET22b using NdeI and XhoI to contain a C-terminal histidine tag. SwMppP,[114,115] which was used as a hydroxylase control, was obtained through gene synthesis in pUC57-Amp and the *swmppP* gene was sub-cloned into pET28a using NdeI and XhoI to contain an N-terminal histidine tag. The resulting vectors were transformed into chemically competent *E. coli* DH5α by heat shocking followed by selection on LB plates

containing ampicillin (100 μg/mL) or kanamycin (50 μg/mL). Colonies were screened by colony PCR with T7 primers (**Table 3.3**). Plasmids from successful colonies were purified using the QIAprep Spin Miniprep Kit from Qiagen and confirmed by sequencing. Successful plasmids were transformed into chemically competent *E. coli* BL21 (DE3) by heat shocking for protein expression. Expression vectors, containing *ind4,*[48] *pel4,*[113] and *plu4* (unpublished) were made previously using pET22b.

Table 3.3: Oligonucleotide primers used in this study.

Primer	Sequence (5′ to 3′)	Description
T7	TAATACGACTCACTATAGGG	T7 promoter for sequencing
T7-term	GCTAGTTATTGCTCAGCGG	T7 terminator for sequencing

3.2.4 Protein purification

Bacterial growth, protein expression, cell lysis and purification were done as described in **Chapter 2**, but the protein was not dialyzed in hydroxylamine to remove all bound PLP and enzymatic products. Instead, fractions collected from the gel filtration column were pooled and directly dialyzed overnight in Buffer D containing 10% glycerol and PLP in five times molar excess of the protein concentration to reconstitute the PLP. The dialyzed samples were concentrated, aliquoted and stored as described in **Chapter 2**. Enzymes P_Mst1 and P_M63 did not express well enough to produce a sufficient amount of material from the Ni-charged column purification step, so work with these proteins was discontinued.

3.2.5 *In vitro* characterization

Each homolog was assayed by ESI-MS, as described in **Chapter 2** for variant characterization. For analysis of products derivatized with *o*-phenylenediamine (OPD), the 100 μL overnight reaction was treated with 100 μL of water and 100 μL of an OPD solution (100 mM OPD in 2 M HCl) for 20 min, at 80 °C, unless otherwise specified. Reactions containing no enzyme, no L-arginine, or boiled enzyme were used as negative controls. The derivatization time was optimized using the SwMppP reaction to reduce the dehydration side reaction and increase detection of **3**. The derivatized reactions were centrifuged to remove denatured protein and 20 μL of the supernatants were analyzed on a 6120 Quadrupole LC-MS system (Agilent) using

a Luna C18(2), 5 μm, 4.6 mm ID × 250 mm column (Phenomenex). Elution was done using a mixed mobile-phase linear gradient of water and acetonitrile ((v/v): 95:5 to 50:50, 0 to 15 minutes; 0:100, 15 to 22 minutes; 95:5 22 to 30 minutes), each containing 0.1% (v/v) formic acid. Detection was done in positive ion mode. The OPD-derivatized reaction products were quantified by high-performance liquid chromatography (HPLC) using a 1260 HPLC apparatus (Agilent) system using the same column and elution conditions listed above for the LC-MS system. The hydroxylation side-product was further confirmed by running the overnight reaction in 50% $H_2{}^{18}O$. Derivatization was done over 2 min at 80 °C to reduce dehydration of derivatized 4-hydroxy-2-ketoarginine and allow for detection of the ^{18}O labelled product.

Plu4 kinetic assays were performed by monitoring the consumption of O_2 using a Clark-type polarographic electrode OXYG1 (Hansatech).[113,116] The electrode was calibrated daily according to manufacturer's instructions using air-saturated water and sodium hydrosulfite. The standard reaction was performed in 1 mL of air-saturated 40 mM TAPS buffer (I = 0.1 M, pH 8.5) at 25°C; buffer and L-arginine mixtures were pre-equilibrated in the reaction chamber prior to reaction initiation by the addition of 2.5 μM Plu4. Reaction rates were measured from the slope at maximum velocity, averaged over at least one minute, and corrected by subtracting the average background oxygen consumption rates prior to reaction initiation. The effect of pH on the rate of Plu4-catalyzed reaction was determined by using air-saturated 20 mM buffers of HEPES (pH 7.5), Tris-HCl (pH 8.0, 8.5 and 9.0), CHES (pH 9.4), and CAPS (pH 10.4). Steady-state kinetic parameters of Plu4 for L-arginine were measured in air-saturated buffer with varying concentrations of L-arginine (5 – 120 μM). The steady-state kinetic parameters for O_2 were evaluated using 150 μM of **1** and initial concentrations of O_2 from 14 to 256 μM. The initial O_2 concentrations were achieved by equilibrating the reaction mixture with humidified mixtures of O_2 and N_2 gasses. Final O_2 levels were normalized to the ambient O_2 level prior to the adjustment. The electrode was equilibrated with air-saturated buffer between runs. Steady-state kinetic parameters were evaluated by fitting the Michaelis-Menten equation to the data using the least-squares fitting of LEONORA.[210]

Plu4 was assayed in the presence of cytochrome c to detect the presence of superoxide. Detection of superoxide was done by including 0.7 mg/mL cytochrome c in a Plu4 reaction containing 5 μM Plu4, 1 mM L-arginine, 25 μg/mL catalase in Tris-HCl pH 8.5. Reactions were initiated by addition of L-arginine, after which UV signals were measured immediately, after 2

min and after 4 min. The absorbance signal was obtained on a Cary 100 UV-Vis spectrophotometer. Negative controls, which excluded either Plu4 or L-arginine, were compared to the samples containing both the enzyme and L-arginine. Rates of cytochrome c reduction were measured from the absorbance signal at 550 nm. The linear trendline was determined from the maximal velocity measured over one minute. Rates were measured in triplicate. The rate of cytochrome c reduction in the presence of 1000U/mL superoxide dismutase (SOD) was also determined in triplicate to determine if SOD inhibits cytochrome c reduction.

3.2.6 Electron paramagnetic resonance (EPR)

Reactions were done by combining 0.5 mM of compound **1**, 10 μM Plu4 and 1 mM cyclic hydroxylamine 1-hydroxy-3-methoxycarbonyl-2,2,5,5-tetramethylpyrrolidine (CMH) in 50 mM Tris pH 8.5 buffer. The buffer was mixed with CHELEX (5 g/ 100 mL) for 7 hours and then mixed with 4 mg DTPA (100 μM) to remove metals that could interfere with CMH. Controls were done using either CMH only, CMH and **1** only, or CMH, **1** and boiled Plu4. The full reaction with active Plu4 and the boiled Plu4 control reaction were done in triplicate. Either 25 μg/mL of catalase or 0.5 mM hydrogen peroxide were used in the assays to test the potential effect of the hydrogen peroxide product on the CMH signal.

Electron paramagnetic resonance (EPR) spectra were collected at room temperature at X-band (9.80-9.84 GHz) using a Bruker EMXplus spectrometer, a PremiumX microwave bridge, and a HS resonator. Samples were prepared by adding an aliquot (~300 μL) of either a control or reaction solution into a quartz flat cell ensuring sufficiently high resonant Q-factor for the measurement of aqueous samples. General experimental parameters: frequency = 9.81 GHz, microwave power = 2.0 mW, time constant = 10.24 ms, modulation amplitude = 3.0 G, where each spectrum is an average of five 15-second scans. All spectral centres were calibrated to a frequency of 9.81 GHz.

3.3 Results

3.3.1 Sequence similarity networks

In order to better understand the relationship between the two types of arginine oxidases, SSNs were made to help map out potential evolutionary relationships within the arginine oxidases. Also, by assigning the unknown enzymes to specific functions, the sequence

similarities and differences could be used to build a hypothesis for the enzymes' reactivity. Using Ind4 as a query, an initial SSN with a threshold of 10^{-30} was made from the top thousand sequences and showed that the known arginine oxidases are highly isolated from the other top hits, which were uncharacterized aminotransferases (**Figure 3.1**). Another SSN was made only using the 296 sequences that clustered together with the known arginine oxidases at a threshold of 10^{-65}, revealing three distinct groups labelled Groups I-III (**Figure 3.1**).

Next, the genomic context of each of these nodes was analyzed to assign probable function to uncharacterized nodes based on neighbouring genes. Over 14 different gene organizations were observed (**Appendix Figure A-1, Appendix Figure A-2, Appendix Figure A-3 and Appendix Figure A-4**), five of which had known functions in the production of four natural products: enduracididine, azomycin, mildiomycin and indolmycin (both the *P. luteoviolacea* and *S. griseus* ATCC 12648 gene organizations were observed). Group II had a large portion of its enzymes (~57%) contained in enduracididine biosynthetic gene clusters,[164,166] but at least two other uncharacterized non-enduracididine gene clusters were identified (**Appendix Figure A-2**). Since the majority of the enzymes in this group are in enduracididine gene clusters, this group is also likely to consist of mostly hydroxylases. Group III contains Plu4 from the *P. luteoviolacea* indolmycin gene cluster (**Appendix Figure A-4**), but the incidence of the indolmycin gene cluster within this group is fairly limited. However, other enzymes in Group III are co-localized with *plu5*-like genes and sometimes *ind3*-like genes (**Appendix Figure A-4**). Currently, colocalization with an *ind5/plu5*-like gene is an invariable quality of the arginine desaturases. Although it is theoretically possible for an arginine desaturase to exist independent of a colocalized *ind5/plu5*-like gene, or for an arginine hydroxylase to have an *ind5/plu5*-like counterpart in its gene cluster, further characterization of the relationship between arginine desaturases and their *ind5/plu5*-like counterparts would be needed to disprove this colocalization pattern as an essential characteristic of arginine desaturase activity *in vivo*. Based on the current knowledge, these colocalized genes suggest that many members of Group III could function as arginine desaturases.

Despite the seemingly consistent clustering of activities in Groups II and III, Group I consisted of diverse gene clusters and previously characterized hydroxylases and desaturases, making the organization of this group non-isofunctional. Therefore, Group I was further analyzed at a 10^{-85} threshold, revealing three isofunctional sub-groups: Ia, Ib, and Ic (**Figure**

3.1). Group Ia contained the characterized enzyme, RohP,[116] and the four-gene azomycin gene cluster[173] was a highly conserved gene organization from this sub-group (**Appendix Figure A-1a**). Additionally, one enzyme in Group Ia was found in the mildiomycin biosynthetic gene cluster (**Appendix Figure A-1b**), but the enzyme has not been characterized.[211] Since mildiomycin contains a 4-hydroxy-2-ketoarginine side chain, this enzyme is also likely a hydroxylase. Therefore, Group Ia consists of known and predicted arginine hydroxylases. Group Ib consisted of only two nodes contained in previously known gene organizations, both of which are from *S. griseus* ATCC 12648-like indolmycin gene clusters[48,113] (**Appendix Figure A-3a**). One other enzyme from Group Ib, Pel4, is from an uncharacterized gene cluster, but has also been shown to have desaturase activity.[113] The gene organization surrounding the *pel4* gene was similar to several other nodes in Group Ib (**Appendix Figure A-3b**). The remaining nodes were found in an unknown gene cluster from *Bacillus* (**Appendix Figure A-3c**), making it unclear if they were desaturases. However, the *Bacillus* nodes were also colocalized with an *ind5*-like gene, suggesting that they are likely to function as arginine desaturases. Therefore, Group Ib consists of nodes that are predicted to be arginine desaturases. The function of Group Ic nodes and their associated gene clusters remain unknown.

Through analysis of these SSNs, it was expected that the hydroxylase nodes would cluster together with higher sequence similarity than with any desaturase nodes. The same pattern was expected for the desaturases. However, the Ind4-like desaturases from Group Ib and the RohP-like hydroxylases from Group Ia displayed the strongest sequence similarity relationship amongst all the groups. (**Figure 3.1**). This sequence similarity patterns suggests that either the hydroxylases or desaturases emerged on at least two different occasions during evolution, which is also mirrored by analysis of the phylogenetic trees (**Figure 3.1** and **Figure 3.2**). This feature could make elucidating the mechanism of differentiation challenging since there could be more than one evolutionary pathway that each group took to reach the same reactivity.

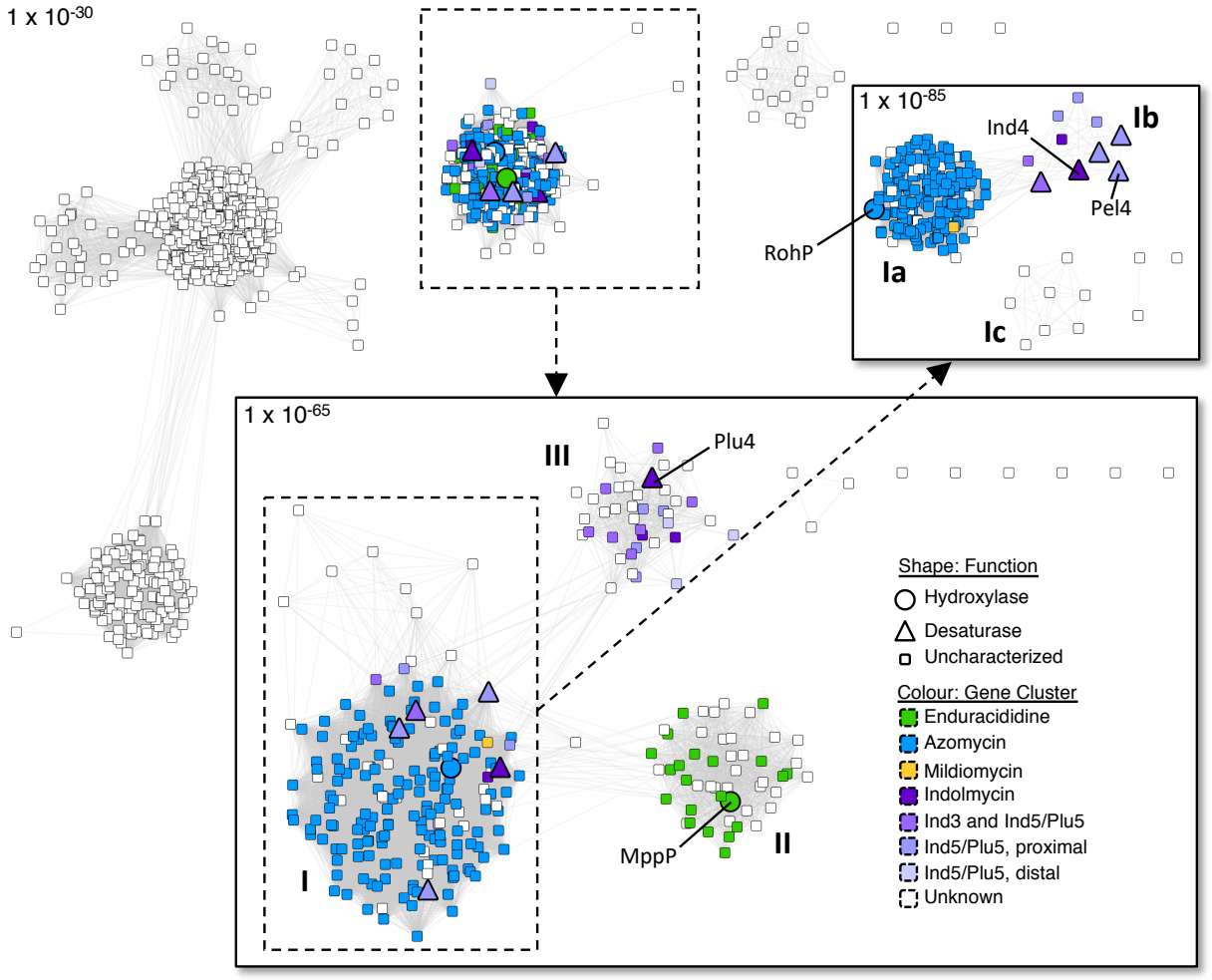


Figure 3.1: Sequence similarity network analysis for the arginine oxidases. Each node represents a different enzyme from the UniProtKB database and edges represent sequence similarity relationships. Node shapes represent the characterized function of the enzymes, while node colours indicate information about surrounding genes in the genome. Previously characterized nodes and newly characterized Plu4 are labelled. Dashed boxes indicate nodes used for smaller networks at a higher threshold cut-off. These smaller networks at thresholds of $1x10^{-65}$ and $1x10^{-85}$ are shown as inset boxes. Each predicted reactivity group is labelled as I-III.

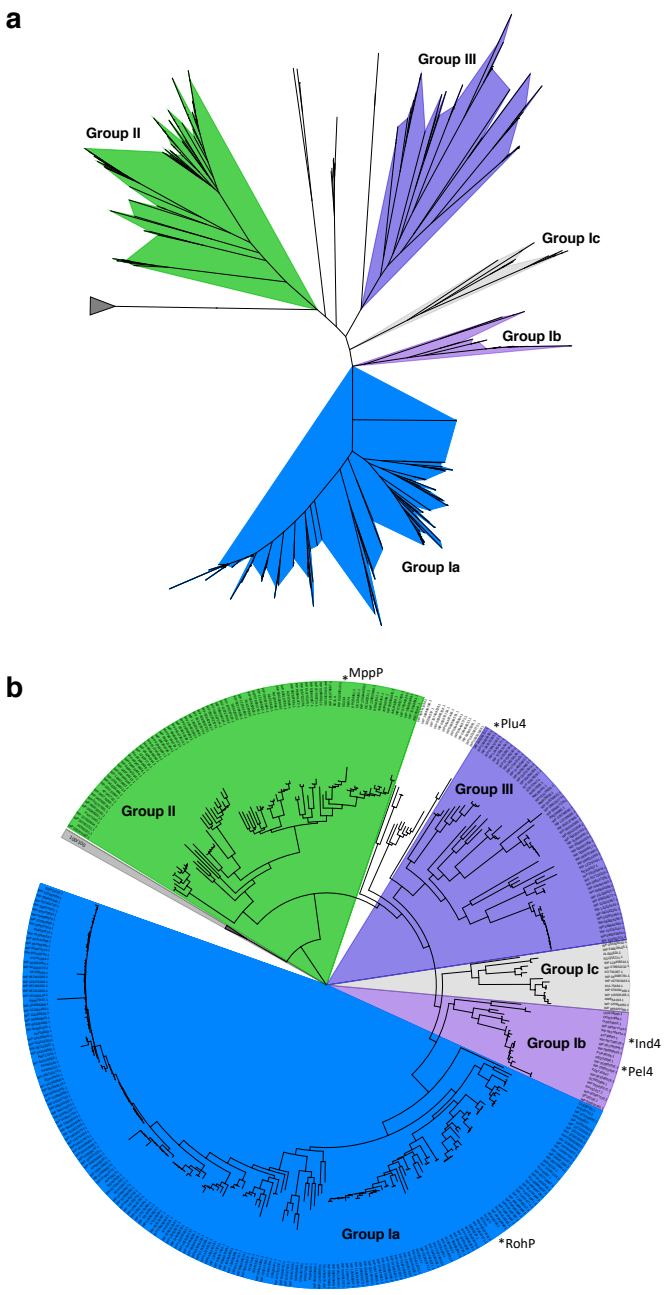


Figure 3.2: Phylogenetic trees of the arginine oxidase family of enzymes. Trees were created from the top 1000 sequences obtained from a BLASTp search of Ind4 (January 21st, 2019) and are shown as either (**a**) unrooted, or (**b**) circular and rooted from distantly related enzyme hits. The distantly related hits are collapsed and shown as a grey triangle for simplicity in both trees. Clades are labelled with their corresponding SSN group and known arginine oxidases are identified with an asterisk and labelled. Uncoloured clades could not be matched to an SSN group for lack of appearance in the UniprotKB database.

3.3.2 *In vitro* desaturase characterization

3.3.2.1 Product analysis and kinetics[5]

When crystals could not be obtained for Ind4, attempts were made to crystallize one of seven other homologs that were identified, cloned and purified (**Chapter 2**). In order to confirm these enzymes as Ind4 homologs, an LC-MS based assay was used to analyze the products of each enzyme, with a particular focus on the crystallized homolog, Plu4. Each purified homolog performs an identical reaction to Ind4, producing 174 *m/*z and 172 *m/*z signals in the presence of catalase and **1** (**Figure 3.3** and **Figure 3.4a**), corresponding to expected $[M+H]^+$ ions for **4** and **5**, respectively (**Figure 2.1**). During shorter reaction times, it was also shown in Plu4 that this reaction likely traverses through **2** (**Figure 2.1**) because the corresponding $[M+H]^+$ 171 *m/z* signal was observed (**Figure 3.5**). In steady-state kinetic analyses using an oxygraph, Plu4 was most active at a pH of 8.5-9.0 (**Figure 3.6**). When assayed using TAPS buffer (I = 0.1; pH 8.5) at 25 °C, Plu4 had K_M and k_{cat} values of 12 ± 1 μM and 12.9 ± 0.3 min^{-1}, respectively, for **1** and K_M and k_{cat} values of 15 ± 1 μM and 14.9 ± 0.3 min^{-1}, respectively, for O_2 (**Figure 3.4b** and **c**). Altogether, these data support that the identified homologs, including Plu4, are members of the arginine desaturases consistent with their predicted functions.

[5] E.K. completed steady-state kinetic analyses and determined kinetic parameters.

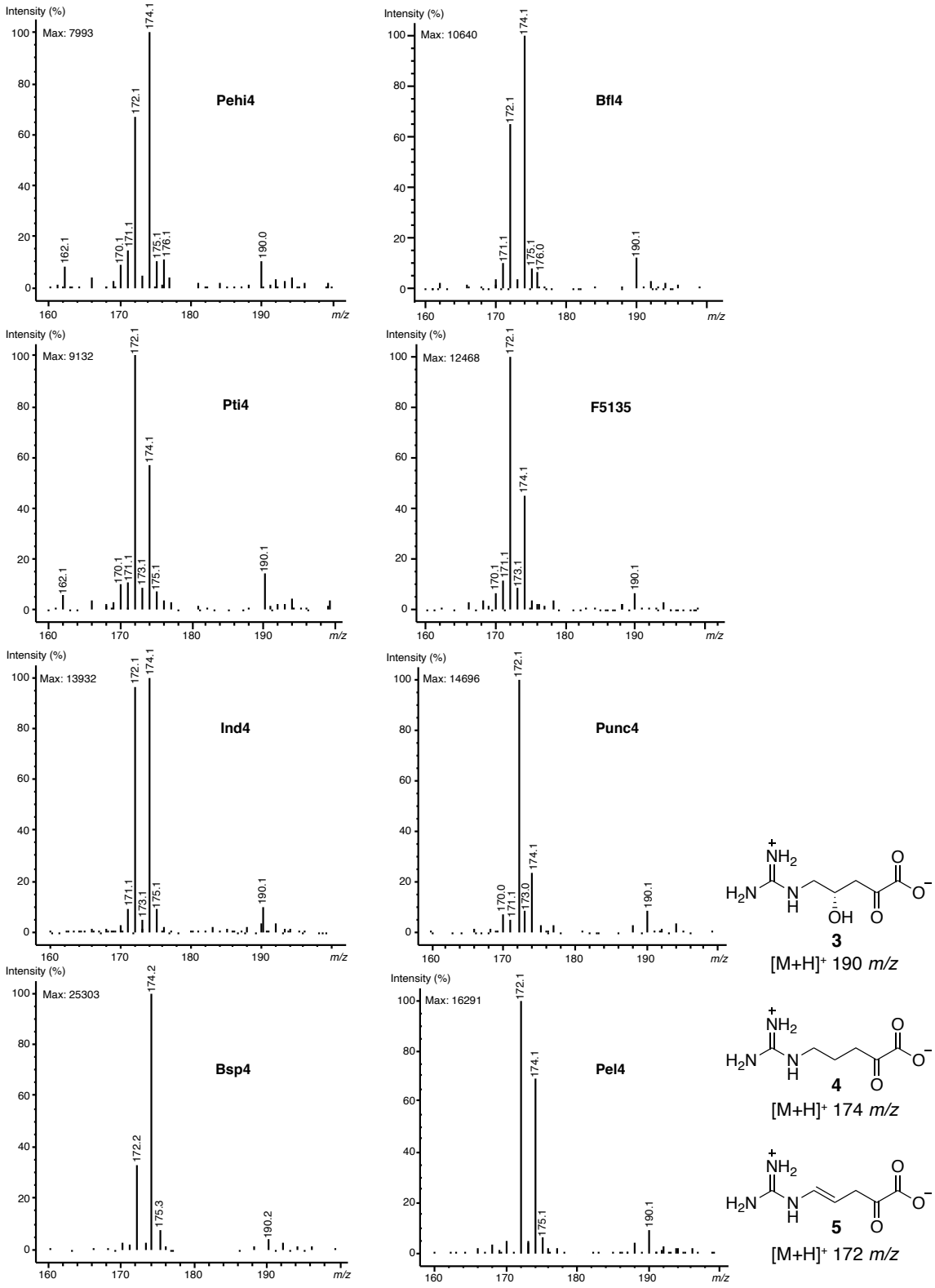


Figure 3.3: LC-MS assay of known and newly discovered arginine desaturases from SSN Group III. Peak heights are not quantitative because ionization efficiencies will differ between each product. The structures of compounds **3**, **4**, and **5** are given to the left with their corresponding $[M+H]^+$ ions.

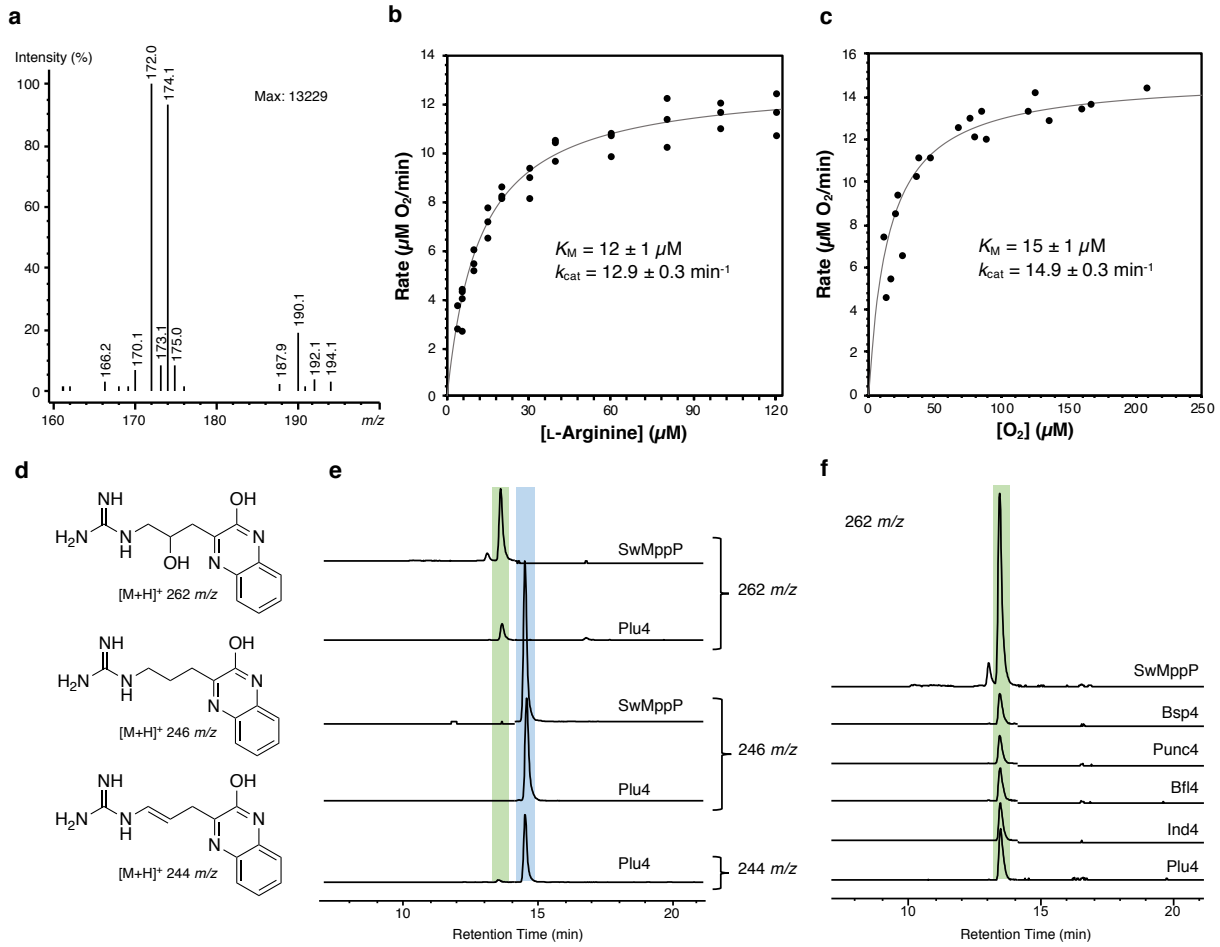


Figure 3.4: Plu4 reaction products and kinetics. (**a**) LC-MS analysis of Plu4 reaction with **1**. (**b** and **c**) Steady-state kinetic analyses of the Plu4-catalyzed reaction, with the lines representing fits of the Michaelis-Menten equation to the data: dependence of initial velocity on (**b**) **1** and (**c**) O_2. Concentrations were measured by the rate of oxygen consumption. (**d**) Structures of OPD-derivatized products and their corresponding $[M+H]^+$ signals. (**e**) Comparison of Plu4 and SwMppP products as extracted ion chromatograms from LC-MS after 20 min OPD-incubations. (**f**) Comparison of OPD-derivatized products from several desaturases and SwMppP. Traces are extracted ion chromatograms at 262 *m/z* from LC-MS. Peaks corresponding to **3** are highlighted in green, while peaks corresponding to **4** and **5** are highlighted in blue in (**e**) and (**f**).

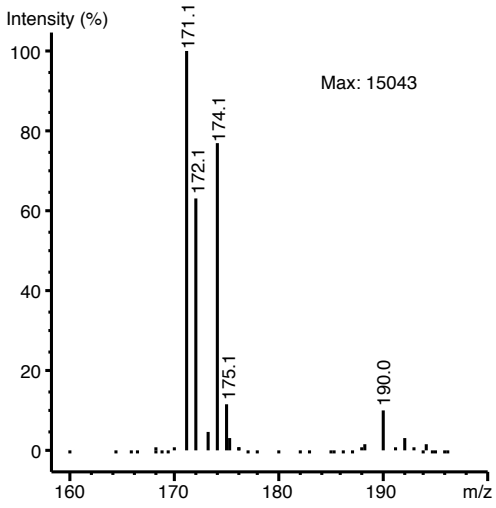


Figure 3.5: Plu4 reaction products measured by LC-MS showing didehydroarginine (based on production of an $[M+H]^+$ ion at *m/z* 171) when the reaction was run for 2.5 h instead of the typical 16 h. All other reaction conditions remained the same as the typical 16 h reaction.

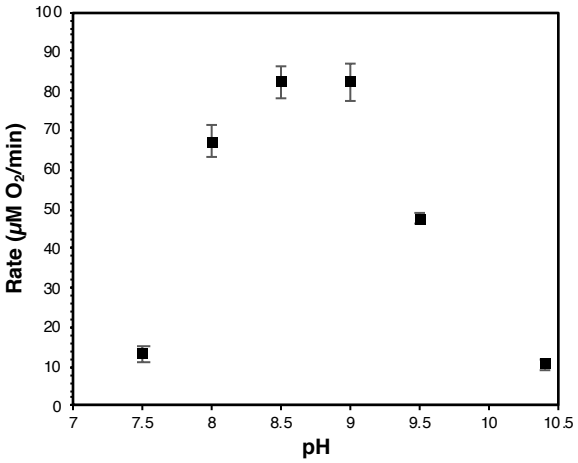


Figure 3.6: Measurement of Plu4 pH optimum. Reaction rate of Plu4 from the rate of oxygen consumption measured by oxygraph in pH 7.5 (HEPES), 8 (Tris), 8.5 (Tris), 9.0 (Tris), 9.5 (CHES), and 10.4 (CAPS). Each data point represents the average of three measurements and error bars represent the standard deviation.

3.3.2.2 Side-product analysis

While characterizing Plu4, there was a minor peak at 190 *m/z*, which matches the mass of **3** (**Figure 3.4a**). This peak appears consistently in the LC-MS traces for all of the desaturases that were tested, including Ind4 **(Figure 3.3).** Therefore, attempts were made to determine if this mass corresponded to a minor hydroxylated side-product made by the desaturases, which could help elucidate the evolutionary history and provide key mechanistic details for the arginine oxidases. To determine the identity of this product, OPD was used to derivatize the products of the Plu4 reaction and separate them by LC-MS (**Figure 3.4d**). Analysis of the derivatized products revealed a small peak with the same $[M+H]^+$ signal of 262 *m/z* and retention time as derivatized **3** from SwMppP (**Figure 3.4e**). The presence of this peak was also confirmed in reactions with other desaturases, including Ind4, under the same reaction conditions (**Figure 3.4**). The new peak observed in the Plu4 reaction does not appear when the enzyme or **1** are omitted, or when the enzyme has been boiled (**Figure 3.7a**), suggesting that it arises from an enzymatic process involving **1**. To further validate that it arises from an enzymatic reaction, the signal from non-derivatized products did not increase after the reaction mixture was separated from the enzyme (**Figure 3.7b**). Additionally, similar to the formation of **3** in the MppP and RohP reactions,[115,116] by assaying the enzyme in 50% $H_2{}^{18}O$ it was demonstrated that this side-product incorporated water during formation (**Figure 3.8**).

SwMppP reactions were also analyzed to see if a product consistent with OPD-derivatized **5** could be detected. The SwMppP reaction produced a signal with an $[M+H]^+$ signal of 244 *m/z*, consistent with the OPD-derivatized **5**; however, the appearance of the 244 *m/z* peak is concurrent with the disappearance of the 262 *m/z* signal during longer derivatization times (**Figure 3.9**). Therefore, it is likely that the 244 *m/z* peak from SwMppP comes from a non-enzymatic dehydration of OPD-derivatized **3** during the derivatization reaction. Using a shorter incubation time of 3 min, the desaturase side-product was quantified by HPLC and compared to the production of **3** from SwMppP. The amount of derivatized **3** was ~35-fold lower than the peak corresponding to derivatized **4** and **5** in Plu4 reactions (**Figure 3.10**). By contrast in SwMppP, the production of derivatized **3** is ~1.7-fold less than derivatized **4** and **5**, suggesting a bona fide difference in reaction outcome for each enzyme, consistent with their biosynthetic roles.

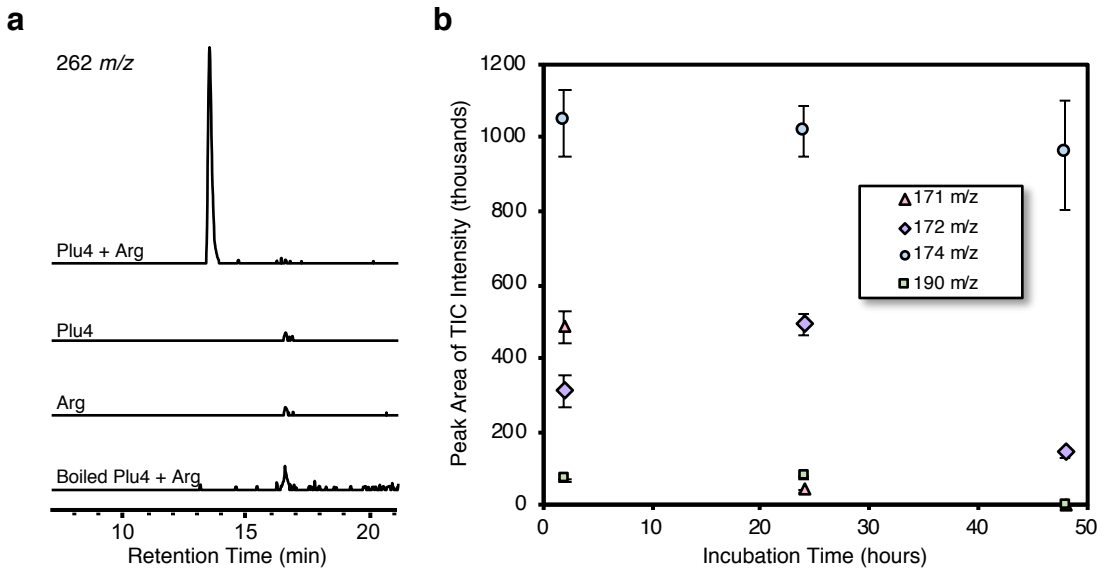


Figure 3.7: Desaturase side-product controls. (**a**) Negative controls for Plu4 reaction. Traces are extracted ion chromatograms at 262 *m/z* from LC-MS for reactions derivatized with OPD. (**b**) Analysis for non-enzymatic production of hydroxylated side-product. Plu4 reaction products analyzed for various amounts of time by LC-MS after separation from enzyme by filtration. Reactions proceeded for 2.5 h before the enzyme was filtered. Catalase was added to the filtrate to prevent decarboxylation of the α-keto acid products. Each data point represents the average of two replicates and error bars represent the standard deviation.

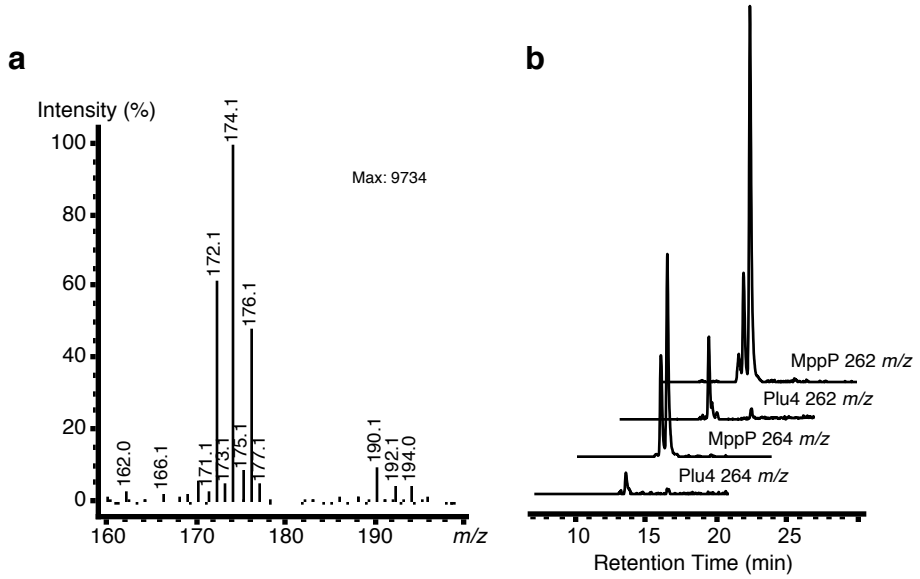


Figure 3.8: Plu4 assay with 50% $H_2{}^{18}O$. (**a**) LC-MS assay with small peaks at 192 *m/z* and 194 *m/z*. (**b**) Comparison of OPD-derivatized products from Plu4 and SwMppP assayed in 50% $H_2{}^{18}O$ by LC-MS analysis. Derivatization reaction was done at 80 °C for 3 min to minimize a dehydration side reaction.

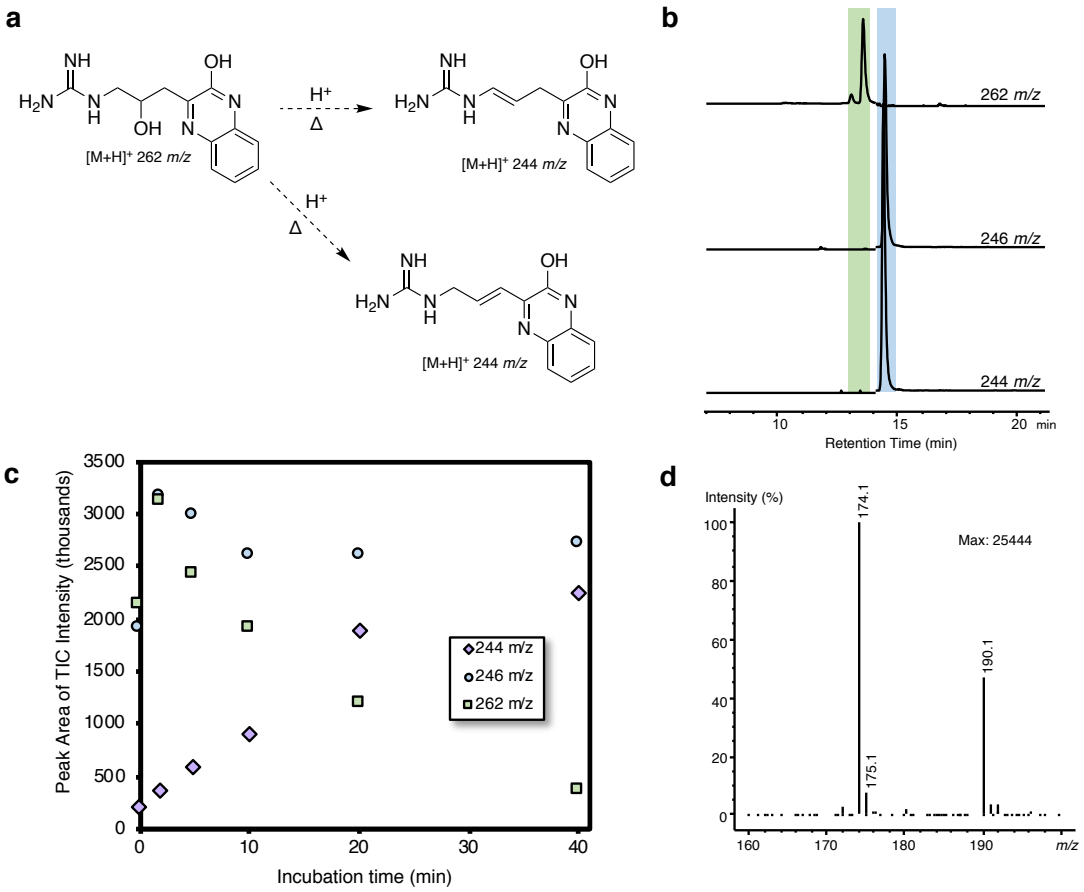


Figure 3.9: Optimization of OPD derivatization conditions using SwMppP. (**a**) Two possible pathways for forming a 244 *m/z* product from hydroxylated product under the OPD derivatization conditions. (**b**) LC-MS analysis OPD-derivatized SwMppP reaction products after a 20 min derivatization. Each trace represents an extracted ion chromatogram. (**c**) OPD-derivatized products from SwMppP at various incubation times at 80°C analyzed by LC-MS. (**d**) LC-MS control showing no products corresponding to dehydro-2-ketoarginine ($[M+H]^+$ 172 *m/z*) before derivatization.

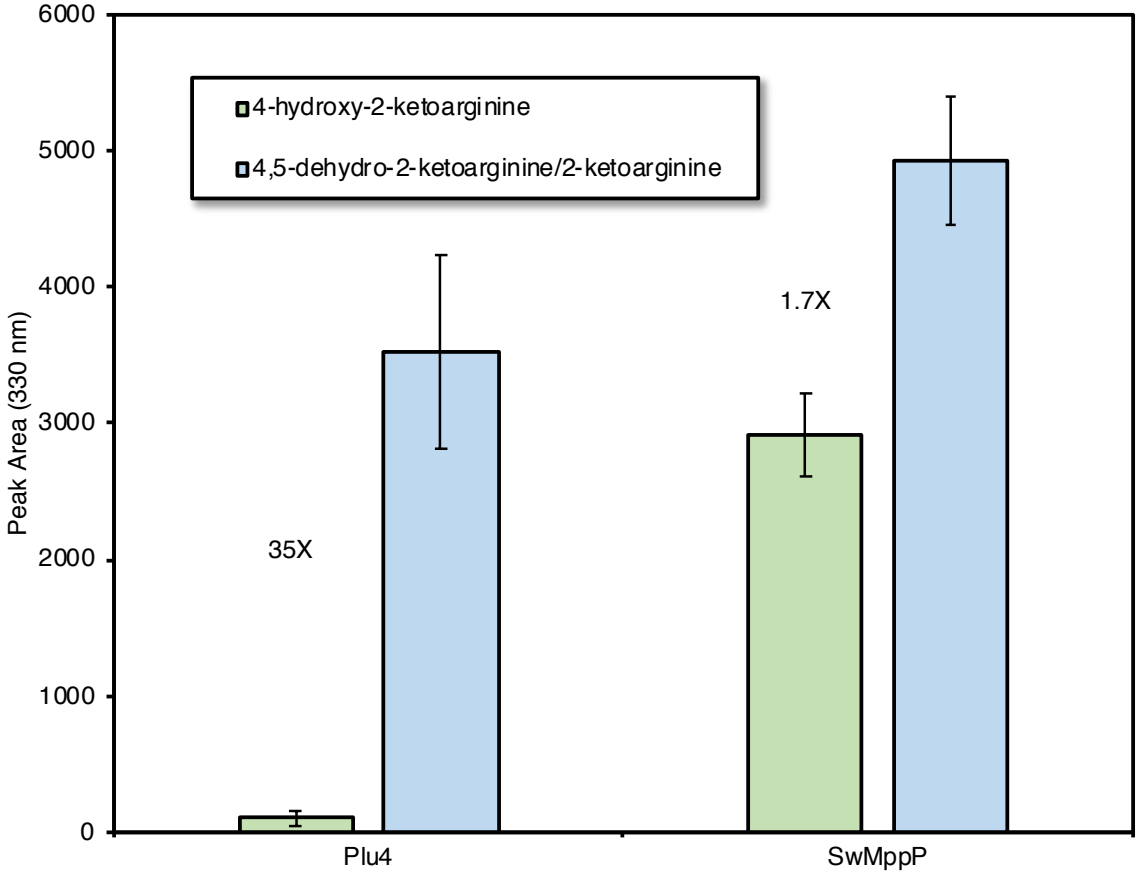


Figure 3.10: Quantification of hydroxylated product comparing Plu4 and SwMppP. Analysis was done on OPD-derivatized products by integrating peak areas from HPLC at 330 nm. 4,5-dehydro-2-ketoarginine and 2-ketoarginine are combined into a single bar because they have the same retention time. Each bar represents the average of two replicates and error bars represent the standard deviations. The fold-difference between the hydroxylated product and the other products is indicated. Derivatization reaction was done at 80°C for 3 min to minimize a dehydration side reaction.

3.3.2.3 Demonstrating the existence of superoxide[6]

To better understand the role of oxygen in the arginine oxidase reactions, EPR and a spin trap, CMH, was used to look for the presence of superoxide during the course of the reaction.[123,212] To do so, 0.5 mM of **1** and 1 mM of Plu4 were combined in the presence of 1 mM CMH to detect superoxide during the reaction. A four-fold enhanced signal for CMH-trapped superoxide was observed when both Plu4 and **1** were present (**Figure 3.11** and **Figure 3.12a**). When catalase was added to the reaction, the observed signal dropped to a two-fold increase and

[6] G.A.M. analyzed the EPR samples and helped with data processing.

the signal increased slightly when only Plu4 and hydrogen peroxide were included (**Figure 3.12b**). These controls indicate that the hydrogen peroxide produced during catalysis is able to react with CMH to some degree, but the remaining two-fold signal from the catalase-containing reaction indicates that superoxide is produced from the Plu4 reaction.

To further confirm the presence of superoxide as an intermediate in the reaction, Plu4 was assayed in the presence of cytochrome c, which uses superoxide to reduce Fe^{3+} to Fe^{2+}, producing a UV-Vis signal at 550 nm (**Figure 3.13**). Superoxide dismutase (SOD) was also included in the assay to observe if Fe^{3+} reduction by cytochrome c was inhibited. The rate of reduction for the SOD-free reaction was 21.1 ± 0.5 μM min^{-1}, while the rate of reduction was 15.4 ± 0.5 μM min^{-1} when SOD was included in the assay. These results demonstrate that superoxide is present as an intermediate in the reaction because the inclusion of SOD resulted in a decreased rate as SOD lowered the amount of superoxide available to reduce the Fe^{3+}. Interestingly, the SOD-free rate of Fe^{3+} reduction was similar to reaction rates measured from the oxygraph experiments.

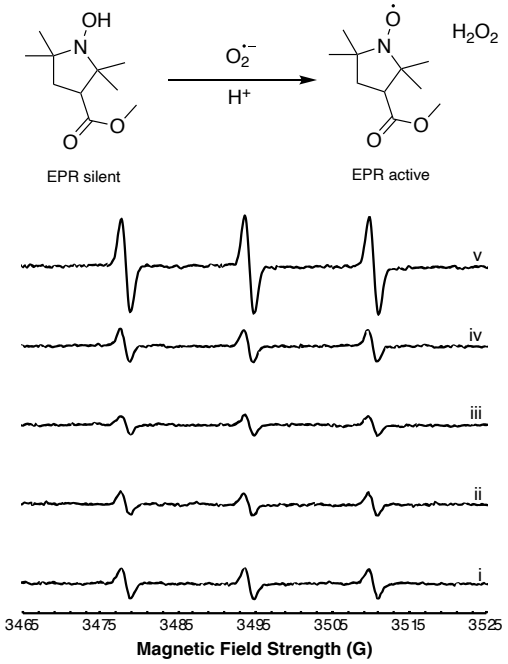


Figure 3.11: EPR measurements demonstrating production of superoxide. EPR spectra is from CMH spin-trap radical adduct, shown above the spectra. Traces show (i) CMH, (ii) CMH and L-Arg, (iii) CMH and Plu4, (iv) CMH, L-Arg and boiled Plu4, (v) CMH, L-Arg and Plu4.

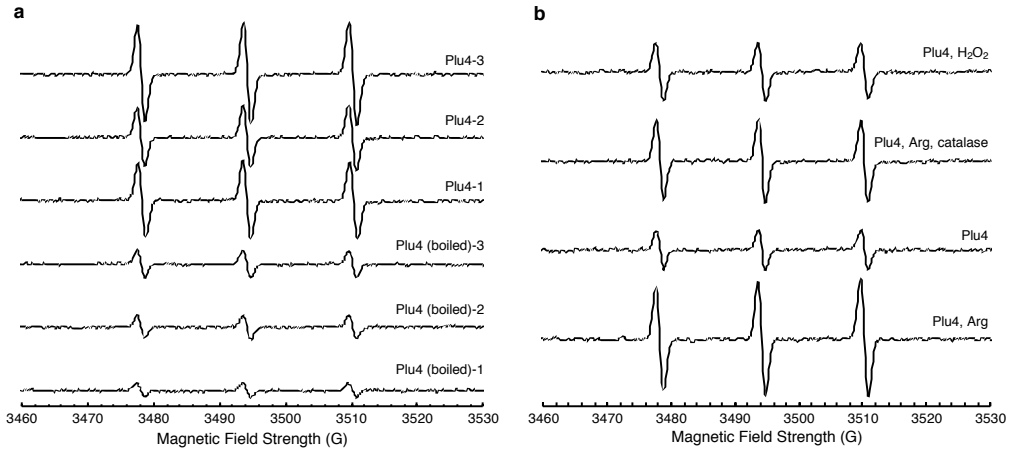


Figure 3.12: EPR spectra of CMH-trapped superoxide during control reactions of L-Arg with Plu4. (**a**) Comparison of reactions with active Plu4 and boiled Plu4 done in triplicate to demonstrate reproducibility. (**b**) Effect of hydrogen peroxide on CMH signal using catalase to consume the hydrogen peroxide by-product of the Plu4 reaction with arginine. A Plu4 sample without arginine was also spiked with hydrogen peroxide.

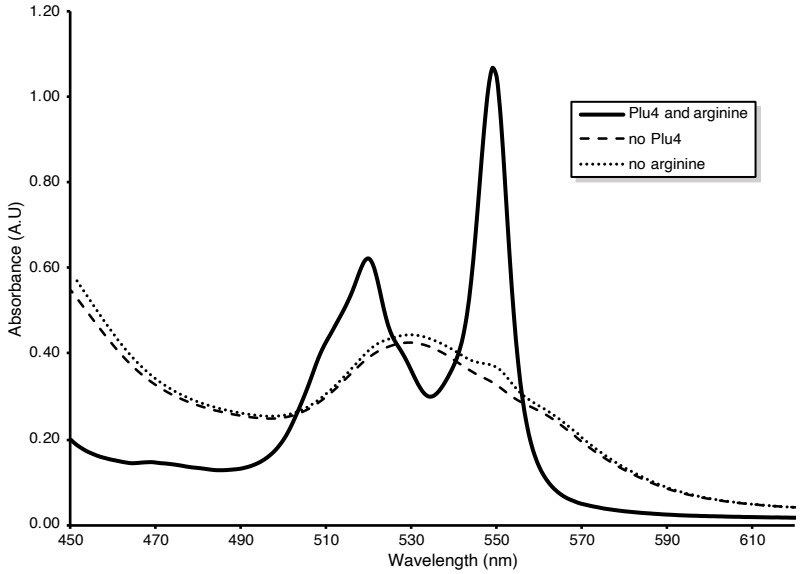


Figure 3.13: Absorbance spectra of Plu4 reaction assayed in the presence of cytochrome c. UV-Vis absorbance at 550 nm is used to detect reduction of Fe^{3+} to Fe^{2+} by superoxide. Cytochrome c reduction is not present when Plu4 or arginine are eliminated from the assay.

3.3.3 Density functional theory calculations[7]

To further investigate the role of superoxide, the likelihood of possible mechanisms was explored though DFT calculations. To summarize this investigation, a superoxide rebound mechanism was shown to be more favourable than a hydrogen abstraction mechanism for both oxygen reaction steps. Energies calculated for each possible peroxy-intermediate suggests that oxygen reacts at the C4′ position of PLP for both Quinonoid I and II. Finally, a comparison of the hydroxylation and deprotonation steps shows that the hydroxylation pathway was enthalpically favoured, which supports the fact that Plu4 has a minor hydroxylation activity despite being a desaturase. The enthalpic favourability of hydroxylation over deprotonation further validates that the key distinction between hydroxylation and double bond formation could be the position of water in the active site of the enzymes during catalysis, a proposal that was initially introduced in **Chapter 2**, **Section 2.4**.

3.4 Discussion

The oxygen reactivity demonstrated by the arginine oxidases shows the catalytic potential of PLP-dependent enzymes in their ability to catalyze challenging oxidation reactions without the use of a metal cofactor. Building upon the structural and mutagenesis work described in **Chapter 2**, the work described here dives further into the biochemical and evolutionary characteristics of these enzymes with a particular focus on the newly characterized Plu4 desaturase from *P. luteoviolacea*. A look at the SSNs for arginine oxidases revealed that there was a stronger evolutionary relationship between the Group Ia hydroxylases and Group Ib desaturases than between the Group Ia hydroxylases and the Group II hydroxylases or between Group Ib desaturases and Group III desaturases. This result demonstrates that there are at least four distinct phylogenetic groups of arginine oxidases (two hydroxylases and two desaturases) that have now been characterized and that either the hydroxylase or desaturase activity emerged separately on more than one occasion, indicating a complex evolutionary history. Furthermore, consistent appearance of a minor hydroxylated side-product in all characterized desaturases indicates that the hydroxylase and desaturase activities are highly related and provide evidence for promiscuous enzyme evolution.[213] The minor appearance of **3** from the desaturases could

[7] K.C.H. completed DFT calculations, described in further detail in publication.

have either served as a starting point for the evolution of hydroxylation reactions from a desaturase ancestor, or it could represent a vestigial feature of a hydroxylating ancestor.

Examples of such a complex evolutionary history and promiscuous enzyme evolution have also been seen for other protein families. The *o*-succinylbenzoate synthase/*N*-succinylamino acid racemase (OSBS/NSAR) family is a similarly complex example of non-isofunctional phylogenetic clustering, in which a group of OSBSs share a closer evolutionary relationship with the NSARs (>40% sequence identity) than they do with other OSBSs (**Figure 3.14a**).[214–216] Additionally, independent evolution of a group of NSARs was also discovered, having a closer evolutionary relationship with L-Ala-D/L-Glu epimerases (AEEs) than with other NSARs.[215] Another example comes from the 6-pyruvoyl-tetrahydropterin synthases (PTPSs),[217,218] which demonstrated that some of the six reactivities observed from these enzymes did not cluster isofunctionally in an SSN, suggesting that each function may have evolved on more than one occasion (**Figure 3.14b**).[217] This complex sequence similarity pattern results in the PTPSs being commonly misannotated without any genomic context.[217,219] Since the arginine oxidase reactivities are most commonly identified based on similar genomic neighbourhoods, automated identification of the arginine desaturases and hydroxylases could also result in misannotations. Additionally, some PTPSs promiscuously catalyze two reactions, pointing towards a potential starting point for evolution of other reactions.[218,220] Promiscuous enzyme evolution is thought to contribute to the appearance of new reactions by acting as a starting point to evolve a side reaction into an enzyme's primary function and contributes to the chemical diversity of natural products.[213,221,222] Brown and Babbitt state that reaction promiscuity and independent appearances of the same reactivity within a superfamily's phylogeny often complicates attempts to understand enzyme evolution and predict reactivity from sequence.[223]

The EPR experiments and cytochrome c assays described here show for the first time that superoxide is an intermediate to reaction within the arginine oxidases. Although only demonstrated for an arginine desaturase, the high similarity with the hydroxylases implies that the hydroxylase reactions also likely include superoxide as an intermediate. A superoxide intermediate has been demonstrated previously for the paracatalytic oxidative deamination reaction of DDC.[125] Additionally, flavin-dependent enzymes and cofactor-independent enzymes also produce superoxide through a single electron transfer to oxygen and will serve as good models for further studies of oxygen activation in the arginine oxidases.[120,121,123,128] DFT

calculations showed that superoxide rebound at the C4′-position of PLP is the most favourable, resulting in a peroxy-intermediate. This result is unprecedented in the arginine oxidases as previous mechanisms have proposed peroxy-adduct formation at the Cα or direct deprotonation of arginine protons by superoxide.[113–115] However, formation of the peroxy-adduct at the C4′-position of PLP can be rationalized by the arginine oxidases' close evolutionary relationship to, and frequent annotation as aspartate aminotransferases, whose mechanisms typically include protonation at the C4′-position of PLP. Alternatively, some of the other oxygen- and PLP-dependent enzymes, such as Cap15[119] and CuaB[117], show evidence that points towards peroxy-intermediate formation at the Cα position of their substrate, based on the appearance of labelled oxygen atoms being incorporated into their products. These results again can be rationalized by the similarity of Cap15 and CuaB with selenocysteine synthases and AONS, respectively (**Table 1.1**), which both involve protonation or carbon-carbon bond formation at the Cα position of the substrate. Therefore, it is possible that the oxygen reactivity sites in the oxygen- and PLP-dependent enzymes could be predicted based on knowing the site of new bond-formation in the nearest characterized relatives.

The key difference between these desaturases and hydroxylases that aids in the proposal of a new, unified mechanism for the arginine oxidases is the use of water. While the hydroxylases require water to produce **3**, water is not a reactant needed to produce **2**, making the movement of water a potentially critical factor in differentiating the arginine oxidase reactivity. As discussed in **Chapter 2**, mutagenesis revealed that there were no single active site residues that differentiated reactivity in the arginine oxidases. Additionally, lower N-terminal B-factors and single conformations of active site hydrophobic residues were observed in the Plu4 crystal structure while the corresponding active site residues in the hydroxylases were observed in multiple conformations and the N-termini had higher B-factors. Therefore, the active sites of the hydroxylases appear to be more dynamic than the desaturases, making it potentially more open to water from outside the active site.

Besides the structural differences, the enzymes have highly similar active sites, including the catalytic His/Asp dyad, which supports that water may be the distinguishing factor. A comparison of the final deprotonation and hydroxylation steps by DFT revealed that the hydroxylation step is more enthalpically favourable than deprotonation. This result indicates that the presence of an ordered water in the active site of a desaturase will produce **3** as a side-

product. These calculations also support that the hydroxylation mechanism does not go through a fully oxidized intermediate with deprotonation at C5 because the hydroxylation step would likely precede deprotonation, since it is more favourable. Therefore, the distinguishing feature between the arginine hydroxylases and desaturases may be an inability to preorganize water in the Plu4 active site, with His27 deprotonating the C5 position of **1** (**Figure 3.15**). Alternatively, the active sites of RohP and SwMppP would be able to preorganize water and the active site His would act to deprotonate water instead of C5. This proposal is not unprecedented as water has been discussed previously to play a role in distinguishing promiscuous activities in some enzymes.[213,224,225]

Overall, the arginine desaturases and hydroxylases show a high degree of similarity despite catalyzing different reactions. A reliance on the preorganization of water is proposed as the reaction differentiation mechanism and accounts for the underlying hydroxylation side reaction in the desaturases. A mechanism for oxygen usage is offered that includes a single electron transfer to oxygen and hydroperoxyl-intermediate formation at the C4′ position of PLP, which suggests that the oxygen reactivity is not reliant on arginine as a substrate and could, theoretically, be engineered into other aminotransferases. Finally, the sequence similarity and phylogenetic analysis reveals that the desaturases and hydroxylases appear to each have two distinct phylogenetic groups, clustered non-isofunctionally, which will be essential to future studies aimed at understanding how the two activities evolved. This new information and proposed mechanism will help guide future efforts in biocatalyst engineering and discovery amongst the arginine oxidases, which will expand the biocatalytic toolbox and its practicality for use by chemists.

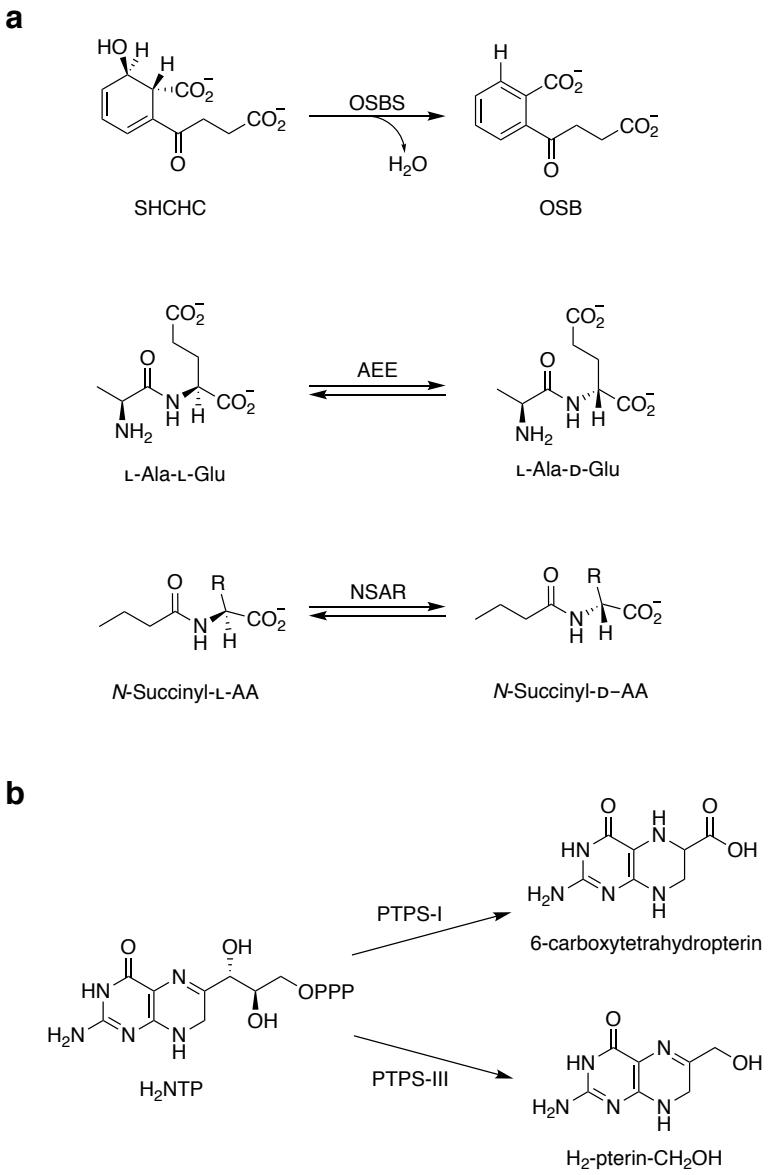


Figure 3.14: Reaction schemes for examples of non-isofunctional phylogenetic clustering. (**a**) *o*-succinylbenzoate synthase (OSBS), L-Ala-D/L-Glu epimerases (AEE) and *N*-succinylamino acid racemase (NSAR) reaction schemes are shown. The OSBSs and NSARs do not cluster isofunctionally and some enzymes show functional promiscuity/bifunctionality with both OSBS and NSAR reaction types.[214,215] (**b**) 6-pyruvoyl-tetrahydropterin synthase (PTPS) reaction schemes are shown for groups I and III, in which some PTPS enzymes catalyze both reactions promiscuously. PTPS-I and -III enzymes do not appear to cluster isofunctionally in a sequence similarity network.[217]

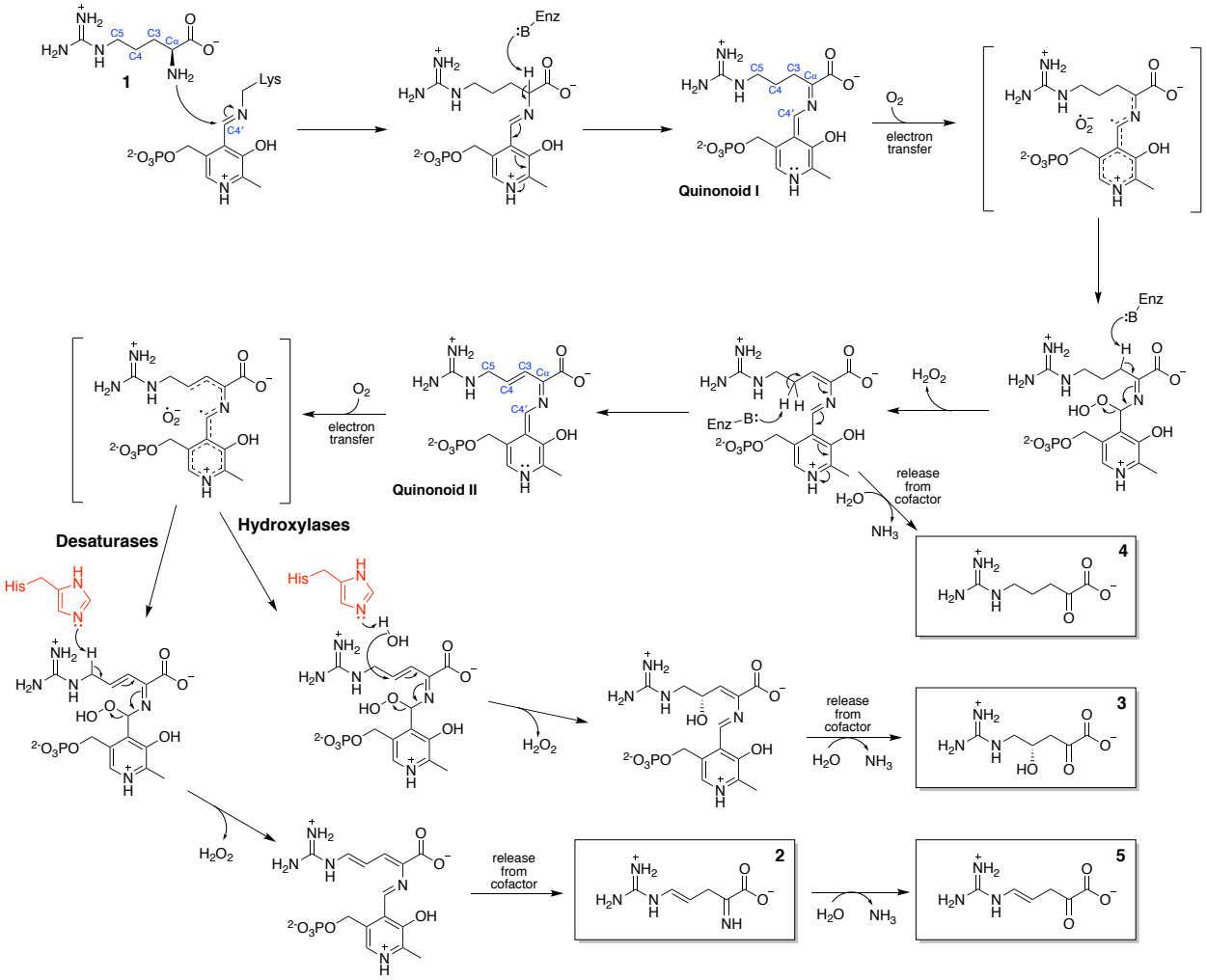


Figure 3.15: Proposed mechanism of arginine desaturases and hydroxylases. Relevant atom labels are shown in blue. Products observed via LC-MS $[M+H]^+$ signals are indicated with a box. The catalytic His residue is highlighted in red.

Chapter 4: Engineered production of indolmycin derivatives through semi-synthesis[8]

4.1 Introduction

As was reviewed in **Chapter 1**, antibiotic resistance threats have reignited interest in indolmycin, leading to the discovery of new activities against MRSA,[38] *Helicobacter pylori*[32] and *Plasmodium falciparum.*[39] Despite this revived interest, there have only been two attempts to synthesize or derivatize indolmycin since the discovery of these new bioactivities.[58,60] These attempts build upon the first indole-substituted derivatives, 5-hydroxy- and 5-methoxy-indolmycin, made biosynthetically from a natural producer, *S. griseus* ATCC 12648. Both of these derivatives showed mild improvements to bioactivity against *S. aureus* and *E. coli*.[67]

Continuing the early biosynthetic work from Hornemann *et al.* (1971),[65] the Ryan group identified the indolmycin gene cluster and elucidated the biosynthetic pathway, demonstrating that indolmycin (**6**) is assembled from tryptophan, **1** and methionine in a three-part process (**Figure 4.1**). In the first part, **1** is oxidized by Ind4 in an oxygen- and PLP-dependent reaction to **2**, which is then reduced by Ind5 and its chaperone, Ind6, to 4,5-dehydro-D-arginine (**7**). In the second part, which occurs parallel to the first, tryptophan (**8**) is deaminated by a PLP-dependent transaminase, giving indole pyruvate (**9**). Compound **9** is then methylated by Ind1, an *S*-adenosyl methionine (SAM)-dependent *C*-methyltransferase, to 3-methyl-indole pyruvate (**10**) which is reduced by Ind2, a nicotinamide adenine dinucleotide (NADH)-dependent ketone reductase, to form indolmycenic acid (**11**). Then, in the third part, compounds **7** and **11** are coupled in an ATP-dependent fashion by Ind3 and Ind6, resulting in an oxazolinone-cyclized molecule, *N*-desmethyl-indolmycin, which is finally *N*-methylated by Ind7, a SAM-dependent *N*-methyltransferase to form **6**.

Here, it is predicted that the indolmycin biosynthetic pathway could be used to make **6** in a heterologous host, such as *E. coli*. Use of *E. coli* as a host would allow pathway modifications and improvements to be made more easily than in other hosts, as the genetic manipulation tools available for *E. coli* are numerous. Based on previously observed acceptance of 5-hydroxy- and 5-methoxy-substituted derivatives,[67] it is also predicted that the biosynthetic enzymes may be able to accept a wider variety of indole-substituted derivatives of **8**, leading to novel indole-

[8] A version of this chapter has been submitted for publication.

substituted derivatives of **6**. Additionally, in order to expand the scope of analogs that could be made, enzymatic methods to generate indole-substituted derivatives of **8** *in vivo* from indole derivatives were explored, which are cheaper and more diverse than derivatives of **8**. Expansion of accessible derivatives of **6** will benefit the search for new antibiotics as some of these derivatives could show bioactivity against new pathogenic targets or they could help overcome the challenges that caused **6** to be forgotten so many years ago.

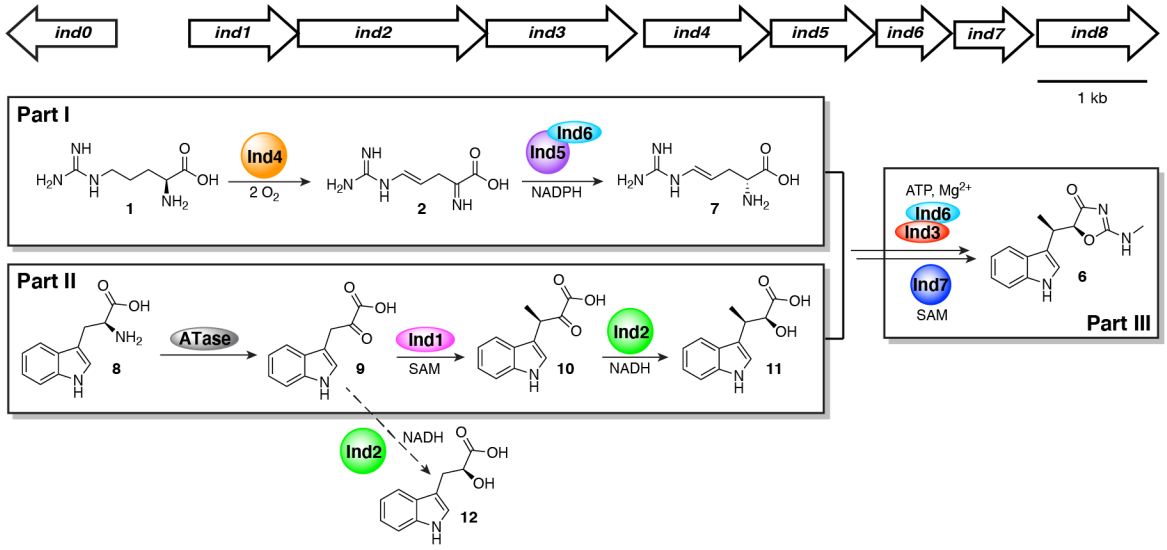


Figure 4.1: Indolmycin gene cluster and biosynthetic pathway from *S. griseus* ATCC 12648. A dashed arrow indicates a predicted reaction based on observed products by LC-MS analysis.

4.2 Materials and methods

4.2.1 General methods, materials and strains

Reagents were purchased from Sigma-Aldrich, Thermo Fisher Scientific Canada, NEB, Bio-Rad, Bio Basic Inc., Gold Biotechnology, and VWR International. The liquid chromatography column used was obtained from Agilent and the semi-preparative HPLC column was from Phenomex. Primers were purchased from IDT. Sequencing was done through the NAPS Unit in the Sequencing and Bioinformatics Consortium at the University of British Columbia. A list of strains and plasmids is given in **Table 4.1**.

Table 4.1: Strains and plasmids used in this study.

Name	Description	Source
Strains		
E. coli DH5α	General cloning host	Laboratory stock
E. coli BL21 (DE3)	Host for protein expression	Laboratory stock
Streptomyces griseus ATCC 12648	Wild-type indolmycin producer	ATCC
E. coli I120	*E. coli* carrying pI0COLA and pI1I2ET; tested for *in vivo* production of indolmycenic acid	This study
E. coli I1234670P5	*E. coli* carrying pP5I0COLA, pI1I2ET, pI3I4CDF and pI6I7ACYC; used for *in vivo* production of indolmycin	This study
E. coli I1234670TS	*E. coli* carrying pTSI0COLA, pI1I2ET, pI3I4CDF and pI6I7ACYC; used for *in vivo* production of IMA and IMA derivatives	This study
Plasmids		
pCOLAduet-1	Vector for protein expression in *E. coli* with two multiple cloning sites	Laboratory stock
pETduet-1	Vector for protein expression in *E. coli* with two multiple cloning sites	Laboratory stock
pCDFduet-1	Vector for protein expression in *E. coli* with two multiple cloning sites	Laboratory stock
pACYCduet-1	Vector for protein expression in *E. coli* with two multiple cloning sites	Laboratory stock
pET28a-*pel5*	Template for *pel5* amplification	Du *et al.*[113]
pSTB7 (ATCC 37845)	Vector carrying *trpS* from *Salmonella enterica*	ATCC
pI0COLA	Vector for cloning *pel5* and *trpS*	This study
pI1ET	Vector for cloning *ind2*	This study
pI3CDF	Vector for cloning *ind4*	This study
pI6ACYC	Vector for cloning *ind7*	This study
pP5I0COLA	Vector for *ind0* and *pel5* expression	This study
pI1I2ET	Vector for *ind1* and *ind2* expression	This study
pI3I4CDF	Vector for *ind3* and *ind4* expression	This study
pI6I7ACYC	Vector for *ind6* and *ind7* expression	This study
pTSI0COLA	Vector for *ind0* and trpS expression	This study

4.2.2 Cloning

The indolmycin biosynthetic genes (*ind0*, *ind1*, *ind2*, *ind3*, *ind4*, *ind6* and *ind7*) were PCR-amplified from *S. griseus* ATCC 12648 genomic DNA using the primers described in **Table 4.2**. Genes *ind1*, *ind3* and *ind6* were cloned using into the NcoI and HindIII restriction sites of MCS1 of pETduet-1, pCDFduet-1 and pACYCduet-1, respectively, using Gibson Assembly (NEB). Genes *ind0*, *ind2*, *ind4* and *ind7* were cloned using T4 ligase into MCS2 of pCOLAduet-1, pI1ET, pI3CDF and pI6ACYC, respectively, using the NdeI and XhoI restriction sites. The *ind5* homolog, *pel5* from *Paenibacillus elgii* B69, was PCR-amplified from a pET28a vector containing *pel5* described previously,[113] and was cloned using T4 ligase into the NcoI and HindIII restriction sites of MCS1 in pI0COLA. The *trpS* gene from *Salmonella enterica* was

PCR amplified from pSTB7 and cloned using Gibson Assembly into MCS1 of pI0COLA using the NcoI and HindIII restriction sites, replacing *pel5* for the production of **11** and its derivatives. The resulting vectors were transformed into electrocompetent *E. coli* DH5α by electroporation and selection on LB plates containing either kanamycin (50 μg/mL) for pCOLAduet vectors, ampicillin (100 μg/mL) for pETduet vectors, chloramphenicol (34 μg/mL) for pACYCduet vectors, or spectinomycin (100 μg/mL) for pCDFduet vectors. For co-expression of the genes, the appropriate vectors were co-transformed into electrocompetent *E. coli* BL21 (DE3) by electroporation and successful transformants were selected for on LB plates containing kanamycin (25 μg/mL), ampicillin (50 μg/mL), chloramphenicol (17 μg/mL), and spectinomycin (50 μg/mL). See **Table 4.1** for a description of each plasmid combination made in *E. coli*.

Table 4.2: Primers used in this study.

Primer Name	**Sequence (5′ to 3′)**	**Description**
Ind0_F	ATTAGTTAAGTATAAGAAGGAGATATACATATG ATCAAGCTGTCGGGAATCACCC	Cloning *ind0* into pCOLAduet-1
Ind0_R	CGGTTTCTTTACCAGACTCGAGTCAGCCCACTCC GGCGAGTT	Cloning *ind0* into pCOLAduet-1
NcoI-Pel5-F	ACTTTAATAAGGAGATATACCATGGATGAGAAT CATTTATCTGGATCAACCTACCTAC	Cloning *pel5* into pI0COLA
HindIII-Pel5-R	CTTAAGCATTATGCGGCCGCAAGCTTTCATTCGC TTGCCGGCCTCTTC	Cloning *pel5* into pI0COLA
pCOLA-TrpS-F	TAATTTTGTTTAACTTTAATAAGGAGATATACAT GACAACACTTCTCAACCCCTACTTTG	Cloning *trpS* into pI0COLA
pCOLA-TrpS-R	ATTATGCGGCCGCAAGCTTTATGCGCGGCTGGC GGC	Cloning *trpS* into pI0COLA
Ind1-3_F	AACTTTAAGAAGGAGATATACCATGACCAGGAC CGATTTCGCC	Cloning *ind1* into pETduet-1
Ind1-HindIII-R	GCATTATGCGGCCGCAAGCTTCATGAAGCCGGC CC	Cloning *ind1* into pETduet-1
NdeI-Ind2-F	TAGTTAAGTATAAGAAGGAGATATACATATGAA GCTGGACGACAAGAGAATTCTC	Cloning *ind2* into pI1ET
Ind2-XhoI-R	GTTTCTTTACCAGACTCGAGTCATAGGCCCTTGA TGCGACGGG	Cloning *ind2* into pI1ET
Ind3-F	TGTTTAACTTTAATAAGGAGATATACCATGAAG GGTGCGAGACAGGAG	Cloning *ind3* into pCDFduet-1
Ind3-HindIII-R	GCATTATGCGGCCGCAAGCTTTACGGGACCAGG CTGATGATCTGGTTG	Cloning *ind3* into pCDFduet-1
NdeI-Ind4-F2	GTTAAGTATAAGAAGGAGATATACATATGGAAC GGTTCAACAATCTGACG	Cloning *ind4* into pI3CDF
Ind4-XhoI-R2	GTTTCTTTACCAGACTCGAGTCATGAACCGGCGC CCTCG	Cloning *ind4* into pI3CDF
Ind6-F	TGTTTAACTTTAATAAGGAGATATACCATGTCGT TCGATAACCAGAACAGG	Cloning *ind6* into pACYCduet-1

Primer Name	Sequence (5′ to 3′)	Description
Ind6-HindIII-R	GCATTATGCGGCCGCAAGCTTCAGCCGACTCCG ATGTCTCCCAGTC	Cloning *ind6* into pACYCduet-1
NdeI-Ind7-F	GTTAAGTATAAGAAGGAGATATACATATGCACA CGGACTGGGAGAC	Cloning *ind7* into pI6ACYC
Ind7-XhoI-R	GTTTCTTTACCAGACTCGAGTCAGGCTGTTCCGC GCAC	Cloning *ind7* into pI6ACYC
pET Upstream	ATGCGTCCGGCGTAGA	Sequencing *ind1* in pI1ET
pACYCduetUP1	GGATCTCGACGCTCTCCCT	Sequencing *trpS* in pTSI0COLA, *pel5* in pP5I0COLA, *ind3* in pI3CDF, *ind6* in pI6ACYC
DuetDOWN1	GATTATGCGGCCGTGTACAA	Sequencing *trpS* in pTSI0COLA, *ind1* in pI1ET, *ind3* in pI3CDF, *ind6* in pI6ACYC
DuetUP2	TTGTACACGGCCGCATAATC	Sequencing *ind0* in pI0COLA, *ind2* in pI1I2ET, *ind4* in pI3I4CDF, *ind7* in pI6I7CYC
T7-term	GCTAGTTATTGCTCAGCGG	T7 terminator for sequencing *ind0* in pI0COLA, *ind2* in pI1I2ET, *ind4* in pI3I4CDF, *ind7* in pI6I7ACYC
TrpS-seq-F1	GCTCAAACGCTATCGGGATGTTTG	Sequencing *trpS* in pTSI0COLA
TrpS-seq-F2	GTACGCGAATCTGGTGTTCAATAACG	Sequencing *trpS* in pTSI0COLA
TrpS-seq-R1	CCATATAGATACGGCATTTCAGACCCAG	Sequencing *trpS* in pTSI0COLA
TrpS-seq-R2	CGTAGCGTTCCATCAGATTTCCCC	Sequencing *trpS* in pTSI0COLA

4.2.3 Expression and feeding

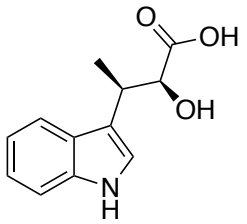
E. coli I1234670P5 and I1234670TS were initially grown in 50 mL of LB media containing kanamycin (25 μg/mL), ampicillin (50 μg/mL), chloramphenicol (17 μg/mL) and spectinomycin (50 μg/mL) for profiling metabolites, analyzing substrate scopes and optimizing the feeding procedure. For larger-scale production of **11** and its derivatives, *E. coli* I1234670TS was grown in two 500 mL portions of LB media containing the antibiotics mentioned above. Cultures were allowed to grow until 0.4-0.6 OD_{600} at 37°C with shaking at 200 rpm, after which the temperature and shaking were lowered to 16°C and 150 rpm, respectively, for 30 min prior to

addition of IPTG at 0.5 mM final concentration to induce expression of the biosynthetic genes. The cultures were incubated for 16 hours at 16°C and 150 rpm. Then, the temperature was raised to 30°C and the cultures were fed solid indoles or substituted indoles to a final concentration of 0.5 mM in each 500 mL culture once per day for two days. Compounds made by the system were harvested 24 hours after the final indole feeding. Use of *trpS* for improved acceptance of substituted indoles was tested with 5-fluorindole fed to 1 mM final concentration. Feeding optimization was done with 5-fluoroindole at 0.05 mM to 2 mM final concentration for one, two or three days to determine what conditions produced the most 5-fluoro-indolmycenic acid (**5F-11**).

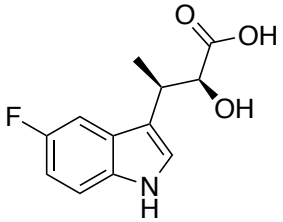
4.2.4 Extraction and purification of indolmycenic acid and its derivatives

Cell cultures were adjusted to pH 3-4 with HCl and then extracted with two volumes of ethyl acetate, which was gravity filtered and evaporated under low pressure. To analyze the metabolite profile and substrate scope, the resulting residues from small-scale test cultures were dissolved in 1 mL of methanol and analyzed by LC-MS using a 6120 Quadrupole LC-MS system (Agilent) operated in positive ion mode on an Agilent 5 TC18(2) 250 x 4.6 mm column using 85:15% water:ACN to 5:95% water:ACN (v/v) with 0.1% (v/v) formic acid run at 1 mL/min for 20 min. Initial tests for biosynthetic production of **6** (**Figure 4.6**) were analyzed by the same system above, but with a Phenomenex Luna C18(2), 5 μm, 4.6 mm ID × 250 mm column using 95:5% water:ACN to 0:100% water:ACN (v/v) with 0.1% (v/v) formic acid run at 0.6 mL/min for 20 min. When testing the optimal feeding conditions, the metabolites were analyzed by HPLC using a 1260 HPLC apparatus (Agilent) system using the Agilent 5 TC18(2) 250 x 4.6 mm column with a 75:25% water:ACN to 50:50% water:ACN (v/v) gradient containing 0.1% (v/v) formic acid run at 1 mL/min for 25 min. The **5F-11** peak eluted at 21.8 min and the area of the peak was used to determine the optimal feeding conditions based on the largest peak area. Testing omission of unnecessary indolmycin genes (**Figure 4.7**), was done with the HPLC system and column mentioned above but with a gradient of 85%:15% water:ACN to 60%:40% water:ACN (v/v) with 0.1% (v/v) formic acid run at 1 mL/min for 25 min. For large-scale growth, the resulting residue was dissolved in 6-7 mL of methanol for purification via semi-preparative HPLC. Semi-preparative HPLC was run with an 6120 Quadrupole LC-MS system (Agilent) using a Luna C18(2), 5 μm, 250 mm × 10 mm column (Phenomenex) using a gradient

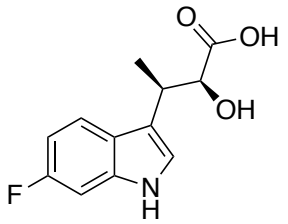
from 70:30% water:ACN (v/v) with 0.1% (v/v) formic acid to 42:58% water:ACN (v/v) at 3.0 mL/min in 12 min. The collected fraction containing the desired peak for **11**, or one of its derivatives, was verified by LC-MS using the conditions described above for analyzing the metabolite profile and substrate scope. Compound **11** and its derivatives were lyophilized and the masses were measured to determine the final amounts obtained. High resolution-mass spectrometry (HR-MS) and nuclear magnetic resonance (NMR) were used to confirm the identity of **11** and its derivatives. HR-MS was done on a Waters/Micromass LCT time of flight-mass spectrometer (TOF-MS) with electrospray ionization in positive ion mode. NMR analysis was done in dimethyl sulfoxide (DMSO)-d_6 with a Bruker Avance 400dir spectrometer.



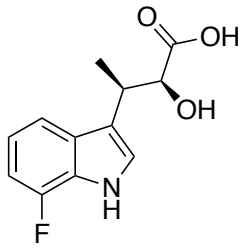
11 (**Appendix Figure B-1**, **Appendix Figure B-2**, **Appendix Figure B-3, Appendix Figure B-4,** and **Appendix Figure B-5**): ^{1}H NMR (400 MHz, DMSO-d_6) δ 10.79 (s, 1H), 7.54 (d, J = 7.8 Hz, 1H), 7.33 (d, J = 8.0 Hz, 1H), 7.14 (d, J = 2.4 Hz, 1H), 7.08 – 7.01 (m, 1H), 6.96 (t, J = 7.6 Hz, 1H), 4.19 (d, J = 4.6 Hz, 1H), 3.42 (qd, J = 7.0, 4.4 Hz, 1H), 1.25 (d, J = 7.1 Hz, 3H). ^{13}C NMR (101 MHz, DMSO) δ 175.26, 136.11, 126.49, 122.40, 120.73, 118.51, 118.14, 116.85, 111.36, 73.99, 34.02, 15.38.



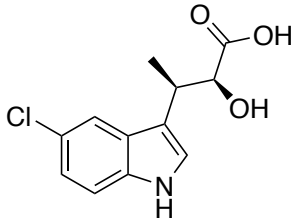
5F-11 (**Appendix Figure B-6** and **Appendix Figure B-7**): ^{1}H NMR (400 MHz, DMSO-d_6) δ 10.91 (s, 1H), 7.31 (dd, J = 8.8, 4.7 Hz, 1H), 7.26 (dd, J = 10.3, 2.6 Hz, 1H), 7.22 (d, J = 2.5 Hz, 1H), 6.88 (td, J = 9.1, 2.6 Hz, 1H), 4.13 (d, J = 4.7 Hz, 1H), 3.37 – 3.31 (m, 1H), 1.24 (d, J = 7.1 Hz, 3H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 175.14, 156.52 (d, J = 230.5 Hz), 132.77, 126.69 (d, J = 9.7 Hz), 124.63, 117.15 (d, J = 4.7 Hz), 112.20 (d, J = 9.7 Hz), 108.79 (d, J = 26.2 Hz), 103.35 (d, J = 23.2 Hz), 74.12, 34.09, 15.44.



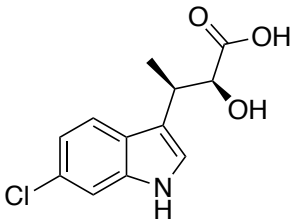
6F-11 (**Appendix Figure B-8** and **Appendix Figure B-9**): ^{1}H NMR (400 MHz, DMSO-d_6) δ 10.87 (s, 1H), 7.52 (dd, J = 8.7, 5.5 Hz, 1H), 7.16 – 7.05 (m, 2H), 6.82 (ddd, J = 9.7, 8.7, 2.4 Hz, 1H), 4.15 (d, J = 4.6 Hz, 1H), 3.38 (dt, J = 12.0, 6.9 Hz, 1H), 1.23 (d, J = 7.1 Hz, 3H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 175.16, 158.68 (d, J = 233.5 Hz), 135.88 (d, J = 12.6 Hz), 123.38, 122.99 (d, J = 3.4 Hz), 119.53 (d, J = 10.3 Hz), 117.13, 106.55 (d, J = 24.4 Hz), 97.20 (d, J = 25.4 Hz), 74.02, 34.01, 15.36.



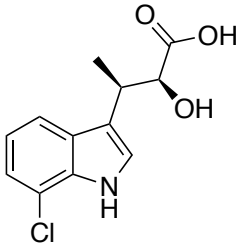
7F-11 (**Appendix Figure B-10**, **Appendix Figure B-11**, **Appendix Figure B-12**, **Appendix Figure B-13** and **Appendix Figure B-14**): ^{1}H NMR (400 MHz, DMSO-d_6) δ 11.30 (s, 1H), 7.36 (d, J = 7.8 Hz, 1H), 7.20 (d, J = 2.4 Hz, 1H), 6.98 – 6.83 (m, 2H), 4.17 (d, J = 4.6 Hz, 1H), 3.41 (qd, J = 7.0, 4.4 Hz, 1H), 1.25 (d, J = 7.1 Hz, 3H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 175.13, 149.31 (d, J = 242.4 Hz), 130.61 (d, J = 6.0 Hz), 123.80 (d, J = 13.0 Hz), 123.66, 118.47 (d, J = 6.1 Hz), 118.07 (d, J = 2.0 Hz), 114.83 (d, J = 3.3 Hz), 105.56 (d, J = 15.9 Hz), 73.89, 34.07, 15.37.



5Cl-11 (**Appendix Figure B-15** and **Appendix Figure B-16**): ^{1}H NMR (400 MHz, DMSO-d_6) δ 11.01 (s, 1H), 7.55 (d, J = 2.0 Hz, 1H), 7.34 (d, J = 8.6 Hz, 1H), 7.21 (d, J = 2.3 Hz, 1H), 7.04 (dd, J = 8.6, 2.1 Hz, 1H), 4.08 (d, J = 4.3 Hz, 1H), 3.40 – 3.33 (m, 1H), 1.21 (d, J = 7.1 Hz, 3H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 175.29, 134.65, 127.75, 124.44, 122.91, 120.71, 118.05, 117.22, 112.95, 74.29, 34.03, 15.35.



6Cl-11 (**Appendix Figure B-17** and **Appendix Figure B-18**): ^{1}H NMR (400 MHz, DMSO-d_6) δ 10.95 (s, 1H), 7.54 (d, J = 8.5 Hz, 1H), 7.36 (d, J = 1.9 Hz, 1H), 7.18 (d, J = 2.3 Hz, 1H), 6.97 (dd, J = 8.4, 1.9 Hz, 1H), 4.12 (d, J = 4.5 Hz, 1H), 3.39 (qd, J = 7.0, 4.2 Hz, 1H), 1.23 (d, J = 7.1 Hz, 3H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 175.76, 137.06, 126.05, 125.94, 124.19, 120.59, 119.04, 117.95, 111.50, 74.66, 34.54, 15.91.



7Cl-11 (**Appendix Figure B-19** and **Appendix Figure B-20**): ^{1}H NMR (400 MHz, DMSO-d_6) δ 11.17 (s, 1H), 7.52 (d, J = 7.9 Hz, 1H), 7.21 (d, J = 2.5 Hz, 1H), 7.13 (d, J = 7.6 Hz, 1H), 6.98 (t, J = 7.8 Hz, 1H), 4.16 (d, J = 4.6 Hz, 1H), 3.41 (dt, J = 7.1, 3.8 Hz, 1H), 1.25 (d, J = 7.1 Hz, 3H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 175.10, 132.82, 128.53, 123.82, 120.28, 119.26, 118.33, 117.73, 115.81, 73.97, 34.09, 15.44.

4.2.5 Synthesis of indolmycin and indolmycin derivatives

The synthetic method for making **6** and its derivatives from **11** and its derivatives was adapted from literature methods to make **6** (**Figure 4.2**).[58,60] To form the indolmycenic acid ethyl ester (**13**), or one of its derivatives, 85 μL of acetyl chloride was mixed into 300 μL of ethanol on ice for 30 min. Compound **11**, or one of its derivatives, was dissolved in 200 μL of ethanol and added slowly to the acetyl chloride and ethanol mixture over 2-3 min. The resulting mixture was then stirred at room temperature for six hours. The solvent was rotary evaporated and the resulting residue was dissolved in 400 μL ethyl acetate, washed with 280 μL saturated sodium bicarbonate and 225 μL of brine, and then dried over anhydrous sodium sulfate before the solvent was speed vacuumed, leaving a residue of **13**, or one of its derivatives.

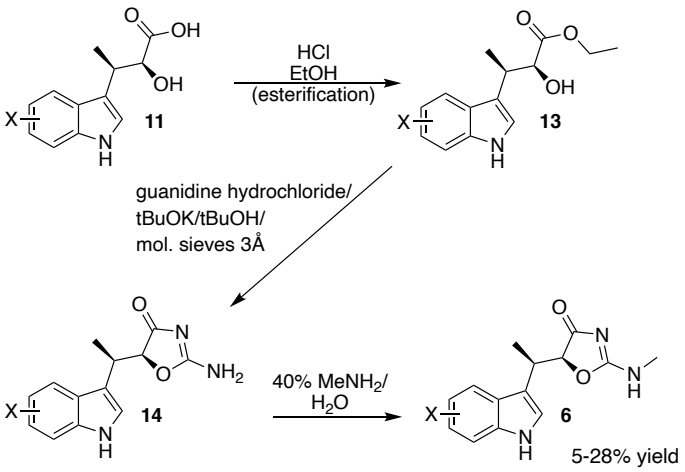


Figure 4.2: Synthetic scheme to **6** from **11**, adapted from literature methods.[58,60]

The oxazolinone ring was cyclized to form *N*-desmethyl-indolmycin (**14**), or one of its derivatives, using 20 mg of guanidine, 20 mg of molecular sieves 3Å, and 240 µL of KOtBu (1 M in THF), which were mixed with 80 µL *tert*-Butanol at room temperature for three days. The residue of **13**, or one of its derivatives, from the esterification step was dissolved in 160 µL of *tert*-butanol (also dried with sieves for three days for derivatives, but not for initial synthesis of **6**) and added to the guanidine/sieves/KOtBu mixture, which was then mixed for 7.5 hours at room temperature. The reaction was stopped with 800 µL of cold, saturated ammonium chloride. The liquids partitioned into two layers, which were separated from each other after centrifugation. The aqueous ammonium chloride layer was washed with 500 µL of ethyl acetate, which was combined with the organic layer. Then the organic layer was washed with 200 µL of 5% sodium bicarbonate and dried with anhydrous sodium sulfate before the solvent was speed vacuumed, leaving a residue of **14**, or one of its derivatives.

Compound **6** and its derivatives were formed by an amine exchange of **14** and derivatives of **14** obtained from the previous reaction. Compound **14**, or one of its derivatives, was dissolved in 65 µL of 40% (w/v) methylamine and incubated at 4°C for five hours. The solvent was evaporated by speed vacuum without heat and the resulting residue was dissolved in 500 µL of methanol for purification by semi-preparative HPLC, as described above for the purification of **11** and its derivatives. To further purify **6** from residual **11** after an initial semi-preparative HPLC purification, semi-preparative HPLC was run again, but at 48%:52% to 54%:46%

ACN:water (v/v) with 0.1% (v/v) formic acid over 12 min. Fractions containing **6**, or one of its derivatives, were verified by LC-MS and lyophilized to measure the final mass and determine the amount obtained. Compound **6** and its derivatives were characterized by HR-MS and NMR to confirm their identity. HR-MS was done on a Waters/Micromass LCT TOF-MS with electrospray ionization in positive ion mode. NMR analysis was done in DMSO-d6 with a Bruker Avance 600 MHz spectrometer or a Bruker Avance 400dir spectrometer. Tautomeric isomers of **6** and its derivatives were observed from all the NMR spectra in a 2:1 ratio, favouring the natural tautomer (**Figure 4.3**), which was confirmed by a varied temperature NMR experiment was done on **5F-6** between 25°C and 115°C with a Bruker Avance 400inv spectrometer (**Appendix Figure B-31**). Observation of these isomers has been reported previously, also in a 2:1 ratio favouring the natural tautomer, however the bioactivity of the non-natural tautomer is not known as the individual tautomers have not been isolated..[31,58,60]

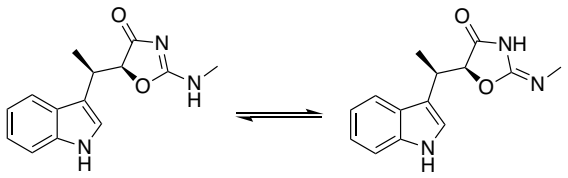
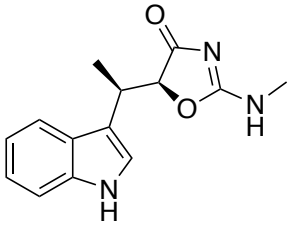
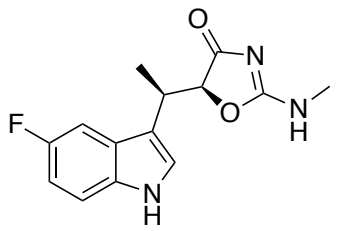


Figure 4.3: Predicted tautomeric isomers of indolmycin observed from ^{1}H NMR spectra. The tautomer on the left is the natural tautomer.

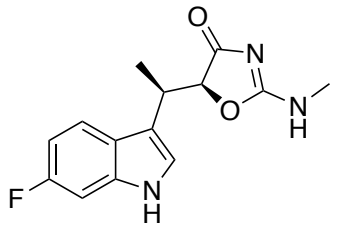


6 (**Appendix Figure B-21**, **Appendix Figure B-22**, **Appendix Figure B-23**, **Appendix Figure B-24** and **Appendix Figure B-25**): ^{1}H NMR (400 MHz, DMSO-d_6) δ 10.91 (s, 1H), 8.75 (s, 1/3H), 8.68 – 8.60 (m, 2/3H), 7.58 (dd, J = 8.0, 4.3 Hz, 1H), 7.35 (d, J = 8.1 Hz, 1H), 7.18 (d, J = 2.4 Hz, 1/3H), 7.15 (d, J = 2.4 Hz, 2/3H), 7.08 (ddt, J = 9.1, 7.8, 1.7 Hz, 1H), 6.99 (t, J = 7.4 Hz, 1H), 4.94 (d, J = 2.6 Hz, 2/3H), 4.91 (d, J = 2.7 Hz, 1/3H), 3.66 – 3.51 (m, 1H), 2.80 (d, J = 4.5 Hz, 2H), 2.77 (s, 1H), 1.25 (d, J = 7.3 Hz, 1H), 1.19 (d, J = 7.1 Hz, 2H). ^{13}C NMR (101

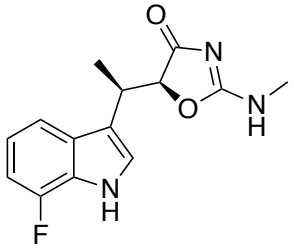
MHz, DMSO-d_6) δ 186.59, 175.68, 136.20, 126.07, 122.44, 121.01, 118.52, 118.46, 115.32, 111.46, 85.35, 31.69, 28.81, 13.56.



5F-6 (**Appendix Figure B-26**, **Appendix Figure B-27**, **Appendix Figure B-28**, **Appendix Figure B-29**, **Appendix Figure B-30** and **Appendix Figure B-31**): ^{1}H NMR (400 MHz, DMSO-d_6) δ 11.02 (d, J = 3.1 Hz, 1H), 8.70 (s, 1/3H), 8.67 – 8.59 (m, 2/3H), 7.38 – 7.19 (m, 3H), 6.96 – 6.85 (m, 1H), 4.93 (d, J = 2.7 Hz, 2/3H), 4.90 (d, J = 2.8 Hz, 1/3H), 3.63 – 3.47 (m, 1H), 2.79 (d, J = 4.4 Hz, 2H), 2.77 (s, 1H), 1.25 (d, J = 7.2 Hz, 1H), 1.19 (d, J = 7.2 Hz, 2H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 187.28, 176.33, 157.27 (d, J = 230.8 Hz), 133.42 (d, J = 8.2 Hz), 126.92 (d, J = 9.8 Hz), 125.24, 116.18 (d, J = 4.9 Hz), 112.97 (d, J = 9.6 Hz), 109.75 (d, J = 26.1 Hz), 104.02 (d, J = 23.4 Hz), 85.84, 32.29, 29.42, 14.32.

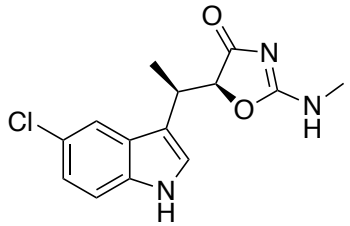


6F-6 (**Appendix Figure B-32** and **Appendix Figure B-33**): ^{1}H NMR (600 MHz, DMSO-d_6) δ 10.97 (s, 1H), 8.69 (s, 1/3H), 8.64 – 8.59 (m, 2/3H), 7.57 – 7.51 (m, 1H), 7.18 – 7.05 (m, 2H), 6.85 – 6.77 (m, 1H), 4.89 (d, J = 2.7 Hz, 2/3H), 4.86 (d, J = 2.8 Hz, 1/3H), 3.60 – 3.48 (m, 1H), 2.76 (d, J = 4.7 Hz, 2H), 2.73 (d, J = 4.1 Hz, 1H), 1.20 (d, J = 7.2 Hz, 1H), 1.14 (d, J = 7.0 Hz, 2H). ^{13}C NMR (151 MHz, DMSO-d_6) δ 186.67, 175.76, 158.79 (d, J = 234.0 Hz), 136.04 (d, J = 12.7 Hz), 123.09 (d, J = 3.3 Hz), 122.98, 119.68 (d, J = 10.1 Hz), 115.58, 106.91 (d, J = 24.3 Hz), 97.36 (d, J = 25.3 Hz), 85.25, 31.67, 28.83, 13.62.

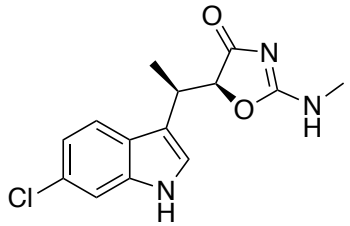


7F-6 (**Appendix Figure B-34** and **Appendix Figure B-35**): ^{1}H NMR (600 MHz, DMSO-d_6) δ 11.44 (s, 1H), 8.73 (s, 1/3H), 8.68 – 8.64 (m, 2/3H), 7.46 – 7.32 (m, 1H), 7.27 – 7.14 (m, 1H), 7.01 – 6.88 (m, 2H), 4.94 (d, J = 2.6 Hz, 2/3H), 4.90 (d, J = 2.8 Hz, 1/3H), 3.65 – 3.51 (m, 1H), 2.79 (d, J = 4.6 Hz, 2H), 2.76 (s, 1H), 1.25 (d, J = 7.2 Hz, 1H), 1.18 (d, J = 7.1 Hz,

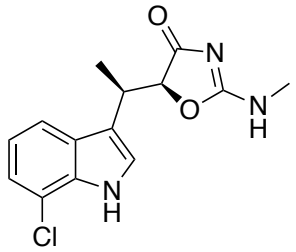
2H). ^{13}C NMR (151 MHz, DMSO-d_6) δ 186.64, 175.77, 149.25 (d, J = 242.6 Hz), 130.15 (d, J = 5.7 Hz), 123.85 (d, J = 19.9 Hz), 123.69, 118.83 (d, J = 6.2 Hz), 116.56, 114.92, 105.88 (d, J = 16.0 Hz), 85.16, 31.71, 28.84, 13.67.



5Cl-6 (**Appendix Figure B-36** and **Appendix Figure B-37**): ^{1}H NMR (600 MHz, DMSO-d_6) δ 11.14 (s, 1H), 8.70 (s, 1/3H), 8.67 – 8.62 (m, 2/3H), 7.64 – 7.57 (m, 1H), 7.39 – 7.31 (m, 1H), 7.27 – 7.20 (m, 1H), 7.10 – 7.01 (m, 1H), 4.94 (d, J = 2.8 Hz, 2/3H), 4.89 (d, J = 2.8 Hz, 1/3H), 3.64 – 3.51 (m, 1H), 2.81 – 2.76 (m, 3H), 1.27 (d, J = 7.2 Hz, 1H), 1.18 (d, J = 7.1 Hz, 2H). ^{13}C NMR (151 MHz, DMSO-d_6) δ 186.66, 175.72, 134.62, 127.24, 124.42, 123.23, 121.02, 117.82, 115.21, 113.01, 85.13, 31.51, 28.84, 13.76.



6Cl-6 (**Appendix Figure B-38** and **Appendix Figure B-39**): ^{1}H NMR (400 MHz, DMSO-d_6) δ 11.07 (s, 1H), 8.70 (s, 1/3H), 8.65 – 8.59 (m, 2/3H), 7.64 – 7.51 (m, 1H), 7.42 – 7.30 (m, 1H), 7.25 – 7.10 (m, 1H), 7.04 – 6.91 (m, 1H), 4.92 (d, J = 2.7 Hz, 2/3H), 4.89 (d, J = 3.0 Hz, 1/3H), 3.63 – 3.48 (m, 1H), 2.78 (d, J = 4.1 Hz, 2H), 2.76 (s, 1H), 1.25 (d, J = 7.2 Hz, 1H), 1.19 (d, J = 7.1 Hz, 2H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 186.58, 175.71, 136.56, 125.74, 124.93, 123.65, 120.06, 118.75, 115.60, 111.02, 85.17, 31.60, 28.80, 13.63.



7Cl-6 (**Appendix Figure B-40** and **Appendix Figure B-41**): ^{1}H NMR (400 MHz, DMSO-d_6) δ 11.30 (s, 1H), 8.69 – 8.63 (m, 1H), 7.62 – 7.50 (m, 1H), 7.26 – 7.09 (m, 2H), 7.05 – 6.95 (m, 1H), 4.94 (d, J = 2.7 Hz, 2/3H), 4.91 (d, J = 2.8 Hz, 1/3H), 3.67 – 3.48 (m, 1H), 2.79 (d, J = 4.2 Hz, 2H), 2.75 (s, 1H), 1.27 (d, J = 7.2 Hz, 1H), 1.19 (d, J = 7.2 Hz, 2H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 186.59, 175.73, 132.92, 128.12, 123.81, 120.58, 119.55, 117.81, 116.78, 115.85, 85.14, 31.70, 28.82, 13.81.

4.2.6 Bioactivity assays[9]

Disk diffusion assays were done on Mueller-Hinton Agar plates with MRSA (*Staphylococcus aureus* ATCC 33591). Each derivative was dissolved in water with 0.5% DMSO and deposited on the disks. Rifampicin (1 μg) was used a positive control and 0.5% DMSO was used as a negative control. Each disk contained either 10 μg of **6**, 20 μg of **6F-6** or **7F-6**, or 30 μg of **5F-6**, **5Cl-6**, **6Cl-6**, or **7Cl-6** and was deposited onto the petri dishes containing a lawn of the MRSA culture. The petri dishes were allowed to grow overnight, after which zones of inhibition were visible. The minimum inhibitory concentration (MIC) measurements were done by diluting an overnight culture of *S. aureus* ATCC 33591 to ~ 1 x 10^5 cells/mL with Mueller-Hinton Broth (MHB). Various dilutions of **1** and its derivatives and the positive control, rifampicin, were prepared. DMSO was used as a negative control. Concentrations ranged from 50 μg/mL to 0.128 ng/mL for **1**, **5F-6**, **6F-6** and **7F-6** and from 200 μg/mL to 0.512 ng/mL for **5Cl-6**, **6Cl-6** and **7Cl-6**. The rifampicin sample concentrations ranged from 10 μg/mL to 0.0256 ng/mL. The 100 μL of *S. aureus* cells and 10 μL of each diluted sample were combined in a 96-well plate and diluted to 200 μL with 90 μL MHB. Each sample concentration for each compound was tested in triplicate. The 96-well plate was covered and incubated at 37°C for 18 h before measuring the OD at 600 nm using a Molecular Devices FilterMax F5 Multi-Mode Microplate Reader. Readings from blank samples were subtracted from each sample reading to remove background signal and the percent of cells alive was calculated in comparison to the DMSO negative control. The MIC_{50} values were determined from the percent alive using the Quest Graph IC50 Calculator online tool from AAT Bioquest.[226] Docking experiments to test analog binding to a TrpRS was done using AutoDock Vina[194] and the structures for **6** and its derivatives were generated using Phenix eLBOW.[178] The TrpRS model used for docking was from *Bacillus stearothermophilus* (PDB: 5DK4) and the cocrystallized ligands, ATP, Mg^{2+} and indolmycin, were removed from the model before docking.

[9] Disk diffusion and MIC assays were performed by J. C.

4.3 Results

4.3.1 Production of indolmycin

4.3.1.1 Heterologous expression of indolmycin genes

Armed with the previously elucidated biosynthetic pathway for **6**, an attempt to create an *in vivo* system to make **6** in *E. coli* was made. First, all necessary genes were cloned into four plasmids and were co-expressed in *E. coli* BL21 (DE3) (**Figure 4.4**). This strain was given the name *E. coli* I1234670P5 (**Table 4.1**). The genes needed to produce indolmycin in *E. coli* were *ind1, ind2, ind3, ind4, ind6, ind7, ind0* and *pel5*, a gene homologous to *ind5* from *Paenibacillus elgii* B69.[113,174] Unfortunately, only a small amount of **6** ($[M+H]^+$ = 258 *m/z*; **Figure 4.5**) was produced (~170 μg/L of culture) and could not be improved (**Figure 4.6**). However, this construct did produce more substantial amounts of **11** ($[M+H]^+$ = 220 *m/z*; **Figure 4.5**) at 40-50 mg/L of culture, along with a shunt product, *C*-desmethyl-indolmycenic acid (**12**; $[M+H]^+$ = 206 *m/z*; **Figure 4.5**).

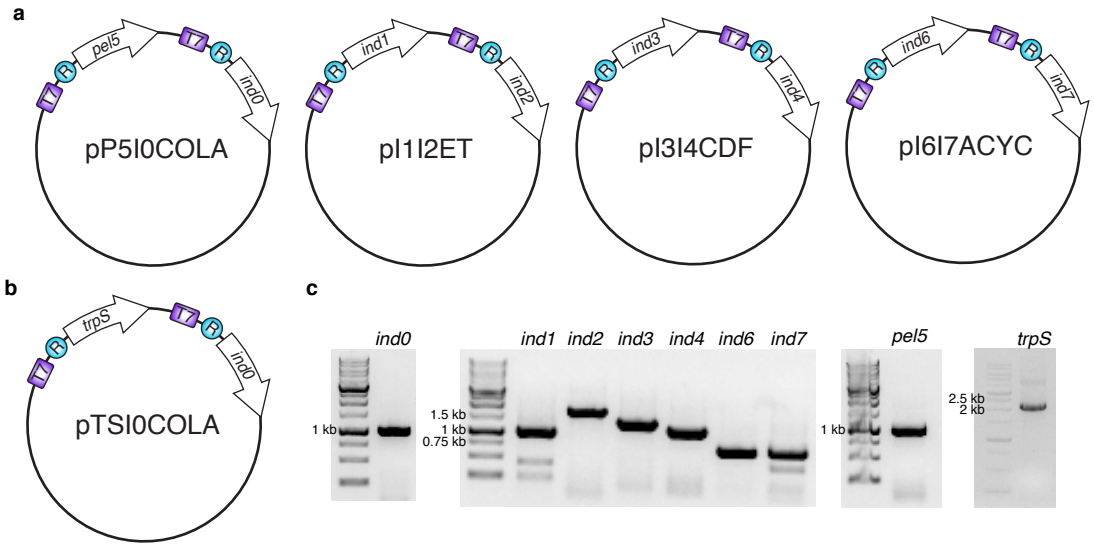


Figure 4.4: Cloning genes for *in vivo* production of **6**, **11** and derivatives of **11**. (**a**) Plasmids made and used in *E. coli* I1234670P5 to make **6** *in vivo*. (**b**) Plasmid used in *E. coli* I1234670TS to replace pP5I0COLA from *E. coli* I1234670P5 for production of **6** and its derivatives. (**c**) Agarose gels of PCR amplifications for each gene cloned into each of the five plasmids. Expected sizes for each gene are *ind0*: 1018 base pairs; *ind1*: 1032 base pairs; *ind2*: 1761 base pairs; *ind3*: 1406 base pairs; *ind4*: 1198 base pairs; *ind6*: 737 base pairs; *ind7*: 760 base pairs; *pel5*: 1011 base pairs; *trpS*: 2050 base pairs.

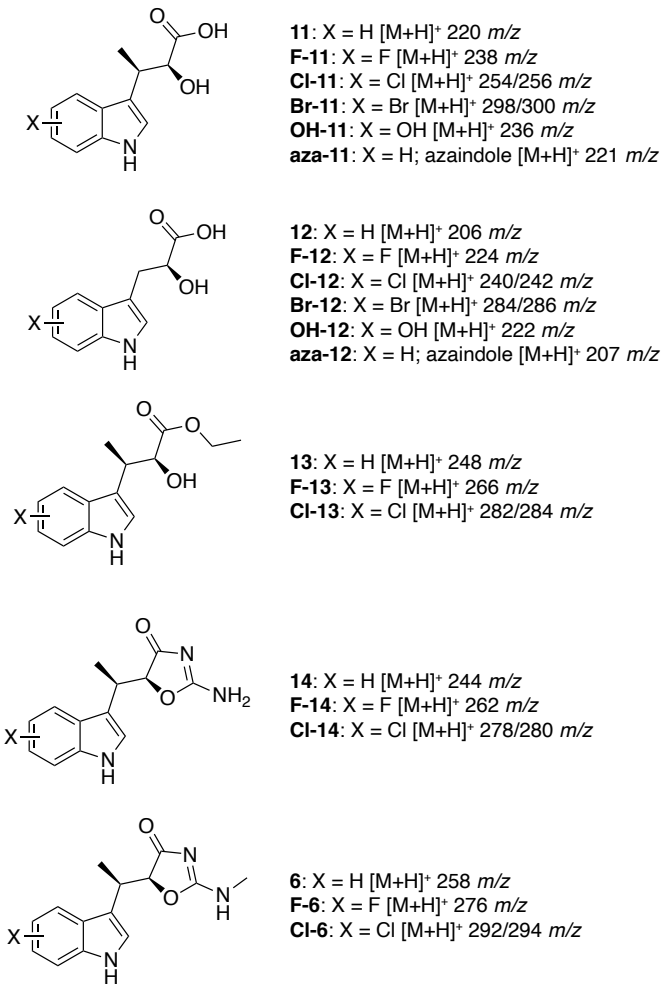


Figure 4.5: Summary of compound $[M+H]^+$ ions.

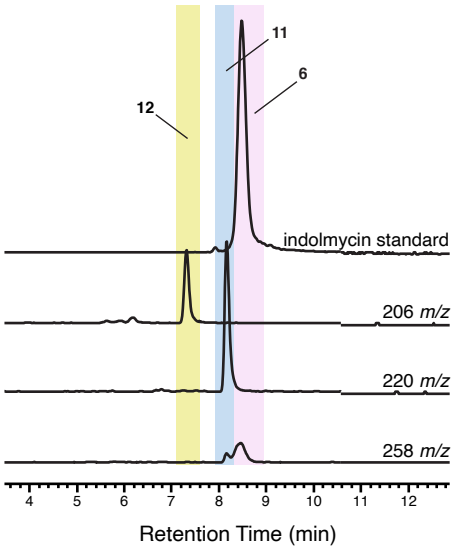


Figure 4.6: LC-MS analysis of observed products from *E. coli* I1234670P5. Peaks are labelled with the corresponding compound number and shown in a coloured box.

4.3.1.2 Indolmycenic acid purification and indolmycin synthesis

Compound **11** has often been a focus of total synthetic efforts towards **6**, as it is a key chiral intermediate.[31,55–60] From **11**, a common method of producing indolmycin involved a three-step synthesis, in which **11** was esterified, the oxazolinone ring was cyclized using guanidine, and the exocyclic nitrogen was methylated (**Figure 4.2**).[58,60] Since production of **11** was much higher than for **6** from *E. coli* I1234670P5, a semi-synthetic method of obtaining **6** was pursued for use in a three-step synthesis to produce **6**. Now focused on **11** rather than **6**, an attempt was made to remove extraneous genes from the biosynthetic platform but was unsuccessful and the changes resulted in reduced amounts of **11** (**Figure 4.7**). Therefore, the full four-gene construct was employed toward production of **11**. Then, the three-step synthesis was used (**Figure 4.2**),[58,60] in which purified **11** was esterified to make the ethyl ester (**13**; $[M+H]^+$ = 248 *m/z*; **Figure 4.5**), cyclized to give *N*-desmethyl-indolmycin (**14**; $[M+H]^+$ = 244 *m/z*; **Figure 4.5**), and methylated at the exocyclic nitrogen to give **6** ($[M+H]^+$ = 258 *m/z*; **Figure 4.5**) (**Figure 4.8**). A total of 2.6 mg of **6** was made with 5.0% yield from **11**. The low yield may be attributed to a high amount of **11** regenerated in the oxazolinone-cyclization step (**Figure 4.8b** and **Figure 4.9**), likely due to insufficient drying of the *tert*-butanol solvent.

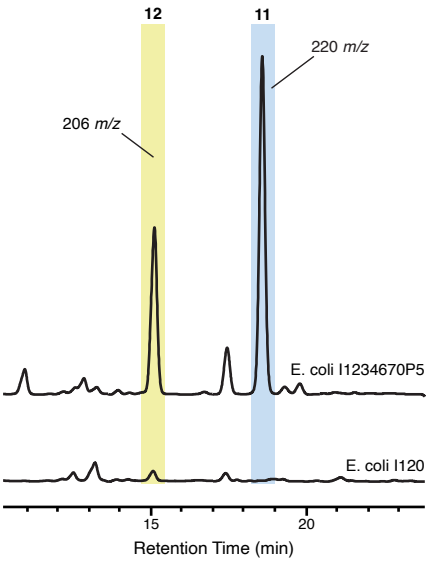


Figure 4.7: HPLC comparison of *E. coli* I120 and *E. coli* I1234670P5 metabolites to understand the effect of removing genes presumably not needed for production of **11**. Relevant peaks identified as indolmycin pathway metabolites are indicated with a coloured box, compound number and their corresponding *m/z* from LC-MS analysis.

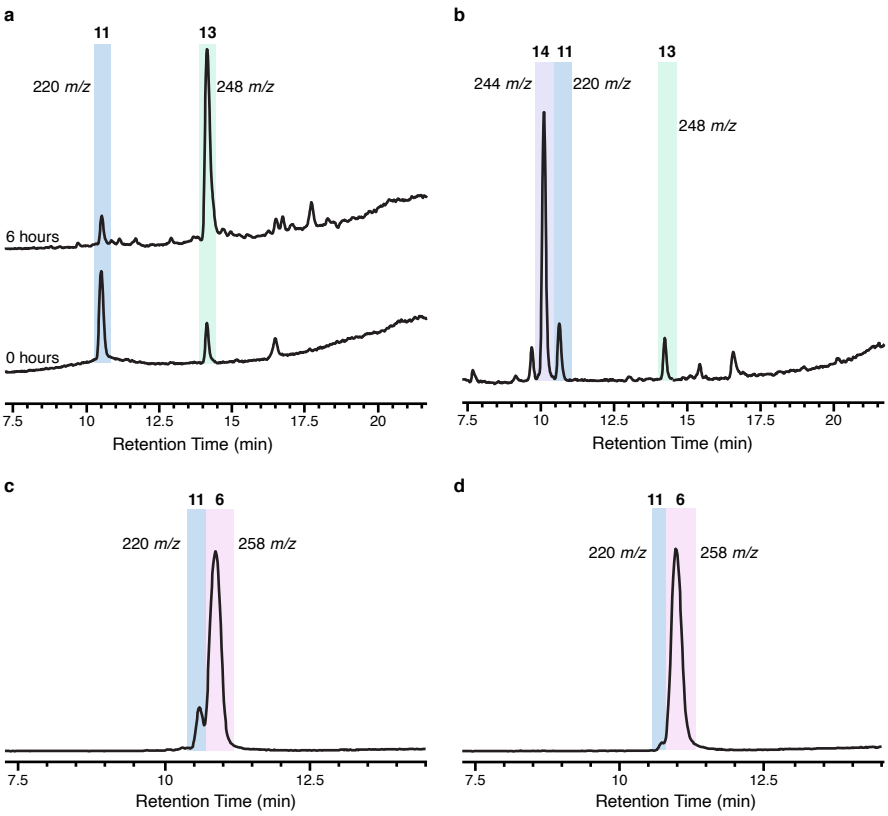


Figure 4.8: LC-MS analysis for synthesis and purification of **6**. (**a**) Total ion chromatograms of the esterification at the beginning (0 hours) and the end (6 hours) of the reaction. (**b**) Total ion chromatograms of the oxazolinone-cyclization reaction. (**c**) Total ion chromatogram of the *N*-methylation reaction after initial purification by semi-preparative HPLC. (**d**) Total ion chromatogram of **6** after final purification by semi-preparative HPLC. Compounds are indicated by a coloured box and numbered. $[M+H]^+$ signals corresponding to each observed compound are indicated beside each coloured box.

Table 4.3: Amounts of **11**-derivatives and **6**-derivatives obtained. Yields given are for synthetic steps from **11** only. Amounts of **11** are from 1 L of bacterial culture. Low yield for underivatized **6** is attributed to improper drying of solvent during production of **14**. Synthetic scheme is given in **Figure 4.2**.

	11 and derivatives (mg)	**6** and derivatives (mg)	% yield
indole	44.1	2.6	5.0
5-fluoroindole	26.9	7.8	25
6-fluoroindole	15.1	2.6	15
7-fluoroindole	24.9	3.1	11
5-chloroindole	9.1	1.7	16
6-chloroindole	10.3	3.3	28
7-chloroindole	10.5	2.6	22

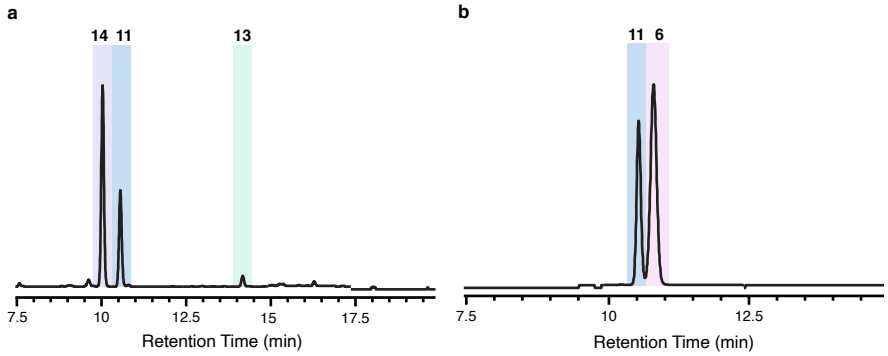


Figure 4.9: UV-Vis traces obtained during LC-MS analysis of synthetic steps showing regeneration of **11**. Absorbance measurements were made at 280 nm. Traces show high levels of **11** regenerated after the oxazolinone-cyclization step in (**a**) and high levels of **11** remaining after *N*-methylation reaction in (**b**).

4.3.2 Production of indolmycin derivatives

4.3.2.1 Use of TrpS for improved indole incorporation

In order to make derivatives of **6**, it was desirable to feed indole-substituted precursors, such as derivatives of **8**, to *E. coli* I1234670P5. However, the lack of accessible derivatives of **8** inspired the use of indole derivatives, which are more widely accessible. The TrpS from *S. enterica* has been previously shown to be able to accept a wide variety of indole derivatives and convert them to derivatives of **8**, *in vivo*, using L-serine.[88] This TrpS was incorporated into the biosynthetic platform to make derivatives of **11** to be used as precursors to synthesize derivatives

of **6**. *pel5* was substituted for *trpS* in this biosynthetic platform (**Figure 4.4b**), which was subsequently named *E. coli* I1234670TS. Improved production of fluorinated metabolites, 5-fluoro-indolmycenic acid (**5F-11**; $[M+H]^+$ = 238 *m/z*; **Figure 4.5**) and 5-fluoro-*C*-desmethyl-indomycenic acid (**5F-12**; $[M+H]^+$ = 224 *m/z*; **Figure 4.5**), was observed from *E. coli* I1234670TS fed 5-fluoroindole (**Figure 4.10a**). 5-fluoroindole was also used to optimize the feeding conditions, which showed that **5F-11** amounts were optimal when fed 0.5 mM of 5-fluoroindole per day over two days (**Figure 4.11**).

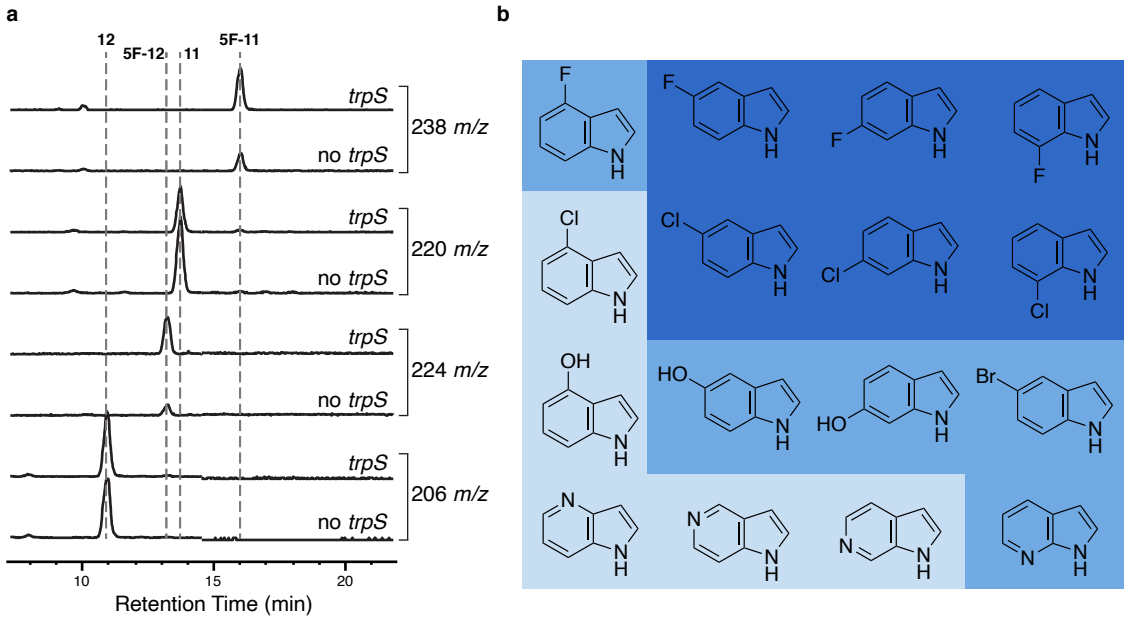


Figure 4.10: TrpS addition to the biosynthetic platform allows incorporation of substituted indoles into **11**. (**a**) LC-MS analysis of strains with and without *trpS* (*E. coli* I1234670TS and *E. coli* I1234670P5, respectively) when fed 5-fluoroindole. Incorporation of fluorinated compounds into the **6**-pathway metabolites (**5F-12** is $[M+H]^+$ = 224 *m/z* and **5F-11** is $[M+H]^+$ = 238 *m/z*) is improved in *E. coli* I1234670TS. (**b**) Indole derivatives tested for incorporation by *E. coli* I1234670TS. Dark blue shows indoles that were detected and further purified, medium blue was possibly detected by LC-MS, but was not further verified, and light blue did not show detectable derivatives of **11**.

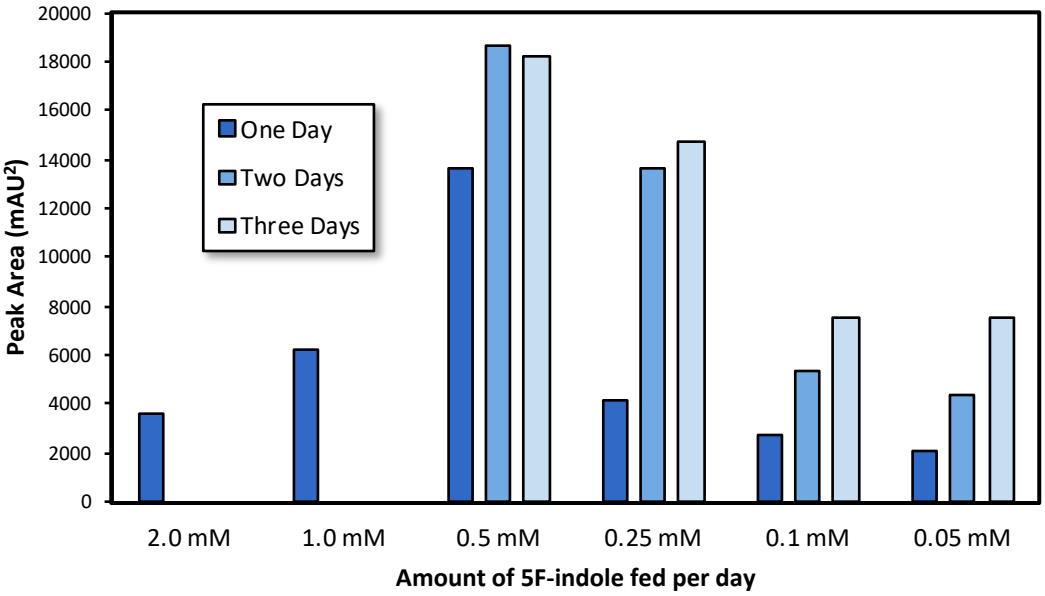


Figure 4.11: HPLC analysis of **5F-11** amounts under different feeding conditions. Amount of **5F-11** was approximated based on peak area measured at 280 nm. Amount of **5F-11** produced for 2.0 mM and 1.0 mM was only tested on the first day.

4.3.2.2 Scope of indole incorporation

To determine the scope of indole derivatives accepted by the biosynthetic platform, a variety of indoles were fed to *E. coli* I1234670TS and the production of **11** and its derivatives was observed. Out of the indoles tested, the fluorinated and chlorinated indoles substituted at the 5-, 6- and 7-positions were the best accepted by the biosynthetic platform (**Figure 4.10b** and **Figure 4.12** and **Figure 4.13**). Lower acceptance of indoles substituted at the 4-position may be due to a steric hindrance, as 4-fluoroindole was moderately accepted, while 4-chloroindole was not utilized at all. Although there was evidence that some of the azaindoles and hydroxyindoles were incorporated, further work would need to be done to confirm, optimize and scale up the purification of these compounds (**Figure 4.13**). It is possible that the hydroxyindoles and azaindoles were less accepted because the mechanism of indole transport into the cells may be disrupted by their different chemical properties or they may have bound unproductively in the active site of TrpS and were unable to be converted to derivatives of **8**. Additionally, reduced nucleophilicity of the azaindoles in particular might have had an effect on the ability for TrpS to convert the indoles to derivatives of **8**, as has been observed previously.[85]

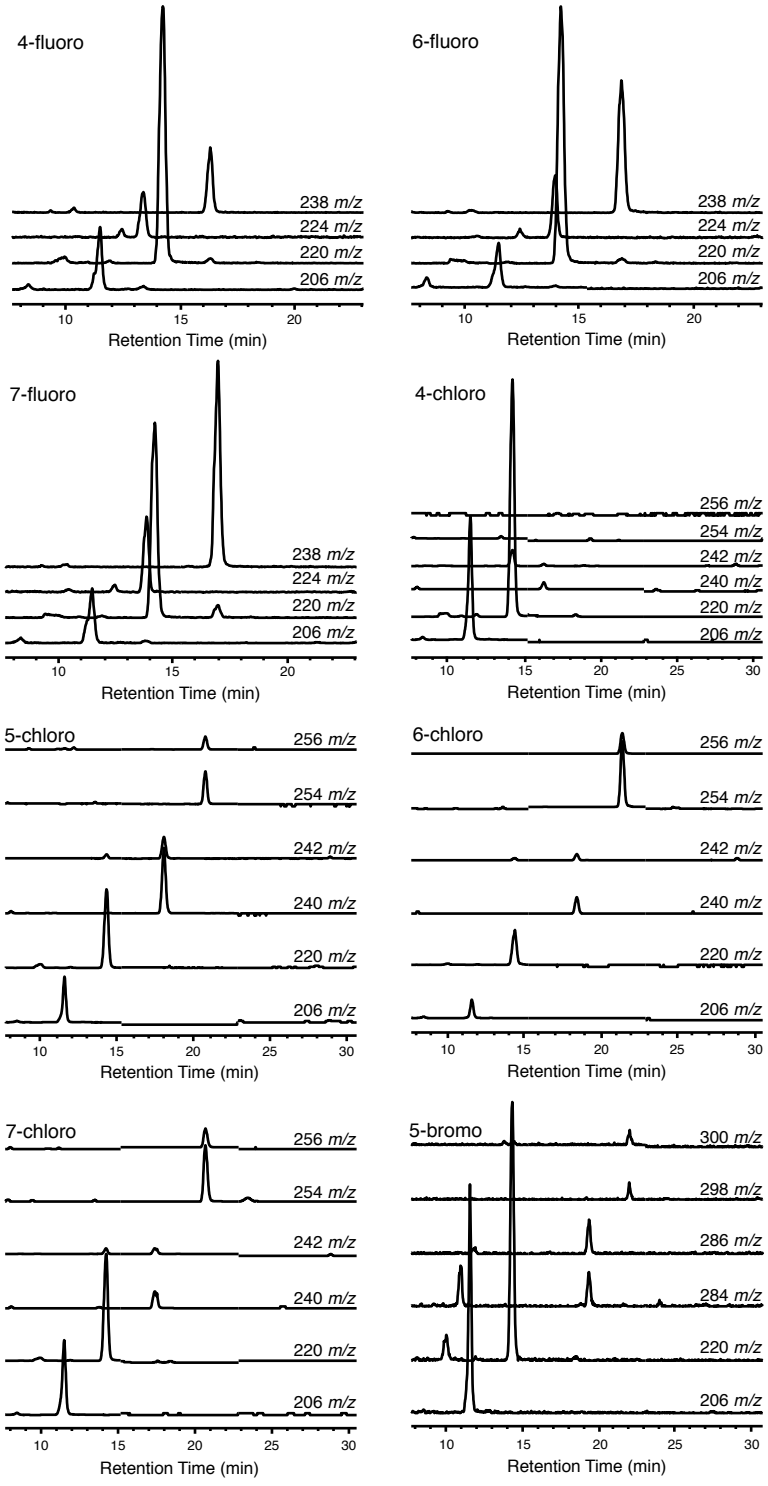


Figure 4.12: LC-MS analysis of halogen-substituted indole incorporation into **11** using *E. coli* I1234670TS. Each trace represents an extracted ion chromatogram of $[M+H]^+$ signals for **12**, **11**, substituted **12** and substituted **11**.

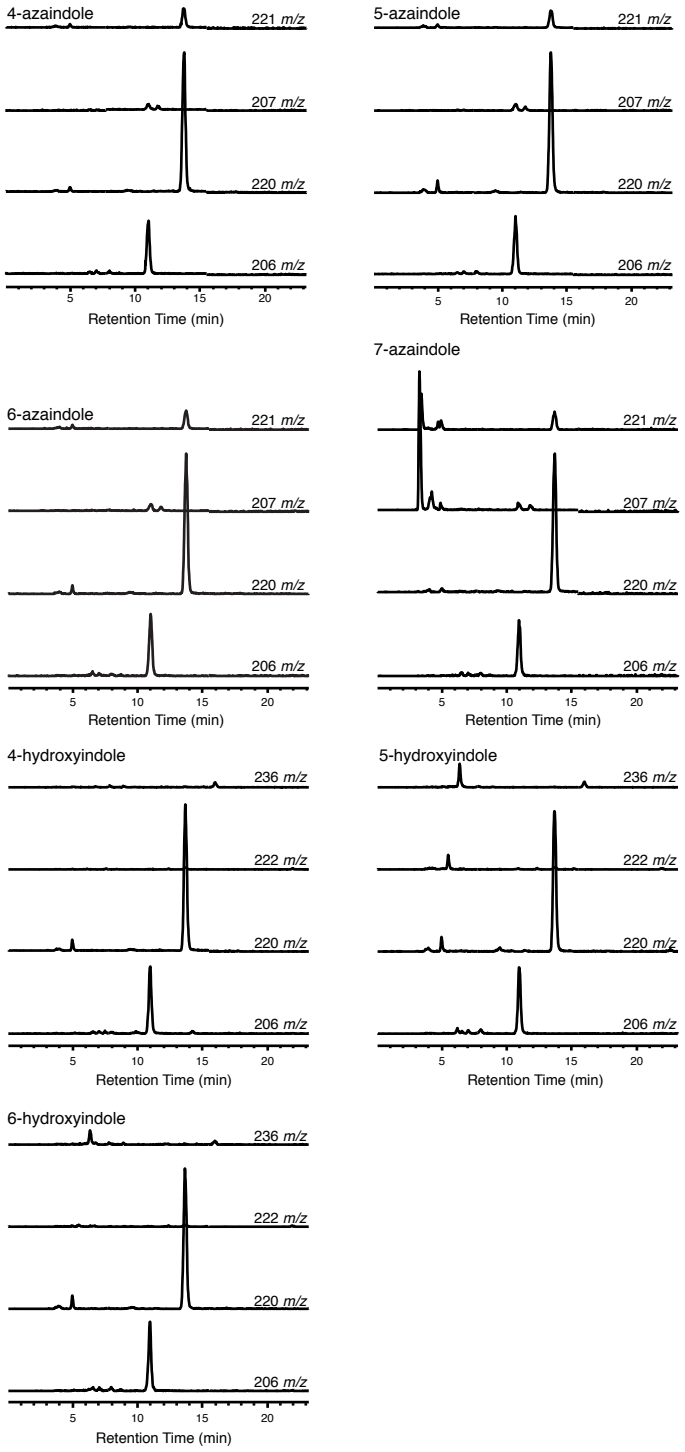


Figure 4.13: LC-MS analysis of hydroxyindole or azaindole incorporation into **11** using *E. coli* I1234670TS. Each trace represents an extracted ion chromatogram of $[M+H]^+$ signals for **12**, **11**, derivatives of **12** and derivatives of **11**. 7-azaindole, 5-hydroxyindole and 6-hydroxyindole show some evidence of incorporation.

4.3.2.3 Production of indolmycin derivatives

Feeding of fluoro and chloroindoles substituted at the 5-, 6- and 7-positions was further scaled up for purification of **11**-derivatives and downstream synthesis of **6**-derivatives (**Figure 4.14**, **Figure 4.15**, **Figure 4.16** and **Figure 4.17**). Each purified derivative of **11** and **6** was characterized by HR-MS and NMR (**Table 4.4** and **Appendix C**)**.** Cultures fed 5-fluoroindole produced the highest amount of any derivative of **11** with 26.9 mg of **5F-11**, but still lower than cultures fed indole, which produced 44.1 mg of **11** (**Table 4.3**). Overall, cultures fed the fluorinated indoles produced a higher amount of **11**-derivatives than the cultures fed with the chlorinated indoles. This result indicates that the larger substitution of the chlorine atom, compared to the fluorine atom, may be causing steric hindrance in some of the enzyme active sites, which reduces the efficiency of turnover. Additionally, growth and feeding conditions were optimized for 5-fluoroindole, so they may have biased the results towards obtaining more **5F-11** than other derivatives of **11**. Derivatives of **6** were made with between 11% and 28% yield from the three-step synthesis (**Figure 4.2**) and between 1.7 mg and 7.8 mg of each derivative was obtained (**Table 4.3**). These improvements over the yield obtained for underivatized-**6** can be attributed to improvements in drying *tert*-butanol in the oxazolinone-cyclization step, leading to less regeneration of **11** and its derivatives as side-products.

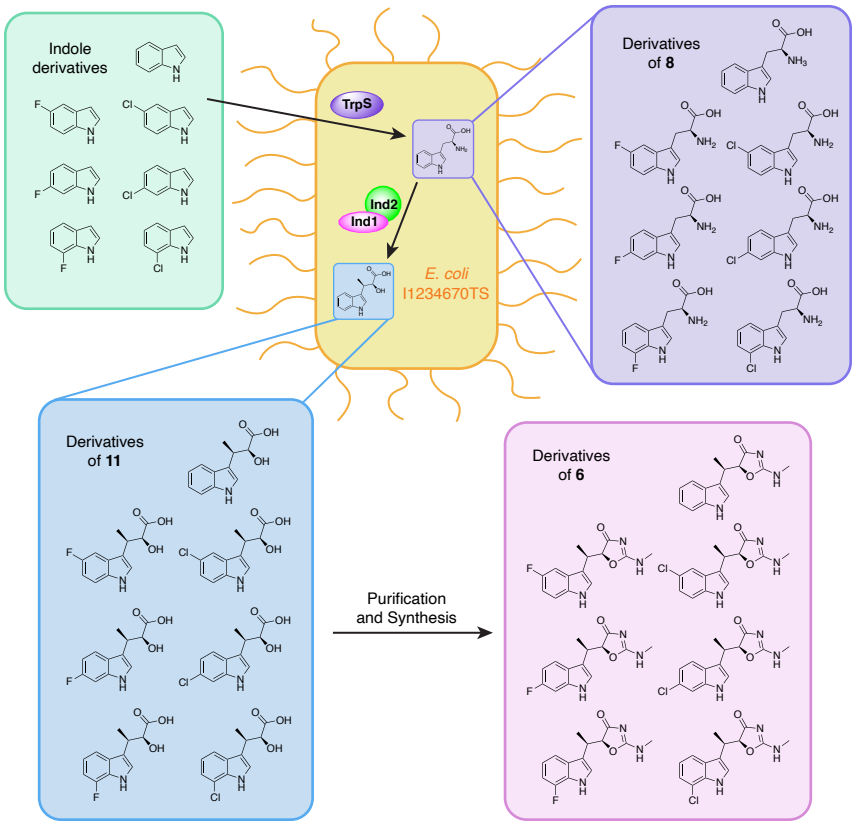


Figure 4.14: Semi-synthetic scheme of producing **6** and its derivatives starting from indole derivatives. **8**, **11** and their derivatives are produced biosynthetically, *in vivo*, from *E. coli* I1234670TS. Compound **11** and derivatives of **11** are purified and synthesized to make **6** and its derivatives.

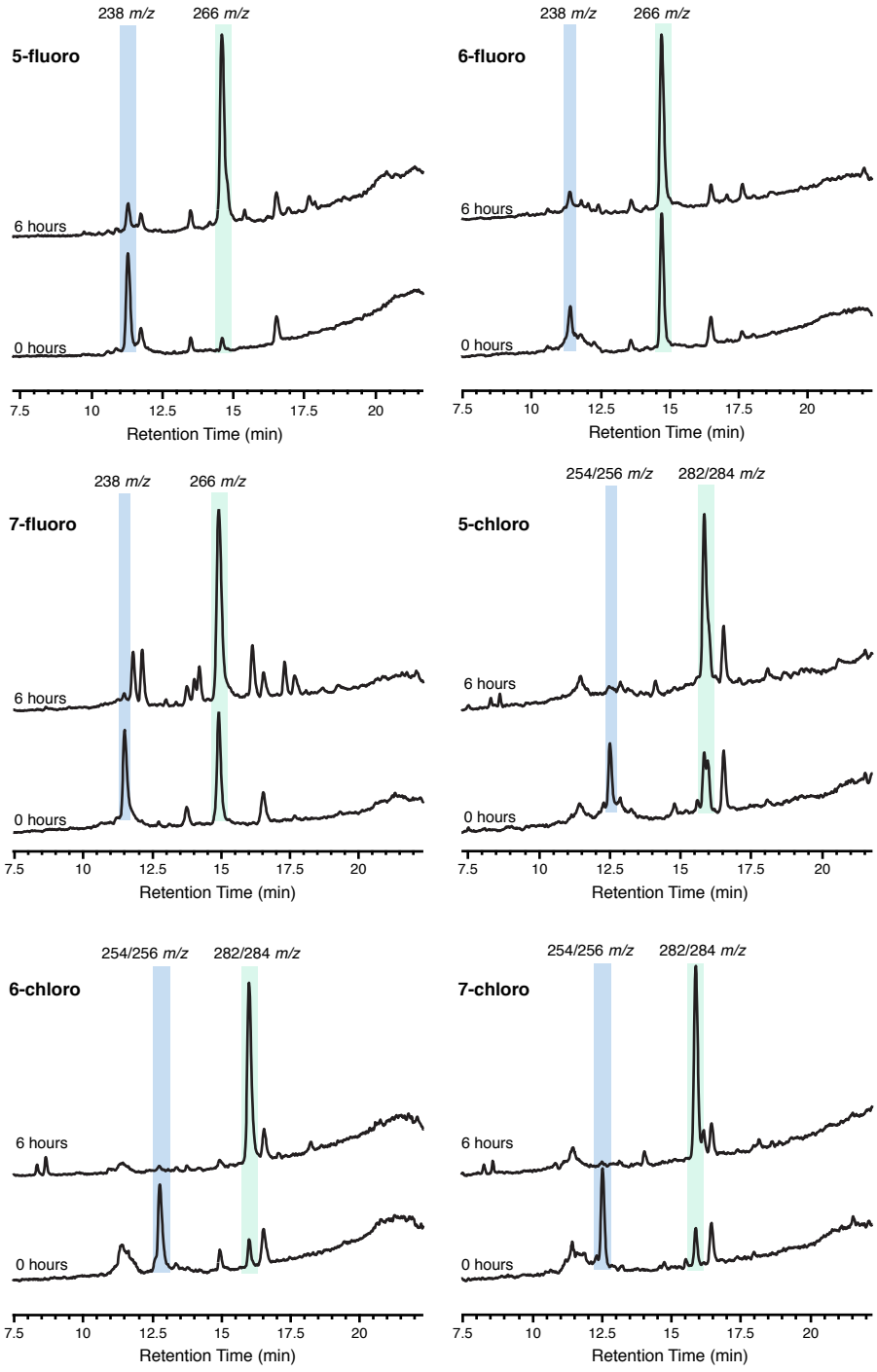


Figure 4.15: LC-MS analysis of **11**-derivative esterification. Total ion chromatograms are shown for each derivative (5-fluoro, 6-fluoro, 7-fluoro, 5-chloro, 6-chloro and 7-chloro) at the beginning of the reaction, immediately after mixing reagents (0 hours) and the end of the reaction (6 hours). The starting materials, **11**-derivatives, and the products, **13**-derivatives, are indicated by a blue or green box, respectively. The *m/z* for each compound is given above the corresponding boxes.

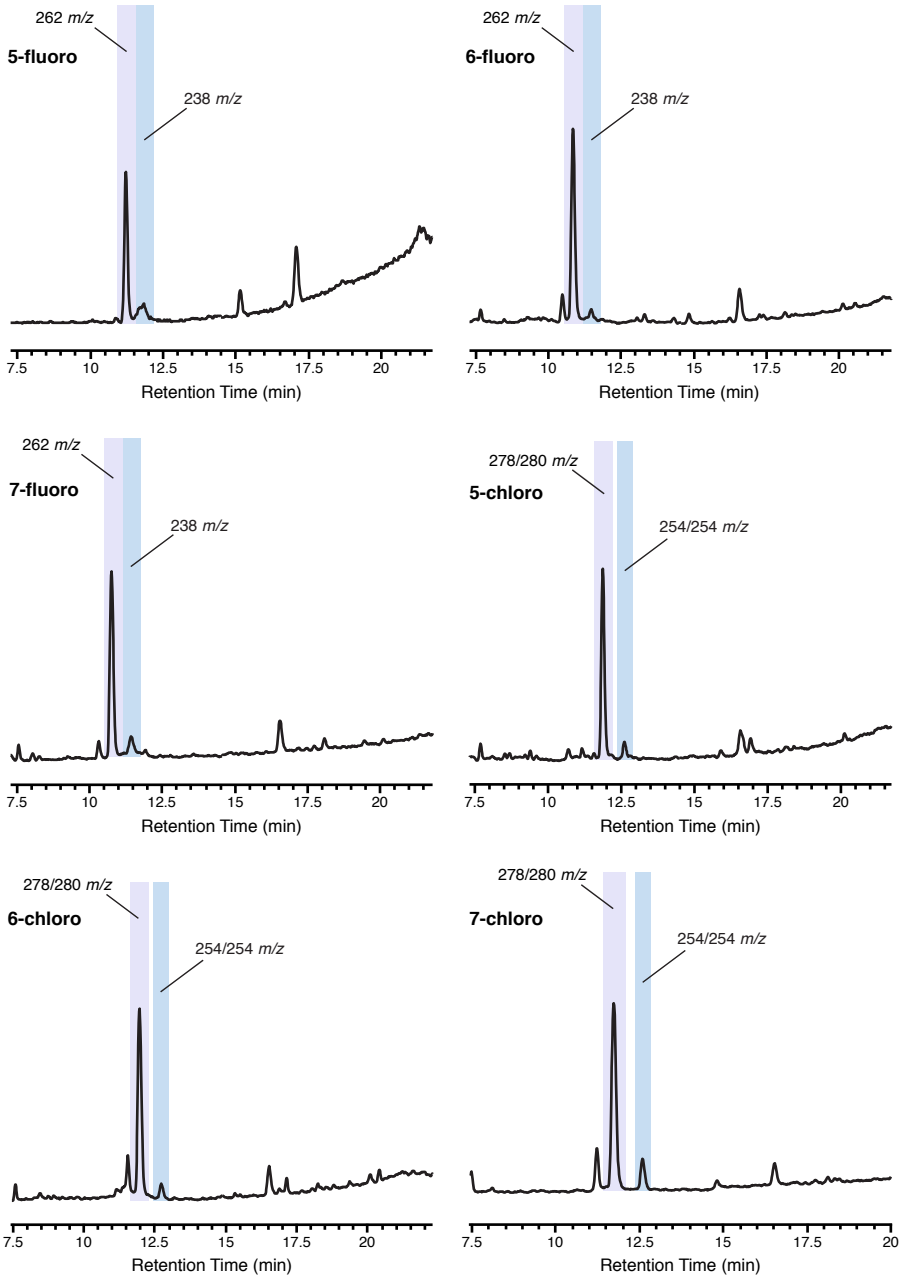


Figure 4.16: LC-MS analysis of **13**-derivative oxazolinone-cyclizations. Total ion chromatograms are shown for each derivative (5-fluoro, 6-fluoro, 7-fluoro, 5-chloro, 6-chloro and 7-chloro). Oxazolinone-cyclized products, **14**-derivatives, are indicated with a purple box and **11**-derivative side-products generated during the reaction are shown with a blue box. The *m/z* for each compound is indicated.

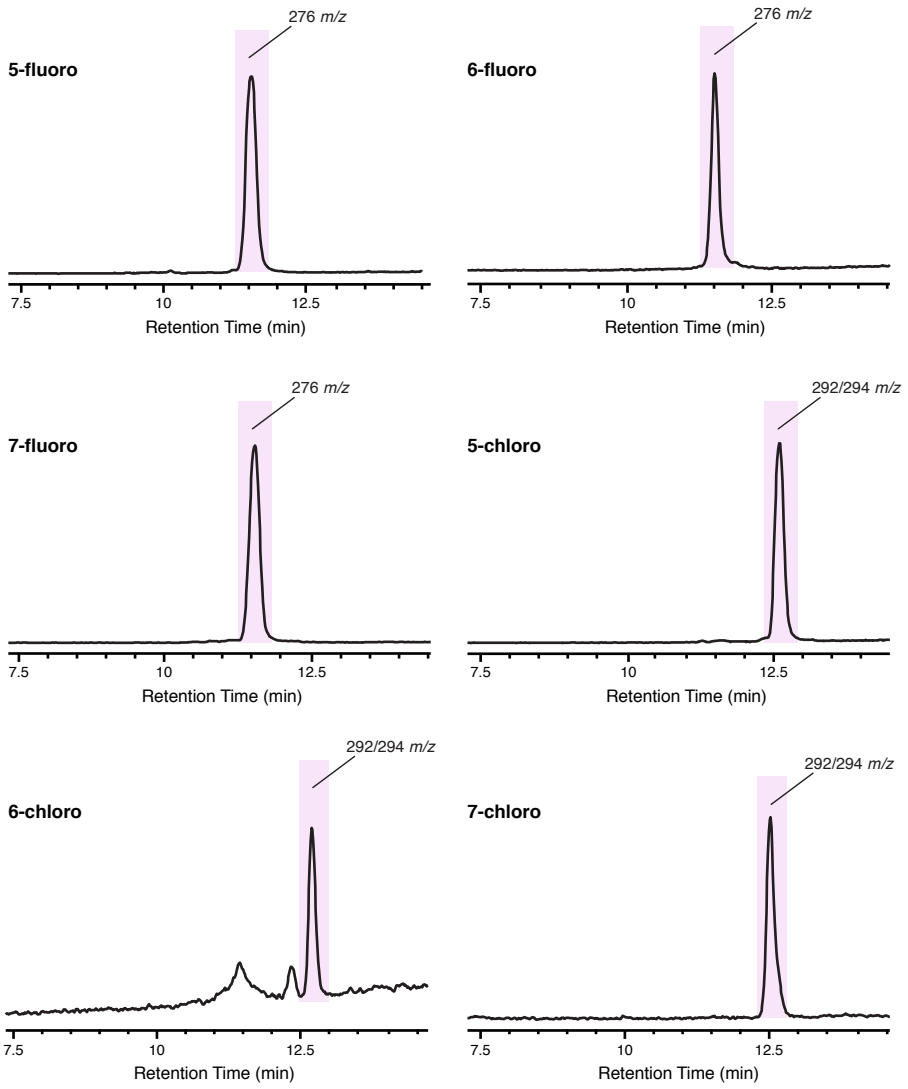


Figure 4.17: LC-MS analysis of **14**-derivative *N*-methylation. Total ion chromatograms are shown for each derivative (5-fluoro, 6-fluoro, 7-fluoro, 5-chloro, 6-chloro and 7-chloro) after semi-preparative HPLC purification. **6**-derivative products are indicated with a pink box. The *m/z* for each product is indicated.

Table 4.4: HR-MS analysis of purified **11**, **6** and their derivatives.

	Observed Mass (*m/z*)	Calculated Mass (*m/z*)	Chemical Formula	Error (ppm)
11	219.0906	219.0895	$C_{12}H_{13}NO_3$	-5
6	257.1167	257.1164	$C_{14}H_{15}N_3O_2$	-1
5F-11	237.0808	237.0801	$C_{12}H_{12}FNO_3$	-3
5F-6	275.1082	275.1070	$C_{14}H_{14}FN_3O_2$	-5
6F-11	237.0809	237.0801	$C_{12}H_{12}FNO_3$	-3
6F-6	275.1078	275.1070	$C_{14}H_{14}FN_3O_2$	-3
7F-11	237.0811	237.0801	$C_{12}H_{12}FNO_3$	-4
7F-6	275.1076	275.1070	$C_{14}H_{14}FN_3O_2$	-2
5Cl-11	253.0514	253.0506	$C_{12}H_{12}ClNO_3$	-3
5Cl-6	291.0777	291.0775	$C_{14}H_{14}ClN_3O_2$	-1
6Cl-11	253.0504	253.0506	$C_{12}H_{12}ClNO_3$	-1
6Cl-6	291.0784	291.0075	$C_{14}H_{14}ClN_3O_2$	-3
7Cl-11	253.0504	253.0506	$C_{12}H_{12}ClNO_3$	-1
7Cl-6	291.0785	291.0075	$C_{14}H_{14}ClN_3O_2$	-3

4.3.3 Bioactivity against MRSA

Finally, the bioactivities of **6** and its derivatives were tested with a disk diffusion assay against MRSA (**Figure 4.18**). Compound **6** showed the highest activity with a zone of inhibition visible when 10 μg of the compound was loaded onto each disk. **6F-6** and **7F-6** showed a zone of inhibition at 20 μg per disk while **5F-6** showed inhibition at 30 μg per disk. The chlorinated derivatives of **6** did not show any bioactivity at the maximum amount tested in the disk diffusion assay (30 μg). MIC_{50} values were determined for each compound (**Table 4.5**). The MIC_{50} values demonstrate **6** is a more potent inhibitor of MRSA than its derivatives, while **6F-6** showed the most potent inhibition of MRSA of any of the derivatives, followed by **7F-6** and **5F-6**. The lack of bioactivity shown by the chlorinated compounds may be due to the bulky chlorinated substituent hindering the compounds' abilities to bind to the TrpRS target, which is supported by docking the analogs into a bacterial TrpRS structure (**Figure 4.19**). However, bioactivity of the fluorinated substituents implies that **4F-6** could also show activity if its production was to be optimized in the future.

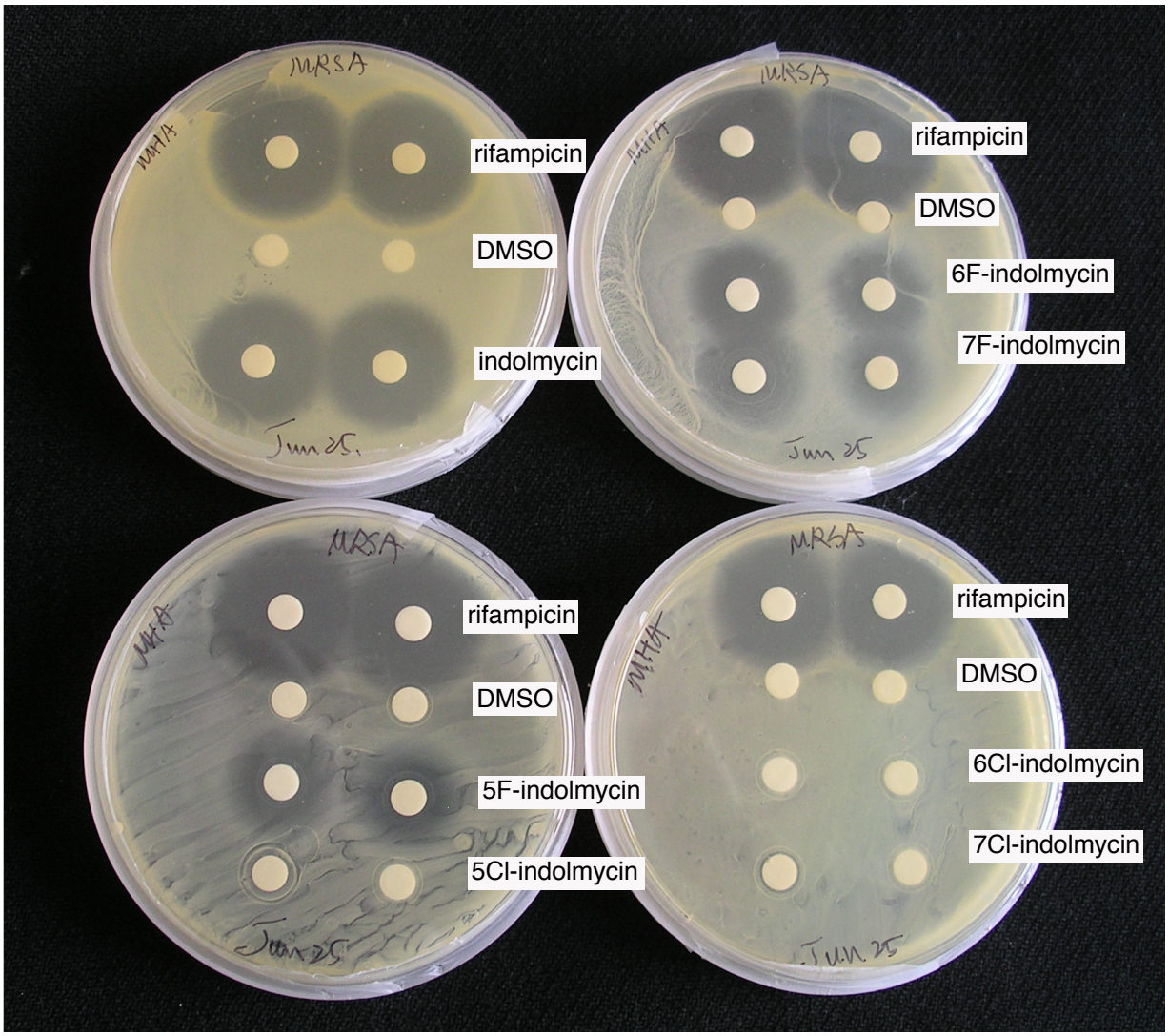


Figure 4.18: Disk diffusion assay of **6** and its derivatives against MRSA. Rifampicin (1 μg) and DMSO (0.5%) were used as positive and negative controls, respectively. 10 μg of **6**, 20 μg of **6F-6** and **7F-6** and 30 μg of **5F-6**, **5Cl-6**, **6Cl-6** and **7Cl-6** were loaded onto each disk.

Table 4.5: MIC_{50} values determined for **6** and its derivatives against MRSA. Values represent the average of three replicates ± the standard deviation. For **6**, **5F-6**, **6F-6** and **7F-6**, concentration ranges between 50 μg/mL to 0.128 ng/mL were tested, and for **5Cl-6**, **6Cl-6** and **7Cl-6**, concentration ranges between from 200 μg/mL to 0.512 ng/mL were tested. Rifampicin was used as a positive control.

	MIC_{50} (μg/mL)
rifampicin	0.0047 ± 0.0013
6	1.21 ± 0.04
5F-6	32.5 ± 19.6
6F-6	6.49 ± 0.03
7F-6	16.7 ± 4.2
5Cl-6	> 200
6Cl-6	> 200
7Cl-6	> 200

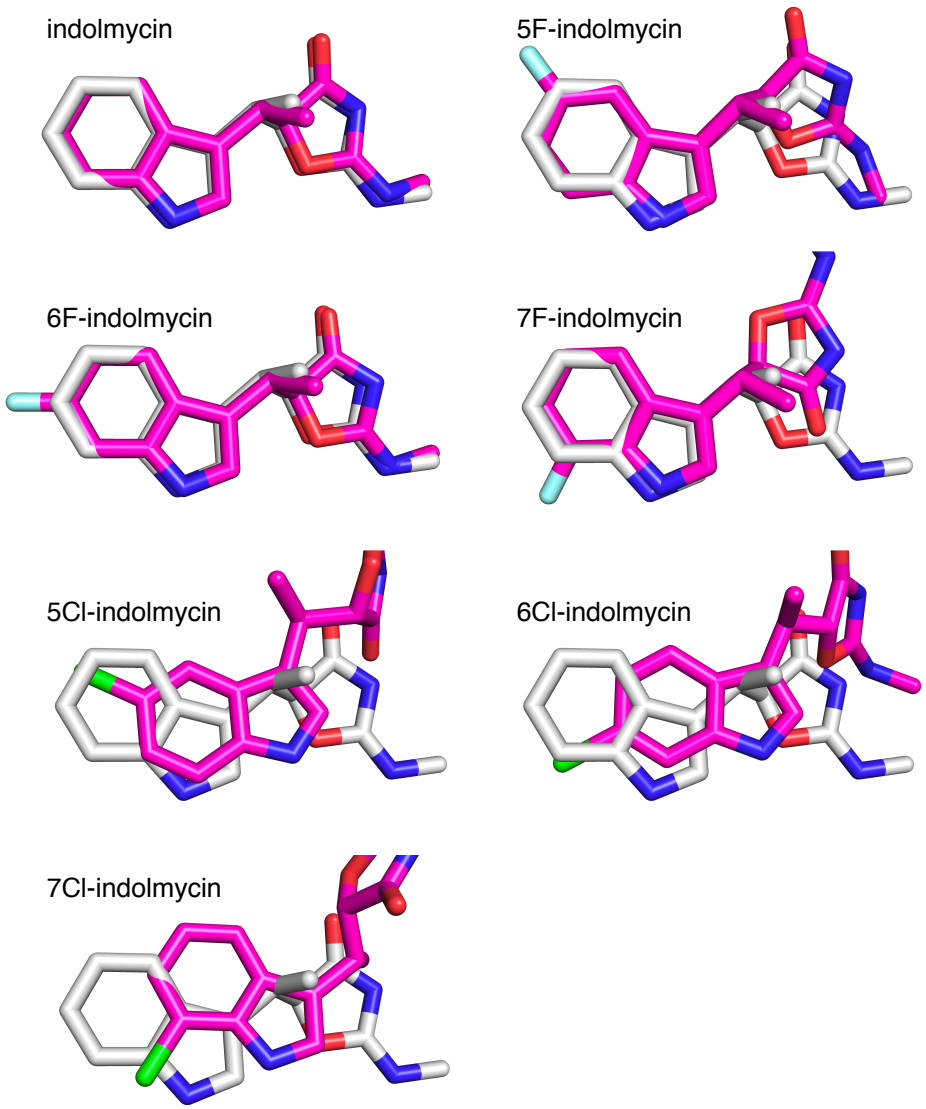


Figure 4.19: Docking of indolmycin derivatives made in this study into an indolmycin-bound crystal structure of TrpRS from *Bacillus stearothermophilus* (PDB: 5DK4). Docked molecules (magenta) are compared with active site-bound indolmycin (white) from the crystal structure. Binding modes for the indole rings of chlorinated analogs could not be matched with the crystal structure. Oxazolinone conformations for fluorinated compounds may have differed from the crystal structure because other ligands in the active site which interact with the oxazolinone ring, including ATP and Mg^{2+}, were not included in the docking model.

4.4 Discussion

By combining biosynthetic and synthetic methods of making **6**, a variety of novel halogenated derivatives of **6** were able to be made, some of which showed bioactivity against MRSA. Attempts of producing derivatives of **6** by either synthesis or biosynthesis have been

made previously. 5-hydroxy and 5-methoxy derivatives of **6** were observed when *S. griseus* ATCC12648 was fed the corresponding derivatives of indole or **8**, but derivatives of **8** and indole substituted at the 6-position were not converted to 6-substituted derivatives of **6**.[67] However, incorporation of indoles substituted at the 6-position was observed in the *E. coli* I1234670TS system, indicating that the semi-synthetic method is able to overcome a limitation of the pure biosynthetic method associated with making 6-substituted derivatives. The enzymes causing this limitation in the natural pathway are likely to be found in Part III of **6** biosynthesis (**Figure 4.1**) because the limitation must come later than Ind1 and Ind2, which are found in Part II. Additionally, purely biosynthetic methods of obtaining **6** and these 5-hydroxy- and 5-methyoxy-derivatives has relied on *Streptomyces* sp. to produce the compounds, which has more limited potential for genetic tailoring than *E. coli* and is also more difficult to grow and maintain. Therefore, the *E. coli* I1234670TS system provides a more versatile method of production by using a more adaptable organism and overcoming natural substrate selectivity by only using two of the indolmycin biosynthetic enzymes.

While the synthetic part of the semi-synthetic method helped to overcome biosynthetic challenges, the biosynthetic portion of the system also overcomes challenges associated with synthetic methods. Synthetic studies were often associated with making the chiral precursor, **11**, through methods such as chiral auxiliary usage,[59] epoxide ring opening,[56,58] lipase-assisted enantioselective acetylation,[60] or a combination of these methods.[57] The *E. coli* I1234670TS biosynthetic method of obtaining **6** and its derivatives offers a simple alternative to synthetic methods by installing chirality *in vivo*. Derivatives made during synthetic efforts were focused on making known metabolites[58] or the 5-methoxy-derivative,[60] previously shown by Werner and Demain to have increased bioactivity.[67] Attempts to synthesize other derivatives, particularly halogenated compounds, were likely not feasible as the methods employed to make these derivatized compounds, such as Larock indole synthesis,[60,61] would be incompatible with halogen substitutions. Therefore, the *E. coli* I1234670TS biosynthetic method of making **11** overcomes the incompatibility of halogenated compounds with synthetic methods for making derivatives of **11**, allowing a variety of novel halogenated derivatives of **6** to be made.

The obtained derivatives were consistent with the initial reports of the promiscuous TrpS from *S. enterica*, which showed the lowest acceptance for 4-chloro, 5-chloro and 5-bromoindoles.[88] While 4-chloro and 5-bromoindoles were not carried forward to larger-scale

growth and purification, the 5-chloroindole feeding resulted in the lowest amount of purified **11**. Since low amounts of 4-fluoroindole incorporation were observed despite Goss's observation of moderate incorporation, it is likely the indolmycin genes themselves that are sensitive to acceptance of substituents at this position. Azatryptophans have been previously made by TrpS but required longer reaction times due to reduced nucleophilicity,[85,227] which explains why they may not have been incorporated by the semi-synthetic system if they were not given a long enough reaction time. Winn *et al.* (2018) was able to produce azaindole derivatives of the agrochemical, thaxtomin, using the *S. enterica* TrpS;[94] however, longer reaction times were used (72 hours to make Trp derivatives and 5-7 days to make thaxtomin derivatives). While hydroxyindoles were not converted to derivatives of **11**, hydroxytryptophans have been produced by the *S. enterica* TrpS *in vitro*;[85] however, there are no examples of them being made *in vivo*. Several other indoles that have been successfully utilized by the *S. enterica* TrpS remain to be tested for acceptance by the semi-synthetic platform, including nitro-, methyl-, ethenyl-, ethynyl-, amino-, methoxy- and dual-substituted indoles,[87–89,94] which also have the potential to show interesting bioactivities. Overall, the substrate acceptance and amounts of **11**-derivatives obtained could be further improved with optimized reaction and purification conditions. Additionally, strains could be engineered to reduce endogenous tryptophan and enzymes, such as TrpS and Ind1/Ind2, could be engineering to improve the substrate scope. TrpS has already been the target of many enzyme engineering studies focused on improving the substrate scope;[83,90,228–232] however, other TrpS variants and homologs were not used for this biosynthetic platform because they were demonstrated for *in vitro* use with purified enzymes, rather than *in vivo*.

Compound **6** and the fluorinated derivatives of **6** showed bioactivity against MRSA with MIC values ranging from 1-33 μg/mL, demonstrating that these newly derivatized compounds may serve as useful molecules for development of new antibiotics. Observed bioactivity from the fluorinated compounds indicates that it may be worthwhile to produce and test other derivatives of **6**. Additionally, the fluorinated compounds could be tested against more of **6**'s identified targets, such as *H. pylori,*[32] and *P. falciparum,*[39] to see if the compounds show maintained, improved or reduced bioactivity. These compounds may also have other useful qualities that remain to be tested, including overcoming the narrow spectrum of activity or the hepatic toxicity of the original compound, which caused original development of **6** to be discontinued during the golden era of antibiotic discovery.

Here, the indolmycin biosynthetic enzymes were used in a semi-synthetic process to make a variety of **6**-derivatives from substituted indoles. This platform applies the best parts of the biosynthetic and synthetic methods towards **6**, while simultaneously pairing them with TrpS to open the door to several new indole-substituted analogs. The biosynthetic aspects of the process allowed for production of **11**, and its derivatives, whose construction was often the aim of synthetic studies focused on making **6** and its derivatives. The synthetic aspects of the process allowed for a greater versatility of derivatives to be explored that the purely biosynthetic system may not have been able to produce, since previous studies indicate that some enzymes from the indolmycin biosynthetic pathway may be selective against indole-substitution at some positions. Overall, this work lays a foundation for making derivatives of **6**, and perhaps analogs of other **8**-derived molecules, to be tested for antimicrobial properties. This work also demonstrates the value of coupling synthetic and biosynthetic methods to make and diversify natural product-derived compounds, beyond just those useful as antimicrobials.

Chapter 5: Conclusions and future directions

5.1 General Conclusions

The exploration of natural product biosynthesis is an important endeavor towards building industrially useful biocatalysts. By understanding how natural products are assembled, new enzymes and new ways of constructing molecules can be discovered that expand the known biocatalytic repertoire. In this work, the biosynthetic enzymes of indolmycin were shown to be useful biochemical tools to better understand the unique oxygen reactivity among some PLP-dependent enzymes and to produce halogenated indolmycin derivatives that were not easily accessible by chemical methods alone. Use of these enzymes as biochemical tools helps to expand the known the biocatalytic repertoire, which may contribute to more widespread use of biocatalysts to solve industrially relevant problems, such as more feasible production of natural products and their derivatives.

During elucidation of indolmycin biosynthesis, a unique enzyme, known as Ind4, was discovered that catalyzes an oxygen- and PLP-dependent desaturation of L-arginine.[113] An interest in studying the mechanism of Ind4, led to the discovery of a homologous, yet evolutionarily distinct enzyme, Plu4, which also catalyzes an oxygen- and PLP-dependent desaturation of L-arginine. The first crystal structure of an oxygen- and PLP-dependent desaturase, described in **Chapter 2**, revealed an active site that is highly similar to other previously studied oxygen- and PLP-dependent enzymes catalyzing L-arginine hydroxylation[114–116] but with some evidence of different dynamics between the two types of enzymes. The crystal structures solved for Plu4 during different stages of catalysis aided mutagenesis studies to understand determinants for catalysis, including a His-Asp catalytic dyad. Although these crystal structures alone could not determine the cause for reactivity differentiation, particularly if the differences between the hydroxylases and desaturases are related to dynamics, this structural information is essential to gain the mechanistic knowledge needed for engineering biocatalysts.[27] Therefore, the structural characterization described in **Chapter 2** lays a foundation for potential engineering efforts that may be needed to make the arginine desaturases a useful biocatalytic tool.

The biochemical and phylogenetic analysis of arginine desaturases in **Chapter 3** supported the findings of the structural analysis in **Chapter 2** by revealing that the hydroxylases and desaturases are more highly similar than previously expected. A minor hydroxylated side-

product was observed from the desaturases suggesting that their mechanisms are closely related. An SSN analysis revealed that some desaturases are more closely related to a group of hydroxylases than they are to other desaturases and vice versa. EPR experiments and cytochrome c assays also suggested that superoxide is an intermediate in in the reaction for the first time in the oxygen- and PLP-dependent enzymes. This information, coupled with computational work, suggests that preorganization of water in the active site of the hydroxylases may be the cause for distinction between the two reactivities. Overall, this new structural and biochemical data has provided valuable information towards identifying a unified mechanism for the arginine oxidases and will aid efforts to engineer these enzymes. Additionally, the discovery and characterization of seven new desaturases will also aid in engineering efforts by providing a pool of candidates to use as starting points.

In order to demonstrate a potential biocatalytic use for the indolmycin biosynthetic genes, in **Chapter 4** the production of indolmycin and various indole-substituted derivatives was attempted. In doing so, the first halogenated derivatives of indolmycin were made using a semi-synthetic approach that overcame compatibility issues associated with solely synthetic methods previously used to make indolmycin. The fluorinated indolmycin compounds made using the semi-synthetic method showed promising bioactivity against MRSA, suggesting that they could be useful for clinical development. Although neither indolmycin nor its derivatives were produced using purely biosynthetic methods *in vivo*, the semi-synthetic approach demonstrated the utility that biocatalysts can have by supplementing current chemical processes to overcome challenges associated with a purely chemical method. Therefore, this work shows that the study of natural products and their biosynthesis can contribute to the expansion of the biocatalytic repertoire, allowing new molecules to be made, tested and used for industrial purposes.

5.2 Future Directions

5.2.1 Arginine oxidases

Further studies regarding the arginine oxidases should be focused on understanding phylogenetic relationships, catalytic mechanisms and potential uses for the enzymes. Through SSN and genomic neighbourhood analysis and the discovery that the arginine oxidases can be organized into subgroups, it is clear that there are still many uncharacterized enzymes with potential arginine oxidase activity whose reactivity cannot be inferred from their genomic

neighbourhood. Many of these uncharacterized enzymes are a part of uncharacterized gene clusters, including the enzymes contained in non-indolmycin gene clusters in Group III and non-enduracididine gene clusters in Group II. Characterizing these unknown enzymes would serve several purposes. First, characterization would allow for further phylogenetic information and sub-group characterization. This information would contribute to a better understanding of how the hydroxylation, desaturation and perhaps overall oxygen reactivities naturally evolved by providing more sequence-activity relationship information. This information could improve the current understanding of how to engineer these reactivities into other enzymes. Second, this characterization would provide a larger pool of enzymes to draw from for future engineering efforts. Work to characterize more of these enzymes, including the work described in **Chapter 3**, would also provide clues to the function of the unknown gene clusters that encode some of these enzymes, which could lead to the discovery of new natural products and their biosynthesis.

Although this work has contributed to a better understanding of the mechanism, there are still many mechanistic details regarding the arginine oxidases that should be elucidated before attempting engineering efforts toward industrial use. Based on these observations, a new hypothesis has been provided for differentiation of the hydroxylase and desaturase reactivities, which suggested that the dynamics of the enzyme or residues found outside the active site must be responsible for distinguishing the two reactivities, perhaps through manipulating preoganization of water. However, this hypothesis should be studied in more detail to determine its validity. This hypothesis could be investigated through molecular dynamics (MD) simulations using the crystal structures described in **Chapter 2**, which would help determine any differences to the dynamics of the desaturases and the hydroxylases. Additionally, MD simulations could help identify key residues that may be responsible for distinguishing their reactivity. Another way to identify these key residues would be to evolve the arginine oxidases in the lab to transform a desaturase into a hydroxylase, or vice versa. Studying the residues that alter the reactivity could provide clues to understand how this differentiation might have occurred naturally among the arginine oxidases. However, a major hurdle to this type of study is the development of an appropriate high-throughput assay that would be needed to distinguish the hydroxylated and desaturated products.

Despite the progress in determining a mechanism for oxygen reactivity, the mechanistic details remain elusive, not only for the arginine oxidases, but for all oxygen-, PLP-dependent

enzymes. Future work in this area should consider using previous studies on flavin-dependent enzymes and cofactor-independent oxygenases as templates for exploring this reactivity.[120,123,233] For example, researchers have done in-depth explorations of redox potentials to investigate oxygen reactivity of the flavin-dependent enzymes to support measured reaction rates and reveal the energetic influence of the enzyme on redox potential. With the arginine oxidases, measuring redox potentials of the enzyme and comparing with several mutagenic variants could help elucidate residues responsible for oxygen reactivity. Studies on flavin-dependent enzymes have also further investigated oxygen reactivity via more global comparisons between the oxygen-utilizing and non-utilizing examples.[120,121] This analysis could also be applied to the arginine oxidases by comparing them with the most closely related PLP-dependent enzymes that do not use oxygen, which would first need to be analyzed phylogenetically to uncover the identity of these potential ancestors. Differences between oxygen usage in oxidases versus oxygenases has also been investigated in the flavin-dependent enzymes,[120,121] which will serve as a useful template for a similar comparison amongst the oxygen- and PLP-dependent oxidases and oxygenases.

The arginine oxidases could be useful biocatalysts in the future as they represent a unique and rare type of reaction amongst the highly versatile PLP-dependent enzymes. In order to be useful as a biocatalyst, it may be important for the oxygen-dependent hydroxylations and desaturations to be able to occur on a variety of different substrates. For this reason, the known arginine oxidases may need to be engineered to accept different substrates. Alternatively, once the basis for oxygen reactivity and the resulting hydroxylation and desaturation are known, the oxidase reactivities could be engineered into other PLP-dependent enzymes that accept other substrates. This type of engineering project would only be possible if it is confirmed that, as the data suggests, these reactivities are not dependent on the L-arginine substrate, itself, being involved in catalysis.

Another potential use for the arginine oxidases is for the production of 2-ketoacids, which are attractive precursors in the production of many compounds because of their chemical versatility.[234] Li *et al.* recently explored the use of aminotransferases for the production of 2-ketoacids by coupling aminotransferases with L-glutamate oxidase, catalase and a catalytic amount of L-glutamate to overcome issues with equilibrium and the need for an excess amount of amino acceptor.[235] With mutation at a key active site His residue, the arginine oxidases are

also able to produce the 2-ketoacid, 2-ketoarginine, but without the need for a co-substrate amino acceptor or a co-substrate recycling enzyme. Engineering the arginine oxidases to accept a wider variety of substrates could enable them to be useful in a more efficient production of 2-ketoacids.

5.2.2 Indolmycin

Further work with the indolmycin biosynthetic pathway and production of indolmycin-like compounds should involve understanding and improving the efficiency of production, the diversity of compounds and the bioactivity of indolmycin-like compounds. To improve the efficiency of production, more information should be gained about the pathway itself. In **Chapter 4**, when unnecessary genes were eliminated from the production of indolmycenic acid, the amount of indolmycenic acid produced was much less than for the system which included all of the genes from the initial attempt to make indolmycin biosynthetically. This finding indicated that the enzymes produced from these genes may in fact be necessary after all, if not catalytically, then perhaps through allosteric regulation. The role that some or all of these proteins may play on modulating the activity of Ind1 or Ind2 remains unclear but could hold key information needed to improve the efficiency of the system. Additionally, the heterologous expression system in *E. coli* produces a large amount of the shunt product, *C*-desmethyl-indolmycenic acid, which could be further optimized by understanding how regulation or environmental conditions within *S. griseus* ATCC 12648 could be mimicked in *E. coli* to improve flux through the desired indolmycenic acid intermediate. Gaining a better understanding of the regulation and optimal conditions for the indolmycin biosynthetic pathway may also enable purely biosynthetic production of indolmycin and its derivatives in the future. Improvements to the chemical synthesis would also result in a better yield for each compound from the semi-synthetic system.

In order to improve the diversity of the indolmycin-like compounds, work should be done to expand the scope and optimize yield. The yield of indolmycin derivatives described in **Chapter 4** could be improved by optimizing growth and isolation methods to increase the indolmycenic acid derivative yields. Another way to improve production of indolmycin derivatives through increasing indolmycenic acid derivative production would be to use a tryptophan auxotrophic strain of *E. coli* which would reduce the production of non-derivatized indolmycenic acid. To produce a greater diversity of indolmycin derivatives, first more indole

derivatives should be tested to better understand what is tolerated by the current system. Then, to improve the scope it will be necessary to understand why certain derivatives are not converted to indolmycenic acid derivatives. If the hurdle to conversion is uptake of the indole derivatives into *E. coli*, it would be useful to engineer a membrane transport system that could allow the indole derivatives to permeate the cell membrane. For indole derivatives whose conversion may be hindered by substrate acceptance from the biosynthetic enzymes, efforts could be placed into finding homologs with better substrate scopes or engineering the biosynthetic enzymes to improve their substrate scopes. Engineering TrpS for wider substrate conversion has already shown some success,[83,90,228–232] but has not yet been attempted for Ind1 or Ind2, as crystal structures have not yet been solved for these enzymes.

Regarding bioactivity, the antibiotic properties observed from the fluorinated indolmycin derivatives in **Chapter 4** indicates that they may be useful in pharmaceutical development against MRSA. Since development of indolmycin was originally discontinued because of potential liver toxicity issues due to an effect on catabolic liver enzymes,[34,36,37] it would be useful to test these fluorinated derivatives *in vitro* against these catabolic liver enzymes. It has also been suggested that *C*-desmethyl-indolmycin has a lower chance of causing hepatic toxicity while keeping some bioactivity against prokaryotic TrpRS,[37] suggesting that it may be useful to optimize the semi-synthetic system to produce *C*-desmethyl-indolmycin derivatives to be tested for bioactivity against MRSA. Additionally, the indolmycin derivatives should also be tested against other pathogens shown to be targets of indolmycin, such as *P. falciparum*[39] or *H. pylori*.[32]

Overall, this work has used the indolmycin biosynthetic pathway to help expand the known biocatalytic repertoire by gaining a better understanding of the oxygen- and PLP-dependent arginine oxidases and by producing new indolmycin derivatives semi-synthetically. By expanding the known biocatalytic repertoire, use of biocatalysts as chemical tools makes production of natural products and their derivatives more feasible. Industrial interest in natural products should increase as a result of greater accessibility to these complex molecules. This interest and accessibility to natural products and their derivatives is beneficial to new drug development, especially antibiotics, as nearly half of all drugs are derived from natural products.[2,9–11] With such a global focus on drug development, perhaps now more than ever, it is

possible that natural products may be indispensable in helping to deal with current and future global health crises.

References

(1) Katz, L.; Baltz, R. H. Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* **2016**, *43*, 155–176.

(2) Dias, D. A.; Urban, S.; Roessner, U. A historical overview of natural products in drug discovery. *Metabolites* **2012**, *2* (2), 303–336.

(3) Awan, A. R.; Shaw, W. M.; Ellis, T. Biosynthesis of therapeutic natural products using synthetic biology. *Adv. Drug Deliv. Rev.* **2016**, *105*, 96–106.

(4) Baltz, R. H. Molecular engineering approaches to peptide, polyketide and other antibiotics. *Nat. Biotechnol.* **2006**, *24* (12), 1533–1540.

(5) Cragg, G. M.; Newman, D. J. Biodiversity: a continuing source of novel drug leads. *Pure Appl. Chem.* **2005**, *77* (1), 7–24.

(6) Shen, B. A new golden age of natural products drug discovery. *Cell* **2015**, *163* (6), 1297–1300.

(7) Hutchings, M.; Truman, A.; Wilkinson, B. Antibiotics: past, present and future. *Curr. Opin. Microbiol.* **2019**, *51*, 72–80.

(8) Li, J. W.-H.; Vederas, J. C. Drug discovery and natural products: end of an era or an endless frontier? *Science* **2009**, *325*, 161–165.

(9) Paterson, I.; Anderson, E. A. The renaissance of natural products as drug candidates. *Science* **2005**, *310*, 451–453.

(10) Newman, D. J.; Cragg, G. M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.* **2012**, *75* (3), 311–335.

(11) Newman, D. J.; Cragg, G. M. Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* **2016**, *79* (3), 629–661.

(12) Baltz, R. H. Renaissance in antibacterial discovery from actinomycetes. *Curr. Opin. Pharmacol.* **2008**, *8*, 557–563.

(13) Pelaez, F. The historical delivery of antibiotics from microbial natural products - Can history repeat? *Biochem. Pharmacol.* **2006**, *71* (7), 981–990.

(14) Wencewicz, T. A. New antibiotics from nature's chemical inventory. *Bioorg. Med. Chem.* **2016**, *24*, 6227–6252.

(15) Durand, G. A.; Raoult, D.; Dubourg, G. Antibiotic discovery: history, methods and perspectives. *Int. J. Antimicrob. Agents* **2019**, *53*, 371–382.

(16) Davies, J, & Davies, D. Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* **2010**, *74* (3), 417–433.

(17) Aminov, R. I. A brief history of the antibiotic era: lessons learned and challenges for the future. *Front. Microbiol.* **2010**, *1*, 134.

(18) Singh, S. B.; Barrett, J. F. Empirical antibacterial drug discovery - Foundation in natural products. *Biochem. Pharmacol.* **2006**, *71* (7), 1006–1015.

(19) Prescott, J. F. The resistance tsunami, antimicrobial stewardship, and the golden age of microbiology. *Vet. Microbiol.* **2014**, *171*, 273–278.

(20) Fryszkowska, A.; Devine, P. N. Biocatalysis in drug discovery and development. *Curr. Opin. Chem. Biol.* **2020**, *55*, 151–160.

(21) B. Pickens, L.; Tang, Y.; Chooi, Y.-H. Metabolic engineering for the production of natural products. *Annu. Rev. Chem. Biomol. Eng.* **2011**, *2* (3), 211–236.

(22) Yang, D.; Park, S. Y.; Park, Y. S.; Eun, H.; Lee, S. Y. Metabolic engineering of *Escherichia coli* for natural product biosynthesis. *Trends Biotechnol.* **2019**, *38* (7), 745–765.

(23) Kennedy, J. Mutasynthesis, chemobiosynthesis, and back to semi-synthesis: combining synthetic chemistry and biosynthetic engineering for diversifying natural products. *Natural Product Reports*. Royal Society of Chemistry February 5, 2008, pp 25–34.

(24) Fischbach, M. A.; Walsh, C. T.; Clardy, J. The evolution of gene collectives: how natural selection drives chemical innovation. *Proc. Natl. Acad. Sci.* **2008**, *105* (12), 4601–4608.

(25) Winter, J. M.; Tang, Y. Synthetic biological approaches to natural product biosynthesis. *Curr. Opin. Biotechnol.* **2012**, *23*, 736–743.

(26) 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.

(27) Tibrewal, N.; Tang, Y. Biocatalysts for natural product biosynthesis. *Annu. Rev. Chem. Biomol. Eng.* **2014**, *5*, 347–366.

(28) Marsh, W. S.; Garretson, A. L.; Wesel, E. M. PA 155 A, B, and X antibiotics produced by a strain of *Streptomyces albus*. *Antibiot. Chemother.* **1960**, *10*, 316–320.

(29) Werner, R. G.; Thorpe, L. F.; Reuter, W.; Nierhaus, K. H. Indolmycin inhibits prokaryotic tryptophanyl-tRNA ligase. *Eur. J. Biochem.* **1976**, *68* (1), 1–3.

(30) Hirota, A.; Higashinaka, Y.; Sakai, H. Isolation of indolmycin and its derivative as antagonists of L-tryptophan. *Agric. Biol. Chem.* **1978**, *42* (1), 147–151.

(31) Preobrazhenskaya, M. N.; Balashova, E. G.; Turchin, K. F.; Padeiskaya, E. N.; Uvarova, N. V; Pershin, G. N.; Suvorov, N. N. Total synthesis of antibiotic indolmycin and its stereoisomers. *Tetrahedron* **1968**, *24* (19), 6131–6143.

(32) Kanamaru, T.; Nakano, Y.; Toyoda, Y.; Miyagawa, K.-I.; Tada, M.; Kaisho, T.; Nakao, M. *In vitro* and *in vivo* antibacterial activities of TAK-083, an agent for treatment of *Helicobacter pylori* infection. *Antimicrob. Agents Chemother.* **2001**, *45* (9), 2455–2459.

(33) Werner, R. G. Uptake of indolmycin in gram-positive bacteria. *Antimicrob. Agents Chemother.* **1980**, *18* (6), 858–862.

(34) Ochsner, U. A.; Sun, X.; Jarvis, T.; Critchley, I.; Janjic, N. Aminoacyl-tRNA synthetases: essential and still promising targets for new anti-infective agents. *Expert Opin. Investig. Drugs* **2007**, *16* (5), 573–593.

(35) Vondenhoff, G. H. M.; Aerschot, A. Van. Aminoacyl-tRNA synthetase inhibitors as potential antibiotics. *Eur. J. Med. Chem.* **2011**, *46* (11), 5227–5236.

(36) Hurdle, J. G.; O'Neill, A. J.; Chopra, I. Prospects for aminoacyl-tRNA synthetase inhibitors as new antimicrobial agents. *Antimicrob. Agents Chemother.* **2005**, *49* (12), 4821–4833.

(37) Werner, R. G.; Reuter, W. Interaction of indolmycin in the metabolism of tryptophan in rat liver. *Arzneimittelforschung* **1979**, *29* (1), 59–63.

(38) Hurdle, J. G.; O'Neill, A. J.; Chopra, I. Anti-staphylococcal activity of indolmycin, a potential topical agent for control of staphylococcal infections. *J. Antimicrob. Chemother.* **2004**, *54* (2), 549–552.

(39) Pasaje, C. F. A.; Cheung, V.; Kennedy, K.; Lim, E. E.; Baell, J. B.; Griffin, M. D. W.; Ralph, S. A. Selective inhibition of apicoplast tryptophanyl-tRNA synthetase causes delayed death in *Plasmodium falciparum*. *Sci. Rep.* **2016**, *6* (January), 27531.

(40) Kanamaru, T.; Nakao, M.; Tawada, H.; Kamiyama, K. Oxazolone derivatives and their

use as anti-Helicobacter pylori agent. US6169102B1, June 24, 2001.

(41) Kanamaru, T.; Masafumi, N.; Tawada, H.; Kamiyama, K. Oxazolone derivatives and their use as anti-*Helicobacter pylori* agents. US6414003B1, November 9, 2002.

(42) Naotsuka, A.; Izawa, M.; Miyagawa, K. Process for producing indolmycin. US6387687B1, July 29, 2002.

(43) Kamiyama, K.; Nakayama, Y. Process for producing indolmycins. US6265586B1, April 8, 2001.

(44) Rajendran, V.; Kalita, P.; Shukla, H.; Kumar, A.; Tripathi, T. Aminoacyl-tRNA synthetases: structure, function, and drug discovery. *Int. J. Biol. Macromol.* **2018**, *111*, 400–414.

(45) Cochrane, R. V. K.; Norquay, A. K.; Vederas, J. C. Natural products and their derivatives as tRNA synthetase inhibitors and antimicrobial agents. *Medchemcomm* **2016**, *7* (8), 1535–1545.

(46) Gadakh, B.; Van Aerschot, A. Aminoacyl-tRNA synthetase inhibitors as antimicrobial agents: a patent review from 2006 till present. *Expert Opin. Ther. Pat.* **2012**, *22* (12), 1453–1465.

(47) Giegé, R.; Springer, M. Aminoacyl-tRNA synthetases in the bacterial world. *EcoSal Plus* **2016**.

(48) Du, Y.-L.; Alkhalaf, L. M.; Ryan, K. S. *In vitro* reconstitution of indolmycin biosynthesis reveals the molecular basis of oxazolinone assembly. *Proc. Natl. Acad. Sci.* **2015**, *112* (9), 2717–2722.

(49) Xu, X.; Zhou, H.; Liu, Y.; Liu, X.; Fu, J.; Li, A.; Li, Y.; Shen, Y.; Bian, X.; Zhang, Y. Heterologous expression guides identification of the biosynthetic gene cluster of chuangxinmycin, an indole alkaloid antibiotic. *J. Nat. Prod.* **2018**, *81* (4), 1060–1064.

(50) Williams, T. L.; Yin, Y. W.; Carter, C. W. Selective inhibition of bacterial tryptophanyl-tRNA synthetases by indolmycin is mechanism-based. *J. Biol. Chem.* **2016**, *291* (1), 255–265.

(51) Hatch, N. D.; Ouellette, S. P. Inhibition of tRNA synthetases induces persistence in chlamydia. *Infect. Immun.* **2020**, *88* (4), e00943-19.

(52) Windels, E. M.; Michiels, J. E.; Fauvart, M.; Wenseleers, T.; Van den Bergh, B.; Michiels, J. Bacterial persistence promotes the evolution of antibiotic resistance by

increasing survival and mutation rates. *ISME J.* **2019**, *13* (5), 1239–1251.

(53) Dirlam, J. P.; Clark, D. a; Hecker, S. J. New total synthesis of (+/-)-indolmycin. *J. Org. Chem.* **1986**, *51* (25), 4920–4924.

(54) Shue, Y. K. Total synthesis of (±) indolmycin. *Tetrahedron Lett.* **1996**, *37* (36), 6447–6448.

(55) von Wittenau, M. S.; Els, H. Chemistry of indolmycin. *J. Am. Chem. Soc.* **1963**, *85* (21), 3425–3421.

(56) Chan, T. H.; Hill, K. K. The absolute configuration of indolmycin. *J. Org. Chem.* **1970**, *35* (10), 3519–3521.

(57) Akita, H.; Kawaguchi, T.; Enoki, Y.; Oishi, T. Formal total synthesis of (-)-indolmycin. *Chem. Pharm. Bull.* **1990**, *38* (2), 323–328.

(58) Hasuoka, A.; Nakayama, Y.; Adachi, M.; Kamiguchi, H.; Kamiyama, K. Development of a stereoselective practical synthetic route to indolmycin, a candidate anti-*H. pylori* agent. *Chem. Pharm. Bull.* **2001**, *49* (12), 1604–1608.

(59) Takeda, T.; Mukaiyama, T. Asymmetric total synthesis of indolmycin. *Chem. Lett.* **1980**, *9* (27), 163–166.

(60) Sutou, N.; Kato, K.; Akita, H. A concise synthesis of (-)-indolmycin and (-)-5-methoxyindolmycin. *Tetrahedron: Asymmetry* **2008**, *19* (15), 1833–1838.

(61) Larock, R. C.; Yum, E. K. Synthesis of indoles via palladium-catalyzed heterannulation of internal alkynes. *J. Am. Chem. Soc.* **1991**, *113* (17), 6689–6690.

(62) Harnden, M. R.; Wright, N. D. Total synthesis of thiazolinone analogues of indolmycin. *J. Chem. Soc. Perkin Trans. 1* **1977**, 1012–1018.

(63) Corrodi, H.; F Hansson, L. C.; Carnmalm, B.; Jacupovic, E.; Johansson, L.; de Paulis, T.; Rámsby, S.; Stjernström, N. E.; Renyi, A. L.; Ross, S. B.; Ógren, S.; Med Chem, J.; Cripps, N.; Williams, J. K.; Sharkey, W. H.; Am, J.; Harnden, M. R.; Bailey, S.; Boyd, M. R.; Taylor, D. R.; Wright, N. D. Thiazolinone analogues of indolmycin with antiviral and antibacterial activity. *J. Med. Chem.* **1978**, *21* (1), 82–87.

(64) Witty, D. R.; Walker, G.; Bateson, J. H.; O'hanlon, P. J.; Eggleston, D. S.; Haltiwanger, R. C. Synthesis of conformationally restricted analogues of the tryptophanyl tRNA synthetase inhibitor indolmycin. *Tetrahedron Lett.* **1996**, *37* (17), 3067–3070.

(65) Hornemann, U.; Hurley, L. H.; Speedie, M. K.; Floss, H. G. The biosynthesis of

indolmycin. *J. Am. Chem. Soc.* **1971**, *93* (12), 3028–3035.

(66) Speedie, M. K.; Hornemann, U.; Floss, H. G. Isolation and characterization of tryptophan transaminase and indolepyruvate *C*-methyltransferase. *J. Biol. Chem.* **1975**, *250* (19), 7819–7825.

(67) Werner, R. G.; Demain, A. L. Directed biosynthesis of new indolmycins. *J. Antibiot. (Tokyo).* **1981**, *XXXIV* (5), 551–554.

(68) Du, Y.-L.; Higgins, M. A.; Zhao, G.; Ryan, K. S. Convergent biosynthetic transformations to a bacterial specialized metabolite. *Nat. Chem. Biol.* **2019**, *15*, 1043–1048.

(69) Dunathan, H. C. Conformation and reaction specificity in pyridoxal phosphate enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **1966**, *55* (4), 712–716.

(70) Eliot, A. C.; Kirsch, J. F. Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu. Rev. Biochem.* **2004**, *73* (1), 383–415.

(71) Rocha, J. F.; Pina, A. F.; Sousa, S. F.; Cerqueira, N. M. F. S. A. PLP-dependent enzymes as important biocatalysts for the pharmaceutical, chemical and food industries: a structural and mechanistic perspective. *Catal. Sci. Technol.* **2019**, *9*, 4864–4876.

(72) Hoegl, A.; Nodwell, M. B.; Kirsch, V. C.; Bach, N. C.; Pfanzelt, M.; Stahl, M.; Schneider, S.; Sieber, S. A. Mining the cellular inventory of pyridoxal phosphate-dependent enzymes with functionalized cofactor mimics. *Nat. Chem.* **2018**, *10* (12), 1234–1245.

(73) Du, Y.-L.; Ryan, K. S. Pyridoxal phosphate-dependent reactions in the biosynthesis of natural products. *Nat. Prod. Rep.* **2019**, *36* (3), 430–457.

(74) Yang, G.; Ding, Y. Recent advances in biocatalyst discovery, development and applications. *Bioorg. Med. Chem.* **2014**, *22* (20), 5604–5612.

(75) Nestl, B. M.; Hammer, S. C.; Nebel, B. A.; Hauer, B. New generation of biocatalysts for organic synthesis. *Angew. Chem. Int. Ed.* **2014**, *53*, 3070–3095.

(76) Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; Devine, P. N.; Huisman, G. W.; Hughes, G. J. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **2010**, *329*, 305–309.

(77) Midelfort, K. S.; Kumar, R.; Han, S.; Karmilowicz, M. J.; McConnell, K.; Gehlhaar, D. K.; Mistry, A.; Chang, J. S.; Anderson, M.; Villalobos, A.; Minshull, J.; Govindarajan, S.; Wong, J. W. Redesigning and characterizing the substrate specificity and activity of

Vibrio fluvialis aminotransferase for the synthesis of imagabalin. *Protein Eng. Des. Sel.* **2013**, *26* (1), 25–33.

(78) Raboni, S.; Bettati, S.; Mozzarelli, A. Tryptophan synthase: a mine for enzymologists. *Cell. Mol. Life Sci.* **2009**, *66* (14), 2391–2403.

(79) Watkins-Dulaney, E. J.; Straathof, S.; Arnold, F. H.; Watkins-Dulaney, E.; Arnold, F.; Watkins-Dulaney, E.; Arnold, F.; Straathof, S. Tryptophan synthase: biocatalyst extraordinaire. *ChemBioChem* **2020**, In Press.

(80) Dunn, M. F. Allosteric regulation of substrate channeling and catalysis in the tryptophan synthase bienzyme complex. *Arch. Biochem. Biophys.* **2012**, *519*, 154–166.

(81) Buller, A. R.; Brinkmann-Chen, S.; Romney, D. K.; Herger, M.; Murciano-Calles, J.; Arnold, F. H. Directed evolution of the tryptophan synthase β-subunit for stand-alone function recapitulates allosteric activation. *Proc. Natl. Acad. Sci.* **2015**, *112* (47), 14599–14604.

(82) Murciano-Calles, J.; Romney, D. K.; Brinkmann-Chen, S.; Buller, A. R.; Arnold, F. H. A panel of TrpB biocatalysts derived from tryptophan synthase through the transfer of mutations that mimic allosteric activation. *Angew. Chem. Int. Ed.* **2016**, *55* (38), 11577–11581.

(83) Murciano-Calles, J.; Buller, A. R.; Arnold, F. H. Directed evolution of an allosteric tryptophan synthase to create a platform for synthesis of noncanonical amino acids. In *Directed Enzyme Evolution: Advances and Applications*; Springer International Publishing, 2017; pp 1–16.

(84) Buller, A. R.; Van Roye, P.; Cahn, J. K. B.; Scheele, R. A.; Herger, M.; Arnold, F. H. Directed evolution mimics allosteric activation by stepwise tuning of the conformational ensemble. *J. Am. Chem. Soc.* **2018**, *140* (23), 7256–7266.

(85) Phillips, R. S. Synthetic applications of tryptophan synthase. *Tetrahedron: Asymmetry* **2004**, *15* (18), 2787–2792.

(86) Goss, R. J. M.; Newill, P. L. A. A convenient enzymatic synthesis of L-halotryptophans. *Chem. Commun.* **2006**, No. 47, 4924–4925.

(87) Winn, M.; Roy, A. D.; Grüschow, S.; Parameswaran, R. S.; Goss, R. J. M. A convenient one-step synthesis of L-aminotryptophans and improved synthesis of 5-fluorotryptophan. *Bioorg. Med. Chem. Lett.* **2008**, *18* (16), 4508–4510.

(88) Smith, D. R. M.; Willemse, T.; Gkotsi, D. S.; Schepens, W.; Maes, B. U. W.; Ballet, S.; Goss, R. J. M. The first one-pot synthesis of L-7-iodotryptophan from 7-iodoindole and serine, and an improved synthesis of other L-7-halotryptophans. *Org. Lett.* **2014**, *16* (10), 2622–2625.

(89) Corr, M. J.; Smith, D. R. M.; Goss, R. J. M. One-pot access to L-5,6-dihalotryptophans and L-alknyltryptophans using tryptophan synthase. *Tetrahedron* **2016**, *72* (46), 7306–7310.

(90) Francis, D.; Winn, M.; Latham, J.; Greaney, M. F.; Micklefield, J. An engineered tryptophan synthase opens new enzymatic pathways to β-methyltryptophan and derivatives. *ChemBioChem* **2017**, *18* (4), 382–386.

(91) Parmeggiani, F.; Casamajo, A. R.; Walton, C. J.; Galman, J. L.; Turner, N. J.; Chica, R. A. One-pot biocatalytic synthesis of substituted D-tryptophans from indoles enabled by an engineered aminotransferase. *ACS Catal.* **2020**, *9*, 3482–3486.

(92) McDonald, A. D.; Perkins, L. J.; Buller, A. R. Facile *in vitro* biocatalytic production of diverse tryptamines. *ChemBioChem* **2019**, *20* (15), 1939–1944.

(93) Blei, F.; Baldeweg, F.; Fricke, J.; Hoffmeister, D. Biocatalytic production of psilocybin and derivatives in tryptophan synthase-enhanced reactions. *Chem. - A Eur. J.* **2018**, *24* (40), 10028–10031.

(94) Winn, M.; Francis, D.; Micklefield, J. De novo biosynthesis of "non-natural" thaxtomin phytotoxins. *Angew. Chem. Int. Ed.* **2018**, *57* (23), 6830–6833.

(95) Abell, L. M.; Schloss, J. V. Oxygenase side reactions of acetolactate synthase and other carbanion-forming enzymes. *Biochemistry* **1991**, *30* (32), 7883–7887.

(96) Bunik, V. I.; Schloss, J. V; Pinto, J. T.; Dudareva, N.; Cooper, A. J. L. A survey of oxidative paracatalytic reactions catalyzed by enzymes that generate carbanionic intermediates: implications for ROS production, cancer etiology, and neurodegenerative diseases. *Adv. Enzymol. Relat. Areas Mol. Biol.* **2011**, *77*, 307–360.

(97) Christen, P. Paracatalytic enzyme modification by oxidation of enzyme—substrate carbanion intermediates. *Methods Enzymol.* **1977**, *46*, 48–54.

(98) Cogoli-Greuter, M.; Hausner, U.; Christen, P. Irreversible inactivation of pyruvate decarboxylase in the presence of substrate and an oxidant. An example of paracatalytic enzyme inactivation. *Eur. J. Biochem.* **1979**, *100* (1), 295–300.

(99) Bertoldi, M.; Dominici, P.; Moore, P. S.; Maras, B.; Voltattorni, C. B. Reaction of dopa decarboxylase with α-methyldopa leads to an oxidative deamination producing 3,4-dihydroxyphenylacetone, an active site directed affinity label. *Biochemistry* **1998**, *37* (18), 6552–6561.

(100) Bertoldi, M.; Carbone, V.; Voltattorni, C. B. Ornithine and glutamate decarboxylases catalyse an oxidative deamination of their alpha-methyl substrates. *Biochem. J.* **1999**, *342* (Pt 3), 509–512.

(101) Kaminaga, Y.; Schnepp, J.; Peel, G.; Kish, C. M.; Ben-Nissan, G.; Weiss, D.; Orlova, I.; Lavie, O.; Rhodes, D.; Wood, K.; Porterfield, D. M.; Cooper, A. J. L.; Schloss, J. V; Pichersky, E.; Vainstein, A.; Dudareva, N. Plant phenylacetaldehyde synthase is a bifunctional homotetrameric enzyme that catalyzes phenylalanine decarboxylation and oxidation. *J. Biol. Chem.* **2006**, *281* (33), 23357–23366.

(102) Sakai, M.; Hirata, H.; Sayama, H.; Sekiguchi, K.; Itano, H.; Asai, T.; Dohra, H.; Hara, M.; Watanabe, N. Production of 2-phenylethanol in roses as the dominant floral scent compound from L-phenylalanine by two key enzymes, a PLP-dependent decarboxylase and a phenyl acetaldehyde reductase. *Biosci. Biotechnol. Biochem.* **2007**, *71* (10), 2408–2419.

(103) Torrens-Spence, M. P.; Gillaspy, G.; Zhao, B.; Harich, K.; White, R. H.; Li, J. Biochemical evaluation of a parsley tyrosine decarboxylase results in a novel 4-hydroxyphenylacetaldehyde synthase enzyme. *Biochem. Biophys. Res. Commun.* **2012**, *418* (2), 211–216.

(104) Torrens-Spence, M. P.; Pluskal, T.; Li, F.-S.; Carballo, V.; Weng, J.-K. Complete pathway elucidation and heterologous reconstitution of *Rhodiola* salidroside biosynthesis. *Mol. Plant* **2018**, *11* (1), 205–217.

(105) Torrens-Spence, M. P.; Liu, C.-T.; Pluskal, T.; Chung, Y. K.; Weng, J.-K. Monoamine biosynthesis via a noncanonical calcium-activatable aromatic amino acid decarboxylase in psilocybin mushroom. *ACS Chem. Biol.* **2018**, *13* (12), 3343–3353.

(106) Torrens-Spence, M. P.; Chiang, Y.-C.; Smith, T.; Vicent, M. A.; Wang, Y.; Weng, J.-K. Structural basis for independent origins of new catalytic machineries in plant AAAD proteins. *bioRxiv* **2018**, 404970.

(107) Gutensohn, M.; Klempien, A.; Kaminaga, Y.; Nagegowda, D. A.; Negre-Zakharov, F.;

Huh, J.-H.; Luo, H.; Weizbauer, R.; Mengiste, T.; Tholl, D.; Dudareva, N. Role of aromatic aldehyde synthase in wounding/herbivory response and flower scent production in different Arabidopsis ecotypes. *Plant J.* **2011**, *66*, 591–602.

(108) Torrens-Spence, M. P.; von Guggenberg, R.; Lazear, M.; Ding, H.; Li, J. Diverse functional evolution of serine decarboxylases: identification of two novel acetaldehyde synthases that uses hydrophobic amino acids as substrates. *BMC Plant Biol.* **2014**, *14* (1), 247.

(109) Vavricka, C.; Han, Q.; Huang, Y.; Erickson, S. M.; Harich, K.; Christensen, B. M.; Li, J. From L-dopa to dihydroxyphenylacetaldehyde: a toxic biochemical pathway plays a vital physiological function in insects. *PLoS One* **2011**, *6* (1), e16124.

(110) Sakai, K.; Miyasako, Y.; Nagatomo, H.; Watanabe, H.; Wakayama, M.; Moriguchi, M. L-Ornithine decarboxylase from *Hafnia alvei* has a novel L-ornithine oxidase activity. *J. Biochem.* **1997**, *122*, 961–968.

(111) Pontrelli, S.; Fricke, R. C. B.; Teoh, S. T.; Laviña, W. A.; Putri, S. P.; Fitz-Gibbon, S.; Chung, M.; Pellegrini, M.; Fukusaki, E.; Liao, J. C. Metabolic repair through emergence of new pathways in *Escherichia coli*. *Nat. Chem. Biol.* **2018**, *14* (11), 1005–1009.

(112) Wang, M.; Zhao, Q.; Zhang, Q.; Liu, W. Differences in PLP-dependent cysteinyl processing lead to diverse S-functionalization of lincosamide antibiotics. *J. Am. Chem. Soc.* **2016**, *138* (20), 6348–6351.

(113) Du, Y.-L.; Singh, R.; Alkhalaf, L. M.; Kuatsjah, E.; He, H.-Y.; Eltis, L. D.; Ryan, K. S. A pyridoxal phosphate–dependent enzyme that oxidizes an unactivated carbon-carbon bond. *Nat. Chem. Biol.* **2016**, *12* (3), 194–199.

(114) Han, L.; Schwabacher, A. W.; Moran, G. R.; Silvaggi, N. R. *Streptomyces wadayamensis* MppP is a PLP-dependent L-arginine α-deaminase, γ-hydroxylase in the enduracididine biosynthetic pathway. *Biochemistry* **2015**, *54* (47), 7029–7040.

(115) Han, L.; Vuksanovic, N.; Oehm, S. A.; Fenske, T. G.; Schwabacher, A. W.; Silvaggi, N. R. *Streptomyces wadayamensis* MppP is a PLP-dependent oxidase, not an oxygenase. *Biochemistry* **2018**, *57*, 3252–3264.

(116) Hedges, J. B.; Kuatsjah, E.; Du, Y.-L.; Eltis, L. D.; Ryan, K. S. Snapshots of the catalytic cycle of an O_2, pyridoxal phosphate-dependent hydroxylase. *ACS Chem. Biol.* **2018**, *13*, 965–974.

(117) Dai, G. Z.; Han, W. B.; Mei, Y. N.; Xu, K.; Jiao, R. H.; Ge, H. M.; Tan, R. X. Pyridoxal-5'-phosphate-dependent bifunctional enzyme catalyzed biosynthesis of indolizidine alkaloids in fungi. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (2), 1174–1180.

(118) Ringel, M. T.; Dräger, G.; Brüser, T. PvdN enzyme catalyzes a periplasmic pyoverdine modification. *J. Biol. Chem.* **2016**, *291* (46), 23929–23938.

(119) Huang, Y.; Liu, X.; Cui, Z.; Wiegmann, D.; Niro, G.; Ducho, C.; Song, Y.; Yang, Z.; Van Lanen, S. G. Pyridoxal-5′-phosphate as an oxygenase cofactor: discovery of a carboxamide-forming, α-amino acid monooxygenase-decarboxylase. *Proc. Natl. Acad. Sci.* **2018**, 201718667.

(120) Mattevi, A. To be or not to be an oxidase: challenging the oxygen reactivity of flavoenzymes. *Trends Biochem. Sci.* **2006**, *31* (5), 276–283.

(121) Chaiyen, P.; Fraaije, M. W.; Mattevi, A. The enigmatic reaction of flavins with oxygen. *Trends Biochem. Sci.* **2012**, *37* (9), 373–380.

(122) Marcus, R. A. Electron transfer reactions in chemistry. Theory and experiment. *Rev. Mod. Phys.* **1993**, *65* (3), 599–610.

(123) Thierbach, S.; Bui, N.; Zapp, J.; Chhabra, S. R.; Kappl, R.; Fetzner, S. Substrate-assisted O_2 activation in a cofactor-independent dioxygenase. *Chem. Biol.* **2014**, *21* (2), 217–225.

(124) Schloss, J. Oxygen toxicity from plants to people. *Planta* **2002**, *216* (1), 38–43.

(125) Bertoldi, M.; Cellini, B.; Montioli, R.; Borri Voltattorni, C. Insights into the mechanism of oxidative deamination catalyzed by DOPA decarboxylase. *Biochemistry* **2008**, *47* (27), 7187–7195.

(126) Bertoldi, M.; Gonsalvi, M.; Contestabile, R.; Voltattorni, C. B. Mutation of tyrosine 332 to phenylalanine converts dopa decarboxylase into a decarboxylation-dependent oxidative deaminase. *J. Biol. Chem.* **2002**, *277* (39), 36357–36362.

(127) Bertoldi, M.; Cellini, B.; Maras, B.; Voltattorni, C. B. A quinonoid is an intermediate of oxidative deamination reaction catalyzed by Dopa decarboxylase. *FEBS Lett.* **2005**, *579* (23), 5175–5180.

(128) Fetzner, S.; Steiner, R. A. Cofactor-independent oxidases and oxygenases. *Appl. Microbiol. Biotechnol.* **2010**, *86* (3), 791–804.

(129) Bunik, V. I.; Schloss, J. V.; Pinto, J. T.; Gibson, G. E.; Cooper, A. J. L. Enzyme-catalyzed side reactions with molecular oxygen may contribute to cell signaling and

neurodegenerative diseases. *Neurochem. Res.* **2007**, *32* (4–5), 871–891.

(130) Bowes, G.; Ogren, W. L.; Hageman, R. H. Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. *Biochem. Biophydical Res. Commun.* **1971**, *45* (3), 716–722.

(131) Bowes, G.; Ogren, W. L. Oxygen inhibition and other properties of soybean ribulose-1,5-diphosphate carboxylase. *J. Biol. Chem.* **1972**, *247* (7), 2171–2176.

(132) Bertoldi, M.; Borri Voltattorni, C. Reaction of dopa decarboxylase with L-aromatic amino acids under aerobic and anaerobic conditions. *Biochem. J.* **2000**, *352*, 533–538.

(133) Fenalti, G.; Law, R. H. P.; Buckle, A. M.; Langendorf, C.; Tuck, K.; Rosado, C. J.; Faux, N. G.; Mahmood, K.; Hampe, C. S.; Banga, J. P.; Wilce, M.; Schmidberger, J.; Rossjohn, J.; El-Kabbani, O.; Pike, R. N.; Smith, A. I.; Mackay, I. R.; Rowley, M. J.; Whisstock, J. C. GABA production by glutamic acid decarboxylase is regulated by a dynamic catalytic loop. *Nat. Struct. Mol. Biol.* **2007**, *14* (4), 280–286.

(134) Goff, S. A.; Klee, H. J. Plant volatile compounds: sensory cues for health and nutritional value? *Science* **2006**, *311* (5762), 815–819.

(135) Knudsen, J. T.; Tollsten, L.; Bergström, L. G. Floral scents—a checklist of volatile compounds isolated by head-space techniques. *Phytochemistry* **1993**, *33* (2), 253–280.

(136) Torrens-Spence, M. P.; Liu, P.; Ding, H.; Harich, K.; Gillaspy, G.; Li, J. Biochemical evaluation of the decarboxylation and decarboxylation-deamination activities of plant aromatic amino acid decarboxylases. *J. Biol. Chem.* **2013**, *288* (4), 2376–2387.

(137) Liao, C.; Upadhyay, A.; Liang, J.; Han, Q.; Li, J. 3,4-Dihydroxyphenylacetaldehyde synthase and cuticle formation in insects. *Dev. Comp. Immunol.* **2018**, *83*, 44–50.

(138) Sparrow, J. C.; F Wright, T. R. The selection for mutants in *Drosophila melanogaster* hypersensitive to α-methyl dopa, a dopa decarboxylase inhibitor. *Mol. Gen. Genet.* **1974**, *130*, 127–141.

(139) Marsh, J. L.; Erfle, M. P.; Leeds, C. A. Molecular localization, developmental expression and nucleotide sequence of the alpha-methyldopa hypersensitive gene of Drosophila. *Genetics* **1986**, *114*, 453–467.

(140) Wang, D.; Marsh, J. L.; Ayala, F. J. Evolutionary changes in the expression pattern of a developmentally essential gene in three Drosophila species. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (14), 7103–7107.

(141) Wang, D.; Marsh, J. L. Developmental regulation of the α-methyldopa hypersensitive gene of *Drosophila melanogaster*. *Dev. Biol.* **1995**, *168* (2), 598–612.

(142) Liang, J.; Han, Q.; Ding, H.; Li, J. Biochemical identification of residues that discriminate between 3,4-dihydroxyphenylalanine decarboxylase and 3,4-dihydroxyphenylacetaldehyde synthase-mediated reactions. *Insect Biochem. Mol. Biol.* **2017**, *91*, 34–43.

(143) Vavricka, C. J.; Yoshida, T.; Kuriya, Y.; Takahashi, S.; Ogawa, T.; Ono, F.; Agari, K.; Kiyota, H.; Li, J.; Ishii, J.; Tsuge, K.; Minami, H.; Araki, M.; Hasunuma, T.; Kondo, A. Mechanism-based tuning of insect 3,4-dihydroxyphenylacetaldehyde synthase for synthetic bioproduction of benzylisoquinoline alkaloids. *Nat. Commun.* **2019**, *10* (1), 2015.

(144) Spížek, J.; Řezanka, T. Lincosamides: chemical structure, biosynthesis, mechanism of action, resistance, and applications. *Biochem. Pharmacol.* **2017**, *133*, 20–28.

(145) Kamenik, Z.; Kadlcik, S.; Radojevic, B.; Jiraskova, P.; Kuzma, M.; Gazak, R.; Najmanova, L.; Kopecky, J.; Janata, J. Deacetylation of mycothiol-derived "waste product" triggers the last biosynthetic steps of lincosamide antibiotics. *Chem. Sci.* **2016**, *7*, 430–435.

(146) Ushimaru, R.; Lin, C.-I.; Sasaki, E.; Liu, H. Characterization of enzymes catalyzing transformations of cysteine S-conjugated intermediates in the lincosamide biosynthetic pathway. *ChemBioChem* **2016**, *17* (17), 1606–1611.

(147) Lamont, I. L.; Beare, P. A.; Ochsner, U.; Vasil, A. I.; Vasil, M. L. Siderophore-mediated signaling regulates virulence factor production in *Pseudomonas*. *Proc. Natl. Acad. Sci.* **2002**, *99* (10), 7072–7077.

(148) Kirienko, N. V; Ausubel, F. M.; Ruvkun, G. Mitophagy confers resistance to siderophore-mediated killing by *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (6), 1821–1826.

(149) Minandri, F.; Imperi, F.; Frangipani, E.; Bonchi, C.; Visaggio, D.; Facchini, M.; Pasquali, P.; Bragonzi, A.; Visca, P. Role of iron uptake systems in *Pseudomonas aeruginosa* virulence and airway infection. *Infect. Immun.* **2016**, *84* (8), 2324–2335.

(150) Drake, E. J.; Gulick, A. M. Structural characterization and high-throughput screening of inhibitors of PvdQ, an NTN hydrolase involved in pyoverdine synthesis. *ACS Chem. Biol.*

2011, *6* (11), 1277–1286.

(151) Hannauer, M.; Schäfer, M.; Hoegy, F.; Gizzi, P.; Wehrung, P.; Mislin, G. L. A.; Budzikiewicz, H.; Schalk, I. J. Biosynthesis of the pyoverdine siderophore of *Pseudomonas aeruginosa* involves precursors with a myristic or a myristoleic acid chain. *FEBS Lett.* **2012**, *586* (1), 96–101.

(152) Nadal-Jimenez, P.; Koch, G.; Reis, C. R.; Muntendam, R.; Raj, H.; Jeronimus-Stratingh, C. M.; Cool, R. H.; Quax, W. J. PvdP is a tyrosinase that drives maturation of the pyoverdine chromophore in *Pseudomonas aeruginosa*. *J. Bacteriol.* **2014**, *196* (14), 2681–2690.

(153) Drake, E. J.; Gulick, A. M. 1.2 Å resolution crystal structure of the periplasmic aminotransferase PvdN from *Pseudomonas aeruginosa*. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **2016**, *72*, 403–408.

(154) Yamaguchi, H.; Sato, S.; Yoshida, S.; Takada, K.; Itoh, M. Capuramycin, a new nucleoside antibiotic. Taxonomy, fermentation, isolation and characterization. *J. Antibiot. (Tokyo).* **1986**, *39* (8), 1047–1053.

(155) Cai, W.; Goswami, A.; Yang, Z.; Liu, X.; Green, K. D.; Barnard-Britson, S.; Baba, S.; Funabashi, M.; Nonaka, K.; Sunkara, M.; Morris, A. J.; Spork, A. P.; Ducho, C.; Garneau-Tsodikova, S.; Thorson, J. S.; Van Lanen, S. G. The biosynthesis of capuramycin-type antibiotics: identification of the A-102395 biosynthetic gene cluster, mechanism of self-resistance, and formation of uridine-5′-carboxamide. *J. Biol. Chem.* **2015**, *290* (22), 13710–13724.

(156) Funabashi, M.; Nonaka, K.; Yada, C.; Hosobuchi, M.; Masuda, N.; Shibata, T.; Van Lanen, S. G. Identification of the biosynthetic gene cluster of A-500359s in *Streptomyces griseus* SANK60196. *J. Antibiot. (Tokyo).* **2009**, *62* (6), 325–332.

(157) Michael, J. P. Indolizidine and quinolizidine alkaloids. *Nat. Prod. Rep.* **2008**, *25* (1), 139–165.

(158) Michael, J. P. Simple indolizidine and quinolizidine alkaloids. In *The Alkaloids: Chemistry and Biology (Volume 55)*; Academic Press, 2001; pp 91–258.

(159) Chen, S.; Bacauanu, V.; Knecht, T.; Mercado, B. Q.; Bergman, R. G.; Ellman, J. A. New regio- and stereoselective cascades via unstabilized azomethine ylide cycloadditions for the synthesis of highly substituted tropane and indolizidine frameworks. *J. Am. Chem.*

Soc. **2016**, *138* (38), 12664–12670.

(160) Han, W. B.; Lu, Y. H.; Zhang, A. H.; Zhang, G. F.; Mei, Y. N.; Jiang, N.; Lei, X.; Song, Y. C.; Ng, S. W.; Tan, R. X. Curvulamine, a new antibacterial alkaloid incorporating two undescribed units from a Curvularia species. *Org. Lett.* **2014**, *16* (20), 5366–5369.

(161) Ling, L. L.; Schneider, T.; Peoples, A. J.; Spoering, A. L.; Engels, I.; Conlon, B. P.; Mueller, A.; Schäberle, T. F.; Hughes, D. E.; Epstein, S.; Jones, M.; Lazarides, L.; Steadman, V. A.; Cohen, D. R.; Felix, C. R.; Fetterman, K. A.; Millett, W. P.; Nitti, A. G.; Zullo, A. M.; Chen, C.; Lewis, K. A new antibiotic kills pathogens without detectable resistance. *Nature* **2015**, *517* (7535), 455–459.

(162) Tsuchiya, K.; Takeuchi, Y. Enduracidin, an inhibitor of cell wall synthesis. *J. Antibiot. (Tokyo).* **1968**, *21* (6), 426–428.

(163) Singh, M. P.; Petersen, P. J.; Weiss, W. J.; Janso, J. E.; Luckman, S. W.; Lenoy, E. B.; Bradford, P. A.; Testa, R. T.; Greenstein, M. Mannopeptimycins, new cyclic glycopeptide antibiotics produced by *Streptomyces hygroscopicus* LL-AC98: antibacterial and mechanistic activities. *Antimicrob. Agents Chemother.* **2003**, *47* (1), 62–69.

(164) Yin, X.; Zabriskie, T. M. The enduracidin biosynthetic gene cluster from *Streptomyces fungicidicus*. *Microbiology* **2006**, *152*, 2969–2983.

(165) Magarvey, N. A.; Haltli, B.; He, M.; Greenstein, M.; Hucul, J. A. Biosynthetic pathway for mannopeptimycins, lipoglycopeptide antibiotics active against drug-resistant gram-positive pathogens. *Antimicrob. Agents Chemother.* **2006**, *50* (6), 2167–2177.

(166) Burroughs, A. M.; Hoppe, R. W.; Goebel, N. C.; Sayyed, B. H.; Voegtline, T. J.; Schwabacher, A. W.; Zabriskie, T. M.; Silvaggi, N. R. Structural and functional characterization of MppR, an enduracididine biosynthetic enzyme from *Streptomyces hygroscopicus*: functional diversity in the acetoacetate decarboxylase-like superfamily. *Biochemistry* **2013**, *52* (26), 4492–4506.

(167) Haltli, B.; Tan, Y.; Magarvey, N. A.; Wagenaar, M.; Yin, X.; Greenstein, M.; Hucul, J. A.; Zabriskie, T. M. Investigating β-hydroxyenduracididine formation in the biosynthesis of the mannopeptimycins. *Chem. Biol.* **2005**, *12*, 1163–1168.

(168) Voogd, C. E. On the mutagenicity of nitroimidazoles. *Mutat. Res. Genet. Toxicol.* **1981**, *86* (3), 243–277.

(169) Lancini, G. C.; Kluepfel, D.; Lazzari, E.; Sartori, G. Origin of the nitro group of

azomycin. *Biochim. Biophys. Acta - Gen. Subj.* **1966**, *130* (1), 37–44.

(170) Lancini, C. G.; Lazzari, E.; Sartori, C. Microbial oxidation of aminoimidazoles to nitroimidazoles. *J. Antibiot. (Tokyo).* **1968**, *21* (6), 387–392.

(171) Seki, Y.; Nakamura, T.; Okami, Y. Accumulation of 2-aminoimidazole by *Streptomyces eurocidicus*. *J. Biochem.* **1970**, *67* (3), 389–396.

(172) Nakane, A.; Nakamura, T.; Eguchi, Y. A novel metabolic fate of arginine in *Streptomyces eurocidicus*. *J. Biol. Chem.* **1977**, *252* (15), 5267–5273.

(173) Hedges, J. B.; Ryan, K. S. *In vitro* reconstitution of the biosynthetic pathway to the nitroimidazole antibiotic azomycin. *Angew. Chem. Int. Ed.* **2019**, *58* (34), 11647–11651.

(174) Guo, J.; Higgins, M. A.; Daniel-Ivad, P.; Ryan, K. S. An asymmetric reductase that intercepts acyclic imino acids produced in situ by a partner oxidase. *J. Am. Chem. Soc.* **2019**, *141* (31), 12258–12267.

(175) Okonechnikov, K.; Golosova, O.; Fursov, M. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **2012**, *28* (8), 1166–1167.

(176) Robert, X.; Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **2014**, *42*, W320–W324.

(177) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *Phaser* crystallographic software. *J. Appl. Crystallogr.* **2007**, *40* (4), 658–674.

(178) Moriarty, N. W.; Grosse-Kunstleve, R. W.; Adams, P. D. *electronic Ligand Builder and Optimization Workbench* (*eLBOW*): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2009**, *65* (10), 1074–1080.

(179) Terwilliger, T. C.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Moriarty, N. W.; Zwart, P. H.; Hung, L.-W.; Read, R. J.; Adams, P. D. Iterative model building, structure refinement and density modification with the *PHENIX AutoBuild* wizard. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2008**, *64* (1), 61–69.

(180) Emsley, P.; Cowtan, K. *Coot* : model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60* (12), 2126–2132.

(181) Afonine, P. V.; Mustyakimov, M.; Grosse-Kunstleve, R. W.; Moriarty, N. W.; Langan, P.; Adams, P. D. Joint X-ray and neutron refinement with *phenix.refine*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66* (11), 1153–1163.

(182) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani,

G.; Barone, V.; Petersson, G.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A.; Bloino, J.; Janesko, B.; Gomperts, R.; Mennucci, B.; Hratchian, H.; Ortiz, J.; Izmaylov, A.; Sonnenberg, J.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. J.; Peralta, J.; Ogliaro, F.; Bearpark, M.; Heyd, J.; Brothers, E.; Kudin, K.; Staroverov, V.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J.; Iyengar, S.; Tomasi, J.; Cossi, M.; Millam, J.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J.; Mart, R.; Fox, D. Gaussian 09, Revision D. 01. Gaussian, Inc.: Wallingford, CT 2009.

(183) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98* (7), 5648–5652.

(184) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37* (2), 785–789.

(185) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **1980**, *58* (8), 1200–1211.

(186) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* **1994**, *98* (45), 11623–11627.

(187) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105* (8), 2999–3093.

(188) Zhao, Y.; Truhlar, D. G. Computational characterization and modeling of buckyball tweezers: density functional study of concave-convex π···π interactions. *Phys. Chem. Chem. Phys.* **2008**, *10* (19), 2813–2818.

(189) Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Use of solution-phase vibrational frequencies in continuum models for the free energy of solvation. *J. Phys. Chem. B* **2011**, *115* (49), 14556–14562.

(190) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class

functionals and 12 other function. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

(191) Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M. Integrated modeling program, applied chemical theory (IMPACT). *J. Comput. Chem.* **2005**, *26* (16), 1752–1780.

(192) Schrödinger Release 2015–3: MacroModel. Schrödinger, LLC: New York, NY 2015.

(193) Legault, C. CYLview, 1.0b. Université de Sherbrooke: Université de Sherbrooke 2009.

(194) Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455–461.

(195) John, R. A. Pyridoxal phosphate-dependent enzymes. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1995**, *1248* (2), 81–96.

(196) Schneider, G.; Käck, H.; Lindqvist, Y. The manifold of vitamin B6 dependent enzymes. *Structure*. Current Biology Ltd January 1, 2000, pp R1–R6.

(197) Melissa P. Hill; Elizabeth C. Carroll; Michael D. Toney, A.; Larsen, D. S. Rapid photodynamics of vitamin B6 coenzyme pyridoxal 5′-phosphate and its Schiff bases in solution. *J. Phys. Chem. B* **2008**, *112*, 5867–5873.

(198) Mehta, P. K.; Hale, T. I.; Christen, P. Aminotransferases: demonstration of homology and division into evolutionary subgroups. *Eur. J. Biochem.* **1993**, *214* (2), 549–561.

(199) Mcphalen, C. A.; Vincent, M. G.; Jansonius, J. N. X-ray structure refinement and comparison of three forms of mitochondrial aspartate aminotransferase. *J. Mol. Biol.* **1992**, *225*, 495–517.

(200) Mcphalen, C. A.; Vincent, M. G.; Picot, D.; Jansonius, J. N. Domain closure in mitochondrial aspartate aminotransferase. *J. Mol. Biol.* **1992**, *227*, 197–213.

(201) Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): a web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta - Proteins Proteomics* **2015**, *1854* (8), 1019–1037.

(202) Zallot, R.; Oberg, N.; Gerlt, J. A. The EFI web resource for genomic enzymology tools: leveraging protein, genome, and metagenome databases to discover novel enzymes and

metabolic pathways. *Biochemistry* **2019**, *58* (41), 4169–4182.

(203) Zallot, R.; Oberg, N. O.; Gerlt, J. A. 'Democratized' genomic enzymology web tools for functional assignment. *Curr. Opin. Chem. Biol.* **2018**, *47*, 77–85.

(204) Gerlt, J. A. Genomic enzymology: web tools for leveraging protein family sequence–function space and genome context to discover novel functions. *Biochemistry* **2017**, *56* (33), 4293–4308.

(205) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13* (11), 2498–2504.

(206) Hoang, D. T.; Chernomor, O.; von Haeseler, A.; Minh, B. Q.; Vinh, L. S. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **2018**, *35* (2), 518–522.

(207) Kalyaanamoorthy, S.; Minh, B. Q.; Wong, T. K. F.; von Haeseler, A.; Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14* (6), 587–589.

(208) Nguyen, L.-T.; Schmidt, H. A.; von Haeseler, A.; Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32* (1), 268–274.

(209) Letunic, I.; Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **2016**, *44* (Web Server issue), W242–W245.

(210) Cornish-Brown, A. *Analysis of Enzyme Kinetic Data*; Oxford University Press, 1995.

(211) Wu, J.; Li, L.; Deng, Z.; Zabriskie, T. M.; He, X. Analysis of the mildiomycin biosynthesis gene cluster in *Streptoverticillum remofaciens* ZJU5119 and characterization of MilC, a hydroxymethyl cytosyl-glucuronic acid synthase. *ChemBioChem* **2012**, *13* (11), 1613–1621.

(212) Kuzkaya, N.; Weissmann, N.; Harrison, D. G.; Dikalov, S. Interactions of peroxynitrite, tetrahydrobiopterin, ascorbic acid, and thiols: implications for uncoupling endothelial nitric-oxide synthase. *J. Biol. Chem.* **2003**, *278* (25), 22546–22554.

(213) Khersonsky, O.; Tawfik, D. S. Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* **2010**, *79* (1), 471–505.

(214) Glasner, M. E.; Fayazmanesh, N.; Chiang, R. A.; Sakai, A.; Jacobson, M. P.; Gerlt, J. A.;

Babbitt, P. C. Evolution of structure and function in the o-succinylbenzoate synthase/N-acylamino acid racemase family of the enolase superfamily. *J. Mol. Biol.* **2006**, *360* (1), 228–250.

(215) Song, L.; Kalyanaraman, C.; Fedorov, A. A.; Fedorov, E. V; Glasner, M. E.; Brown, S.; Imker, H. J.; Babbitt, P. C.; Almo, S. C.; Jacobson, M. P.; Gerlt, J. A. Prediction and assignment of function for a divergent N-succinyl amino acid racemase. *Nat. Chem. Biol.* **2007**, *3* (8), 486–491.

(216) Martínez-Rodríguez, S.; Soriano-Maldonado, P.; Gavira, J. A. N-succinylamino acid racemases: enzymatic properties and biotechnological applications. *Biochim. Biophys. Acta - Proteins Proteomics* **2020**, *1868* (4), 140377.

(217) Zallot, R.; Harrison, K.; Kolaczkowski, B.; de Crécy-Lagard, V. Functional annotations of paralogs: a blessing and a curse. *Life* **2016**, *6* (3), 39.

(218) Phillips, G.; Grochowski, L. L.; Bonnett, S.; Xu, H.; Bailly, M.; Blaby-Haas, C.; Basma, E. Y.; Iwata-Reuyl, D.; White, R. H.; de Crécy-Lagard, V. Functional promiscuity of the COG0720 family. *ACS Chem. Biol.* **2012**, *7* (1), 197–209.

(219) Kandlinger, F.; Plach, M. G.; Merkl, R. AGeNNT: annotation of enzyme families by means of refined neighborhood networks. *BMC Bioinformatics* **2017**, *18* (1), 1–13.

(220) Miles, Z. D.; Roberts, S. A.; McCarty, R. M.; Bandarian, V. Biochemical and structural studies of 6-carboxy-5,6,7,8-tetrahydropterin synthase reveal the molecular basis of catalytic promiscuity within the tunnel-fold superfamily. *J. Biol. Chem.* **2014**, *289* (34), 23641–23652.

(221) Fischbach, M. A.; Clardy, J. One pathway, many products. *Nat. Chem. Biol.* **2007**, *3* (7), 353–355.

(222) Austin, M. B.; O'Maille, P. E.; Noel, J. P. Evolving biosynthetic tangos negotiate mechanistic landscapes. *Nat. Chem. Biol.* **2008**, *4* (4), 217–222.

(223) Brown, S. D.; Babbitt, P. C. New insights about enzyme evolution from large scale studies of sequence and structure relationships. *J. Biol. Chem.* **2014**, *289* (44), 30221–30228.

(224) Kourist, R.; Bartsch, S.; Fransson, L.; Hult, K.; Bornscheuer, U. T. Understanding promiscuous amidase activity of an esterase from *Bacillus subtilis*. *ChemBioChem* **2008**, *9* (1), 67–69.

(225) Godehard, S. P.; Badenhorst, C. P. S.; Müller, H.; Bornscheuer, U. T. Protein engineering

for enhanced acyltransferase activity, substrate scope and selectivity of the *Mycobacterium smegmatis* acyl-transferase MsAcT. *ACS Catal.* **2020**, *10*, 7552–7562.

(226) Quest Graph™ IC50 Calculator https://www.aatbio.com/tools/ic50-calculator (accessed Jul 9, 2020).

(227) Sloan, M. J.; Phillips, R. S. Enzymatic synthesis of aza-L-tryptophans: the preparation of 5- and 6-aza-L-tryptophan. *Bioorg. Med. Chem. Lett.* **1992**, *2* (9), 1053–1056.

(228) Herger, M.; Van Roye, P.; Romney, D. K.; Brinkmann-Chen, S.; Buller, A. R.; Arnold, F. H. Synthesis of β-branched tryptophan analogues using an engineered subunit of tryptophan synthase. *J. Am. Chem. Soc.* **2016**, *138* (27), 8388–8391.

(229) Romney, D. K.; Murciano-calles, J.; Wehrmüller, J.; Arnold, F. H. Unlocking reactivity of TrpB: a general biocatalytic platform for synthesis of tryptophan analogs. *J. Am. Chem. Soc.* **2017**, *23*, 22–23.

(230) Boville, C. E.; Romney, D. K.; Almhjell, P. J.; Sieben, M.; Arnold, F. H. Improved synthesis of 4-cyanotryptophan and other tryptophan analogues in aqueous solvent using variants of TrpB from Thermotoga maritima. *J. Org. Chem.* **2018**, *83* (14), 7447–7452.

(231) Buller, A. R.; Van Roye, P.; Murciano-Calles, J.; Arnold, F. H. Tryptophan synthase uses an atypical mechanism to achieve substrate specificity. *Biochemistry* **2016**, *55* (51), 7043–7046.

(232) Boville, C. E.; Scheele, R. A.; Koch, P.; Brinkmann-Chen, S.; Buller, A. R.; Arnold, F. H. Engineered biosynthesis of β-alkyl tryptophan analogues. *Angew. Chem. Int. Ed.* **2018**, *57* (45), 14764–14768.

(233) Malmstrom, B. G. Enzymology of Oxygen. *Annu. Rev. Biochem.* **1982**, *51* (1), 21–59.

(234) Jambunathan, P.; Zhang, K. Novel pathways and products from 2-keto acids. *Curr. Opin. Biotechnol.* **2014**, *29* (1), 1–7.

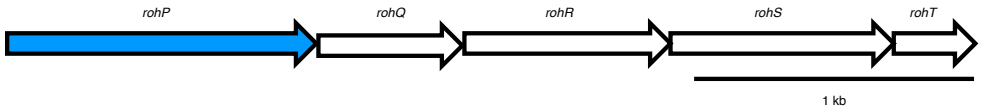
(235) Li, T.; Cui, X.; Cui, Y.; Sun, J.; Chen, Y.; Zhu, T.; Li, C.; Li, R.; Wu, B. Exploration of Transaminase Diversity for the Oxidative Conversion of Natural Amino Acids into 2-Ketoacids and High-Value Chemicals. *ACS Catal.* **2020**, *10* (14), 7950–7957.

(236) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci.* **2001**, *98* (18), 10037–10041.

Appendices

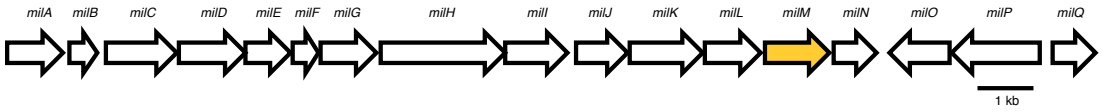
Appendix A: Supplementary information for Chapter 3

a



UniProtKB ID	Size	Protein Family	Predicted Function
G8WNK6	367	PF00155	*rohP*; arginine hydroxylase
F8JPS0	169	N/A	*rohQ*; guanidinoacetaldehyde cyclization
F8JVZ6	267	PF00701	*rohR*; retroaldolase
G8WNK3	289	PF14518	*rohS*; amino oxidase
F8JVZ5	95	PF00355	*rohT*; unknown

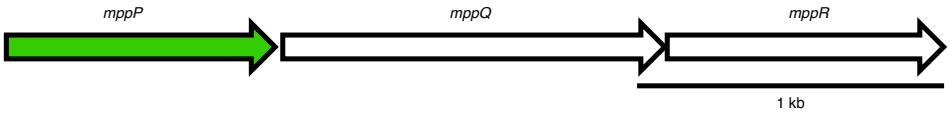
b



UniProtKB ID	Size	Protein Family	Predicted Function
B4Y380	333	PF00303	*milA*; CMP 5-hydroxymethylase
B4Y381	169	PF05014	*milB*: CMP/hydroxymethyl CMP hydrolase
H9BDW2	419	N/A	*milC*; cytosylglucuronic acid/hydroxymethyl cytosylglucuronic acid synthase
H9BDW3	394	PF01041	*milD*; degT/dnrJ/eryC1/strS aminotransferase
H9BDW4	272	PF01636	*milE*; aminoglycoside phosphotransferase
H9BDW5	156	N/A	*milF*; hypothetical protein
H9BDW6	334	PF04055/PF13353	*milG*; radical SAM superfamily
H9BDW7	739	N/A	*milH*; hypothetical protein
H9BDW8	359	PF00550	*milI*; chitinase A
H9BDW9	316	N/A	*milJ*; hypothetical protein
H9BDX0	441	PF07690	*milK*; major facilitator superfamily
H9BDX1	327	N/A	*milL*; hypothetical protein
H9BDX2	388	PF00155	*milM*; aspartate/tyrosine/aromatic aminotransferase
H9BDX3	257	PF00701	*milN*; dihydrodipicolinate synthetase family protein
H9BDX4	355	PF00196	*milO*: regulatory protein LuxR family
H9BDX5	526	PF00005	*milP*: ABC transporter superfamily
H9BDX6	255	PF01636	*milQ*: aminoglycoside phosphotransferase

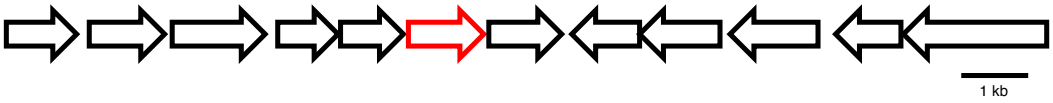
Appendix Figure A-1: Example gene clusters with annotated gene functions for Group Ia. (**a**) Partial azomycin gene cluster from *Streptomyces cattleya* strain ATCC 35852 and table of annotated functions. (**b**) *Streptomyces rimofaciens* mildiomycin gene cluster and table of annotated functions. Sizes are given in number of amino acids.

a



UniProtKB ID	Size	Protein Family	Predicted Function
Q643C0	292	PF00155	*mppP*; PLP-dependent arginine hydroxylase
Q643B9	414	PF00155	*mppQ*; PLP-dependent aminotransferase
Q643B8	301	PF06314	*mppR*; dehydratase/cyclase

b



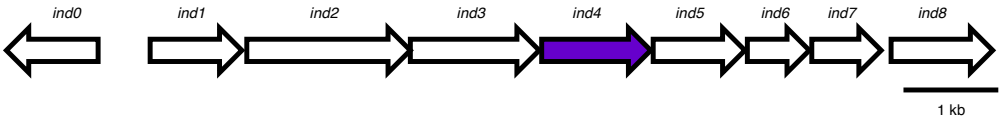
UniProtKB ID	Size	Protein Family	Predicted Function
A0A3R8QEI4	355	PF02826	dihydrofolate reductase
A0A3R8SFS3	391	N/A	hypothetical protein
A0A3R8RHL4	474	PF03241/PF11794	4-hydroxyphenylacetate 3-monooxygenase
A0A426SBR1	317	PF08241	SAM-dependent methyltransferase
A0A426SC37	318	PF00491	agmatinase
A0A3R8QJY3	386	PF00155	hypothetical protein
A0A426SC80	371	PF04339	hypothetical protein
A0A3R8RFS1	339	PF00389/PF02826	phosphoglycerate dehydrogenase family protein
A0A426SBX5	404	PF00465	hypothetical protein
A0A3R8QDV1	452	N/A	transcriptional regulator
A0A3R8RNY0	340	PF11583	hypothetical protein
A0A3R8S471	719	PF00733/PF13537	asparagine synthase (Glutamine-hydrolyzing)

c

UniProtKB ID	Size	Protein Family	Predicted Function
C7Q945	884	PF01804	peptidase S45 penicillin amidase
C7Q944	2249	PF00501/PF00550/ PF00668	amino acid adenylation domain protein
C7Q943	389	PF00155	aminotransferase
C7Q942	342	PF02668	L-ornithine/L-arginine 3-hydroxylase
C7Q941	376	PF00155	aminotransferase class I and II
C7Q940	1338	PF00501/PF00550/ PF00668/PF00975/ PF13193	amino acid adenylation domain protein
C7Q939	192	PF01738	hypothetical protein
C7Q938	226	PF00196	Transcriptional regulator, LuxR family
C7Q937	379	PF00155	aminotransferase class I and II

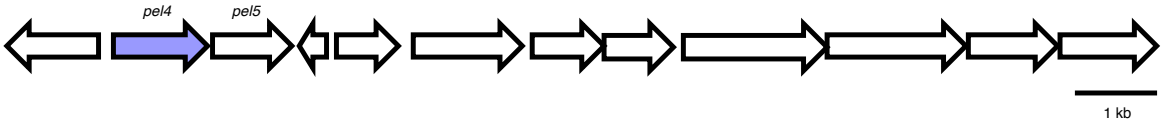
Appendix Figure A-2: Example gene clusters with annotated gene functions for Group II. (**a**) *Streptomyces hygroscopicus* enduracididine gene cluster and table of annotated functions. (**b**) *Streptomyces griseofuscus* unknown gene cluster and table of annotated functions. (**c**) *Catenulispora acidiphila* unknown gene cluster and table of annotated functions. Sizes are given in number of amino acids.

a



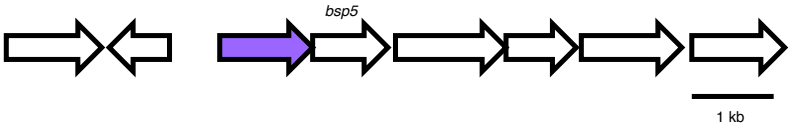
UniProtKB ID	Size	Protein Family	Predicted Function
A0A0D4BSA1	320	PF00579	*ind0*; tryptophanyl-tRNA synthase
A0A0D4BS77	328	PF13649	*ind1*; indolepyruvate C-methyltransferase
A0A0D4BSN8	569	N/A	*ind2*; beta-methylindole-3-pyruvate reductase
A0A0D4BST1	451	N/A	*ind3*; ATP-dependent amide synthetase
A0A0D4BS17	382	PF00155	*ind4*; PLP-dependent arginine desaturase
A0A0D4BSA6	318	PF00389/PF02826	*ind5*; NADPH-dependent imine reductase
A0A0D4BS80	228	N/A	*ind6*; chaperone-like
A0A0D4BSP3	236	PF13649	*ind7*; N-demethylindolmycin N-methyltransferase
A0A0D4BST6	360	PF00155	*ind8*; aminotransferase

b



UniProtKB ID	Size	Protein Family	Predicted Function
A0A161S599	366	PF02875/PF08245	Mur ligase
A0A165QAT3	381	PF00155	*pel4*; aminotransferase class I and II
A0A163V3I0	318	PF00389/PF02826	*pel5*; imine/ketoacid reductase
A0A163V4B7	109	N/A	protein tyrosine phosphatase
A0A165QAS7	256	PF00455/PF08220	DeoR family transcriptional regulator
A0A163V3H0	440	PF01547	ABC transporter substrate-binding protein
A0A163V4A6	292	PF00528	ABC transporter permease
A0A163V3F4	272	PF00528	sugar ABC transporter permease
A0A165QAR7	583	PF00672	histidine kinase
A0A163V3E8	553	PF00072/PF12833	AraC family transcriptional regulator
A0A163V3D9	367	PF01168/PF14031	amino acid processing protein
A0A163V3C6	388	PF01546/PF07687	acetylornithine deacetylase

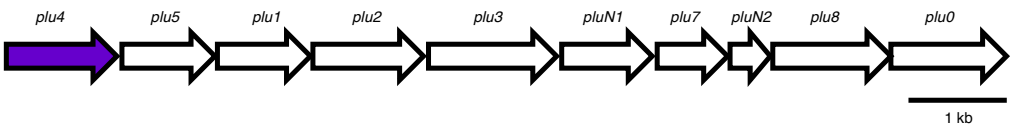
c



UniProtKB ID	Size	Protein Family	Predicted Function
A0A1I4FW50	390	PF00108/PF02803	acetyl-CoA acetyltransferase
A0A1I4FVH3	249	PF13304	predicted ATPase
A0A1I4FUF9	383	PF00155	aminotransferase class I and II
A0A1I4FUG4	305	PF00389/PF02826	*bsp5*; imine reductase
A0A1I4FVH2	455	N/A	phenylacetate-CoA ligase
A0A1I4FXW7	291	PF08443	glutathione synthase/RimK-type ligase, ATP-grasp superfamily
A0A1I4FUJ5	406	PF13535	ATP-grasp domain-containing protein
A0A1I4FXS2	387	PF07690	predicted arabinose efflux permease, MFS family

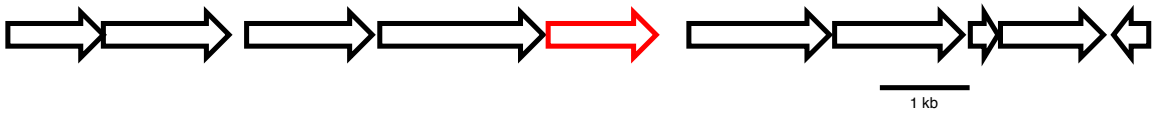
Appendix Figure A-3: Example gene clusters with annotated gene functions for Group Ib. (**a**) *Streptomyces griseus* indolmycin gene cluster and table of annotated functions. (**b**) *Paenibacillus elgii* unknown gene cluster and table of annotated functions. (**c**) *Bacillus* sp. 5mfcol3.1 unknown gene cluster and table of annotated functions. Sizes are given in number of amino acids.

a



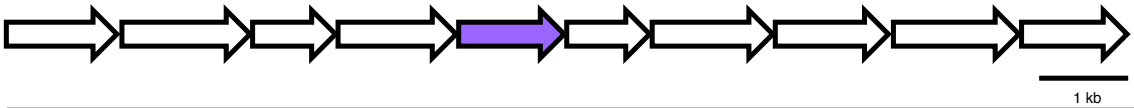
UniProtKB ID	Size	Protein Family	Predicted Function
A0A167HIA4	380	PF00155	*plu4*; PLP-dependent arginine desaturase
A0A167HID2	315	PF00389/PF02826	*plu5*; 2-ketoacid dehydrogenase
A0A167HIE7	318	PF13847	*plu1*; indolepyruvate C-methyltransferase
A0A167HIE9	370	N/A	*plu2*; beta-methylindole-3-pyruvate reductase
A0A167HIF2	441	N/A	*plu3*; ATP-dependent amide synthetase
A0A167HIF8	304	N/A	*pluN1*; unknown
A0A161YDL9	236	PF13649	*plu7*; N-demethylindolmycin N-methyltransferase
A0A167HII1	131	PF00903	*pluN2*; chaperone-like
A0A167HIJ3	394	PF00155	*plu8*; aminotransferase
A0A167HIK5	332	PF00579	*plu0*; tryptophanyl-tRNA synthase

b



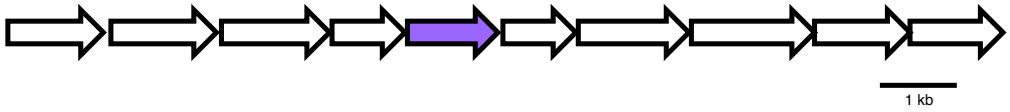
UniProtKB ID	Size	Protein Family	Predicted Function
A0A0D5XXF8	356	PF00389/PF02826	hydroxyacid dehydrogenase
A0A0D5XWM9	460	PF01565/PF08031	FAD-binding protein
A0A0D5XXQ4	459	PF01554	multidrug transporter MATE
A0A0D5XWP5	607	N/A	hypothetical protein
A0A0D5XXS7	389	PF00155	aspartate/tyrosine/aromatic aminotransferase
A0A0D5XXG3	526	PF09098/PF09099/PF09100/PF14930	quinohemoprotein amine dehydrogenase, alpha subunit
A0A0D5XWN3	476	PF04055/PF13353	quinohemoprotein amine dehydrogenase
A0A0D5XXQ9	106	PF08992	quinohemoprotein amine dehydrogenase, gamma subunit
A0A0D5XWQ0	373	N/A	quinohemoprotein amine dehydrogenase subunit beta PeaD
A0A0D5XXT1	132	PF07007	urease-associated protein

c



UnProtKB ID	Size	Protein Family	Predicted Function
A0A1H0X398	423	PF13535	ATP-grasp domain-containing protein
A0A1H0X1Y3	479	N/A	phenylacetate-CoA ligase
A0A1H0X1U5	319	PF02746/PF13378	dipeptide epimerase
A0A1H0X1W9	445	PF00464	serine hydroxymethyltransferase
A0A1H0X1S5	389	PF00155	aspartate/methionine/tyrosine aminotransferase
A0A1H0X2G9	318	PF00389/PF02826	gluconate 2-dehydrogenase
A0A1H0X1T9	450	PF01565	FAD/FMN-containing dehydrogenase
A0A1H0X1R4	423	N/A	hypothetical protein
A0A1H0X2G1	465	PF01554	Multidrug resistance protein, MATE family
A0A1H0X1T0	401	PF13535	biotin carboxylase

d

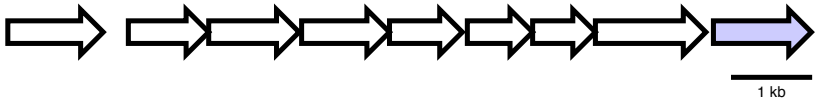


UniProtKB ID	Size	Protein Family	Predicted Function
A0A3M3S133	414	PF00464	putative serine hydroxymethyltransferase
A0A3M3MBM7	461	PF01554	putative multidrug resistance protein NorM
A0A3M3MAJ3	475	N/A	CapK domain protein
A0A3M3MAS6	315	PF00291	Pyridoxal-phosphate dependent enzyme protein
A0A3M3MA06	389	PF00155	aminotransferase class I and II
A0A3M3S0Z5	312	PF00389/PF02826	D-isomer specific 2-hydroxyacid dehydrogenase protein
A0A3M3S0C0	470	PF01565	putative oxidoreductase, FAD-binding
A0A3M3S095	530	PF01593	Amine oxidase, flavin-containing protein
A0A3M3S0Z6	409	PF02771/PF08028	acyl-CoA dehydrogenase protein
A0A3M3S0G4	394	PF02771/PF08028	acyl-CoA dehydrogenase protein

e

UniProtKB ID	Size	Protein Family	Predicted Function
A0A1K1Q0V2	395	PF00155	aminotransferase class I and II
A0A1K1Q0S4	310	PF00389/PF02826	lactate dehydrogenase
A0A1K1Q2L4	255	PF06182	ABC-type uncharacterized transport system, permease component
A0A1K1Q0V4	261	N/A	ABC-2 type transport system permease protein
A0A1K1Q0A1	335	PF00005	ABC-2 type transport system ATP-binding protein
A0A1K1PZV1	486	PF00005	ATPase components of ABC transporters with duplicated ATPase domains
A0A1K1Q049	425	N/A	phenylacetate-CoA ligase

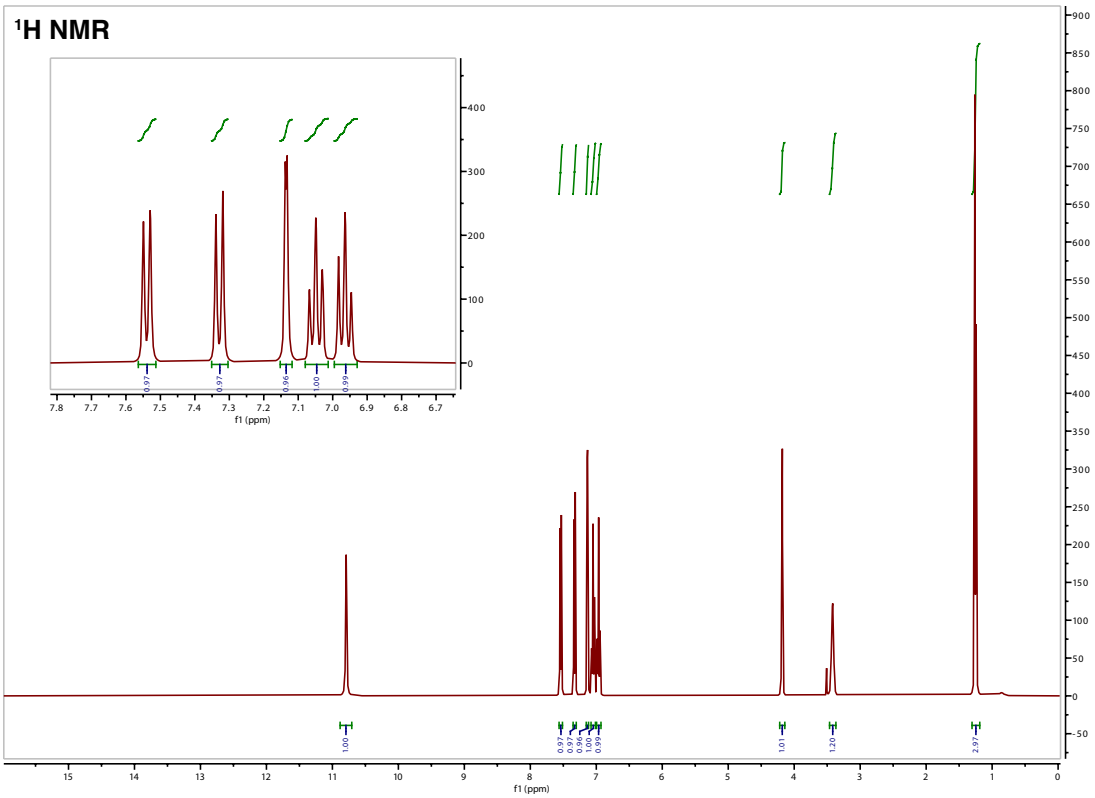
f



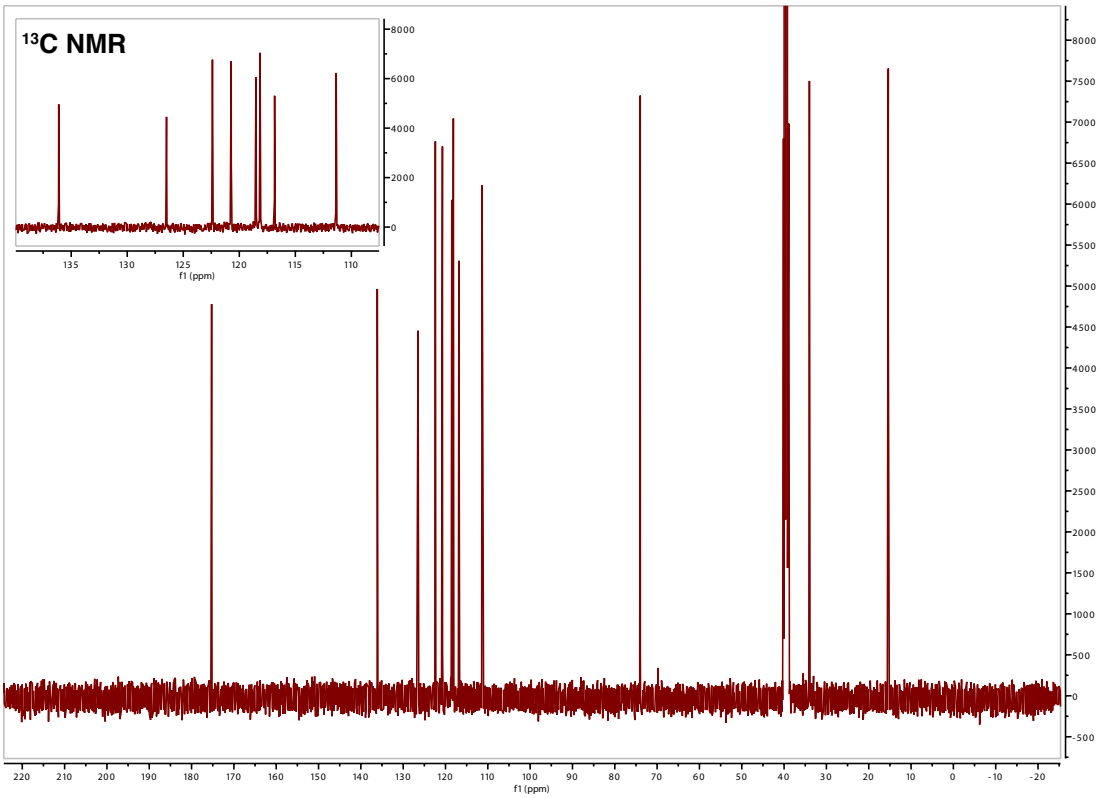
UniProtKB ID	Size	Protein Family	Predicted Function
A0A2S9KP32	395	PF00389/PF02826	hypothetical protein
A0A2S9KP93	323	PF01564	polyamine aminopropyltransferase
A0A2S9KP81	357	N/A	amidinotransferase
A0A2S9KP36	301	PF00005	ABC transporter
A0A2S9KP34	267	PF06182	ABC transporter permease
A0A2S9KP57	251	PF06182	membrane protein
A0A2S9KP38	447	PF00278/PF02784	putative diaminopimelate decarboxylase
A0A2S9KP24	407	PF00155	putative aspartate/tyrosine/aromatic aminotransferase

Appendix Figure A-4: Example gene clusters with annotated gene functions for Group III. (**a**) *Pseudoalteromonas luteoviolacea* CPMOR-1 indolmycin gene cluster and table of annotated functions. (**b**) *Pseudomonas chlororaphis* unknown gene cluster and table of annotated functions. (**c**) *Actinopolyspora xinjiangensis* unknown gene cluster and table of annotated functions. (**d**) *Pseudomonas syringae pv. Apii* unknown gene cluster and table of annotated functions. (**e**) *Pseudomonas sp.* NFACC04-2 unknown gene cluster and table of annotated functions. (**f**) *Burkholderia ambifaria* unknown gene cluster and table of annotated functions. Sizes are given in number of amino acids.

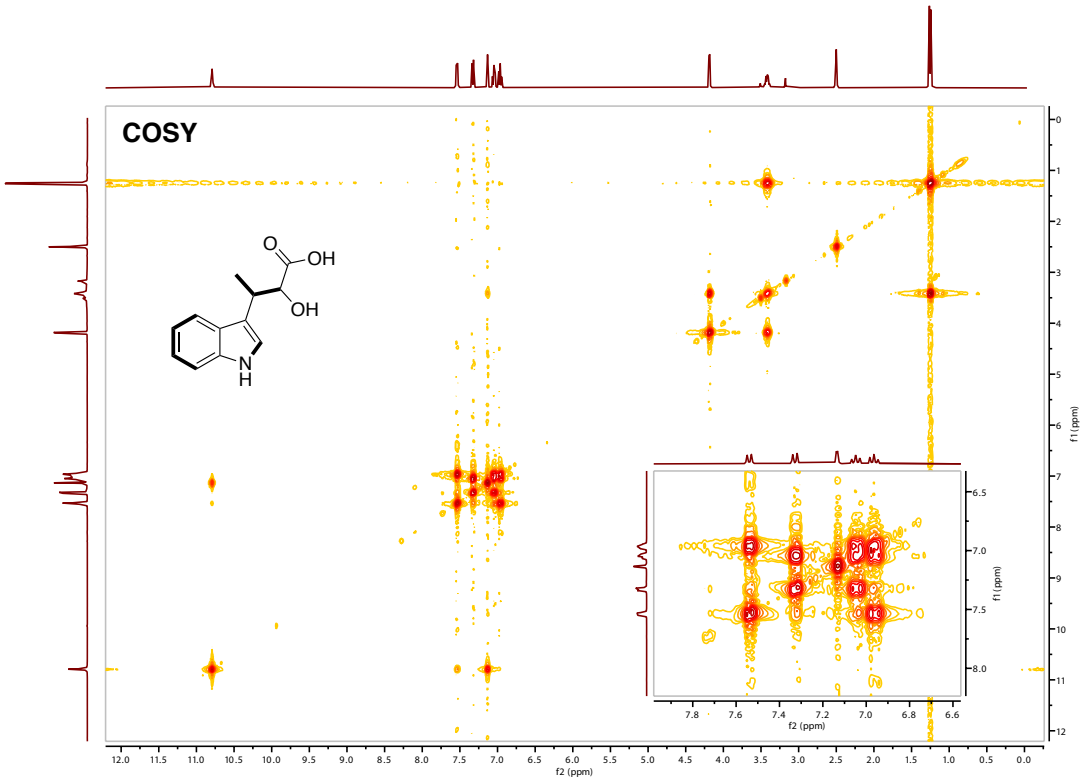
Appendix B: Supplementary information for Chapter 4



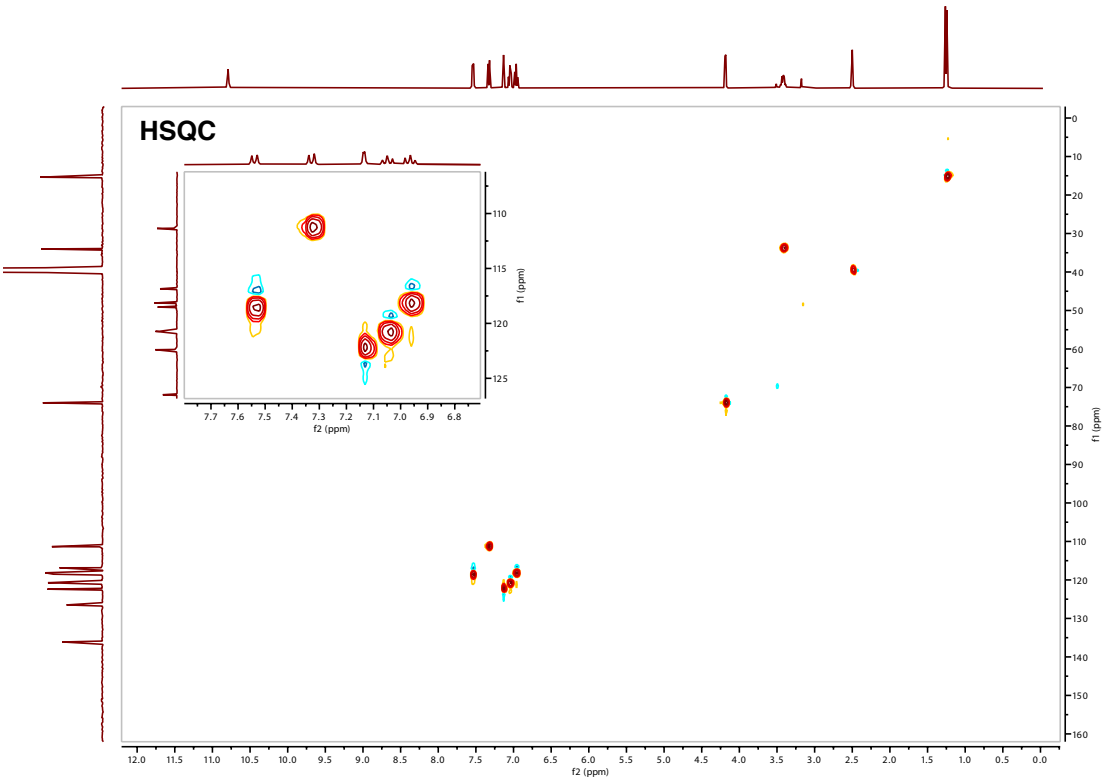
Appendix Figure B-1: ^{1}H NMR (400 MHz, DMSO-d_6) spectrum for indolmycenic acid (**11**).



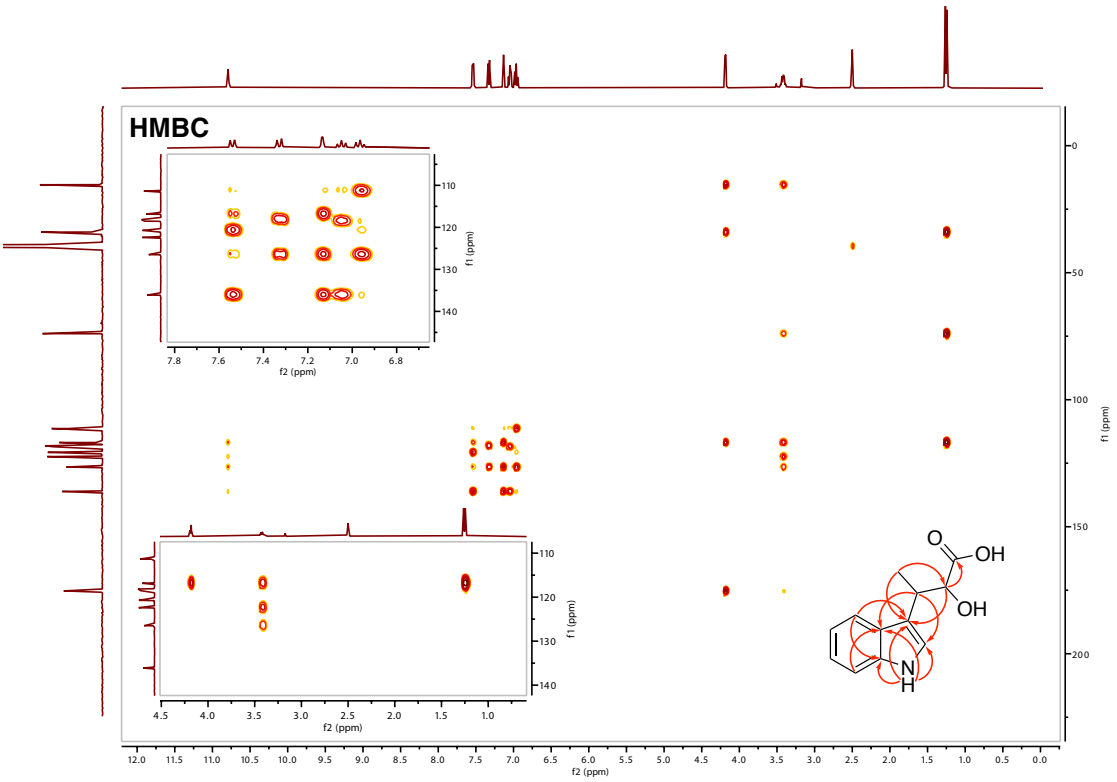
Appendix Figure B-2: ^{13}C NMR (101 MHz, DMSO-d_6) spectrum for **11**.



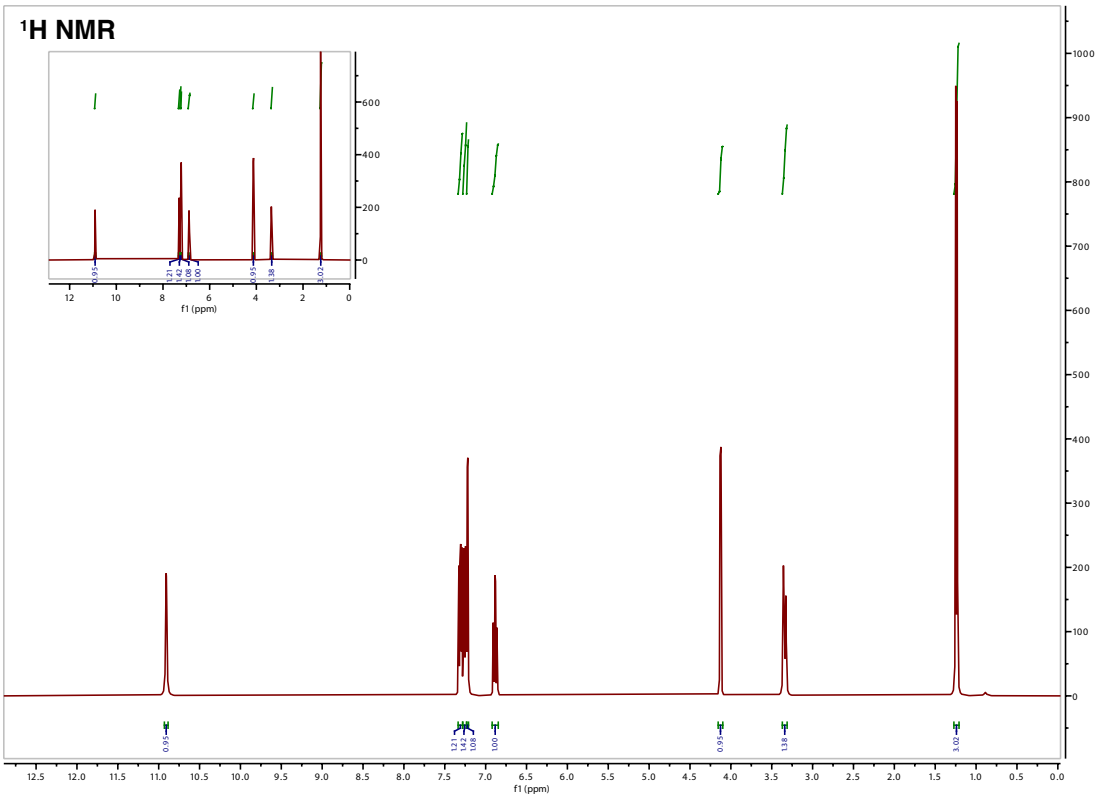
Appendix Figure B-3: COSY NMR (400 MHz, DMSO-d_6), spectrum for **11**.



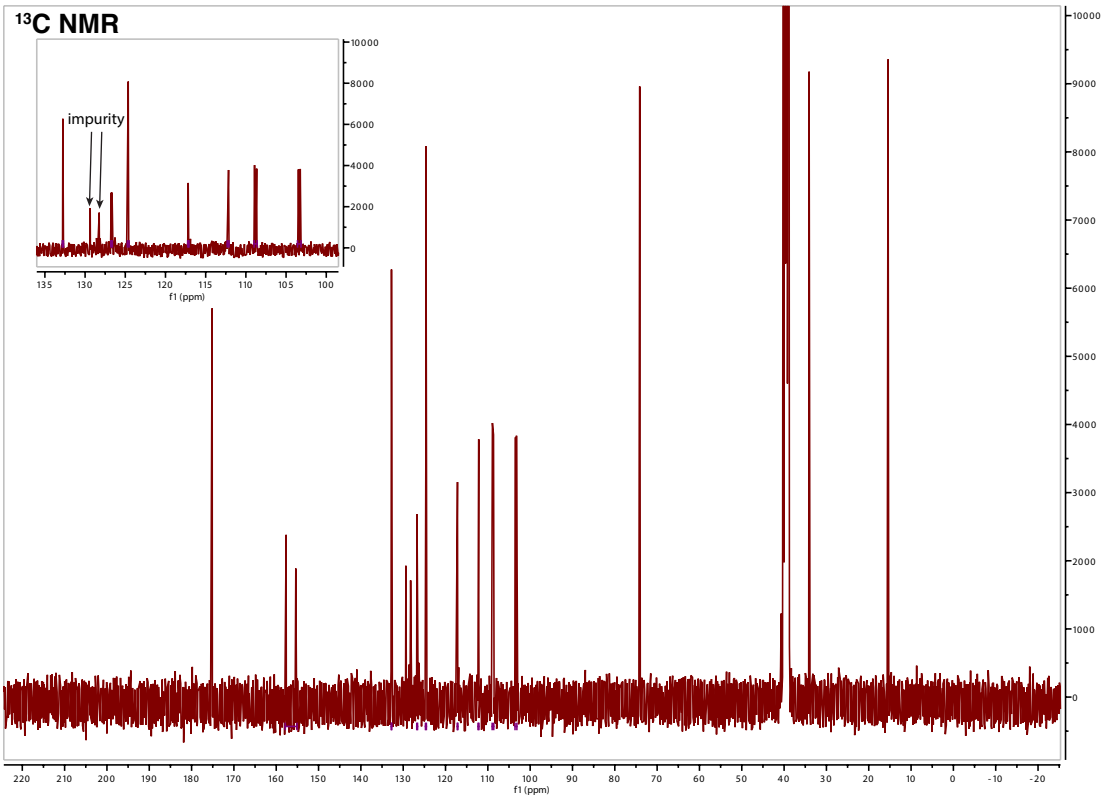
Appendix Figure B-4: HSQC NMR (400 MHz for ^{1}H NMR, 101 MHz for ^{13}C NMR, DMSO-d_6) spectrum for **11**.



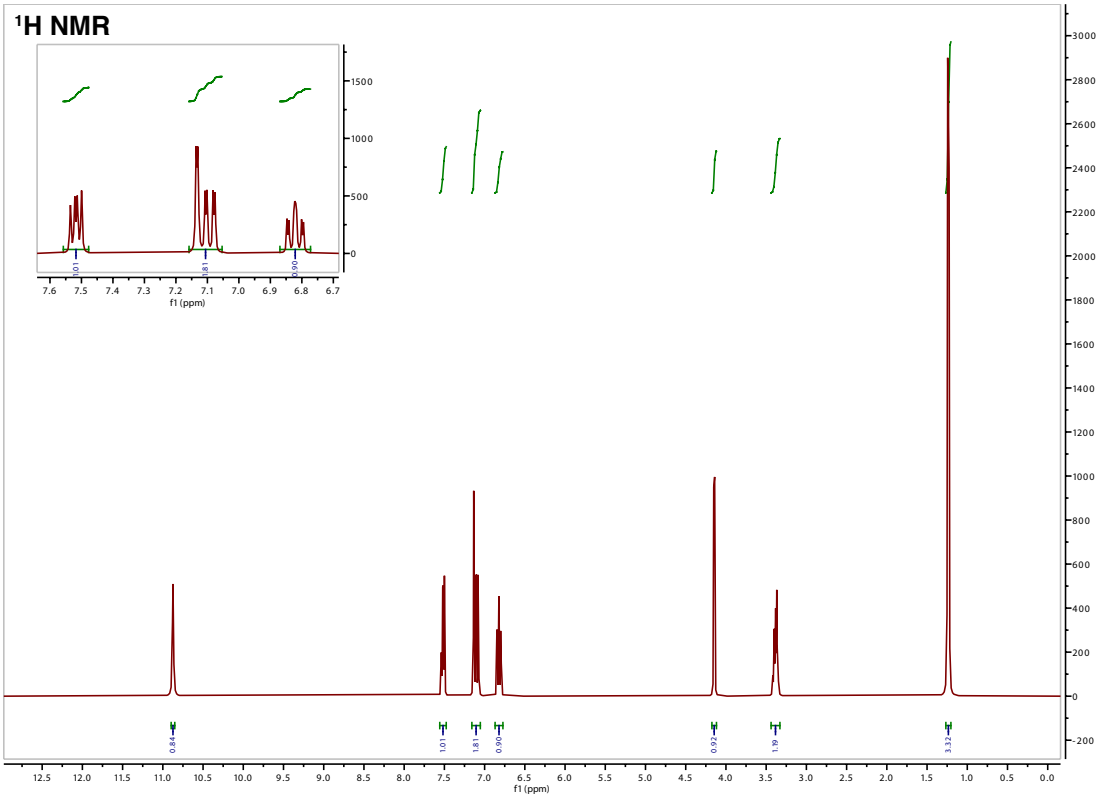
Appendix Figure B-5: HMBC NMR (400 MHz for ^{1}H NMR, 101 MHz for ^{13}C NMR, DMSO-d_6) spectrum for **11**.



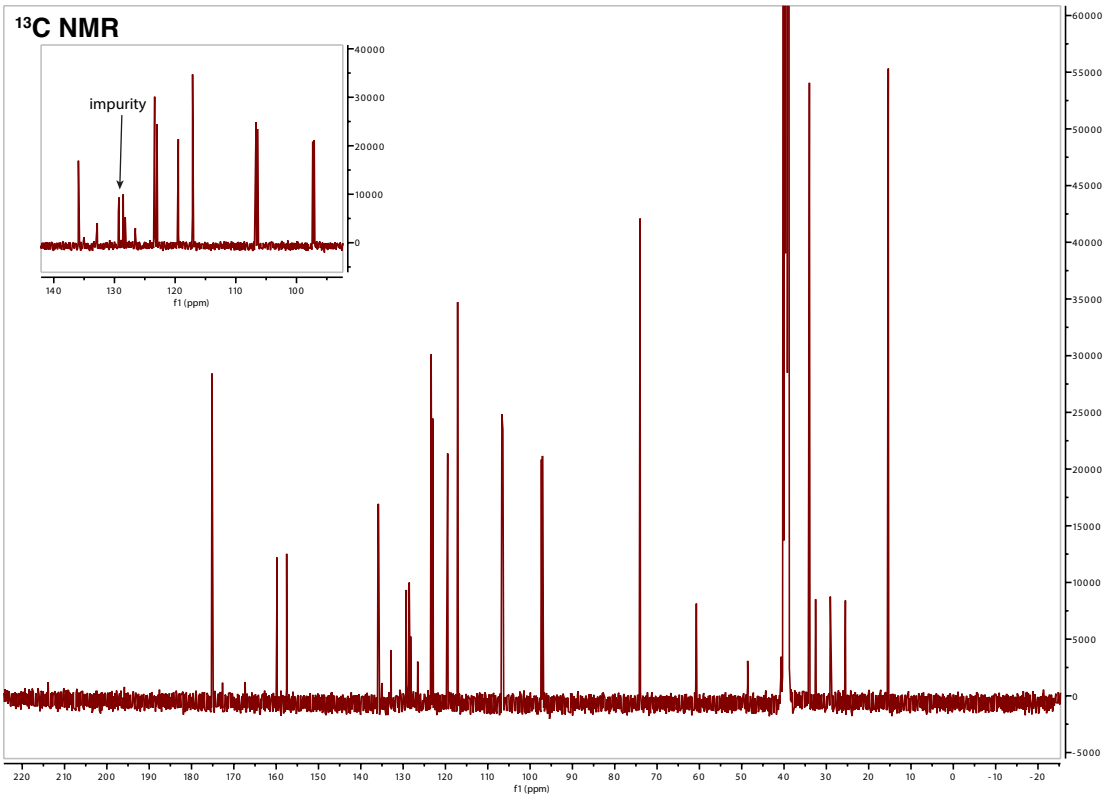
Appendix Figure B-6: ^{1}H NMR (400 MHz, DMSO-d_6) spectrum for 5-fluoro-indolmycenic acid (**5F-11**).



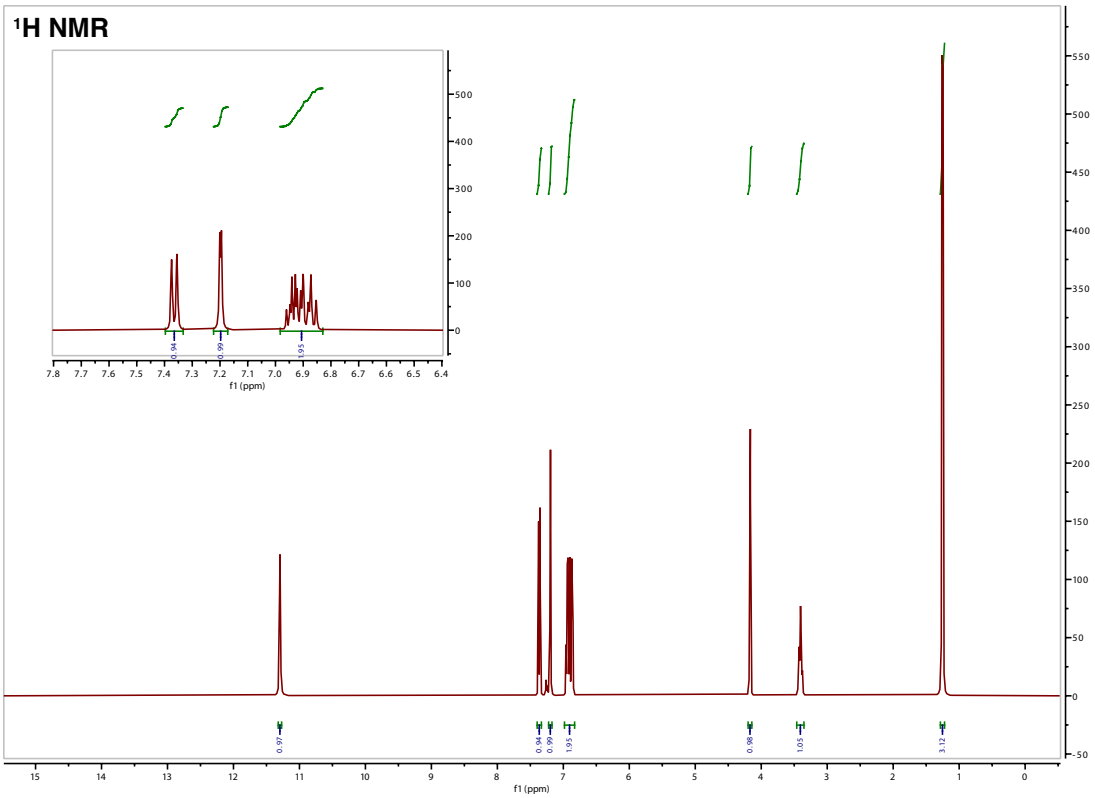
Appendix Figure B-7: ^{13}C NMR (101 MHz, DMSO-d_6) spectrum for **5F-11**.



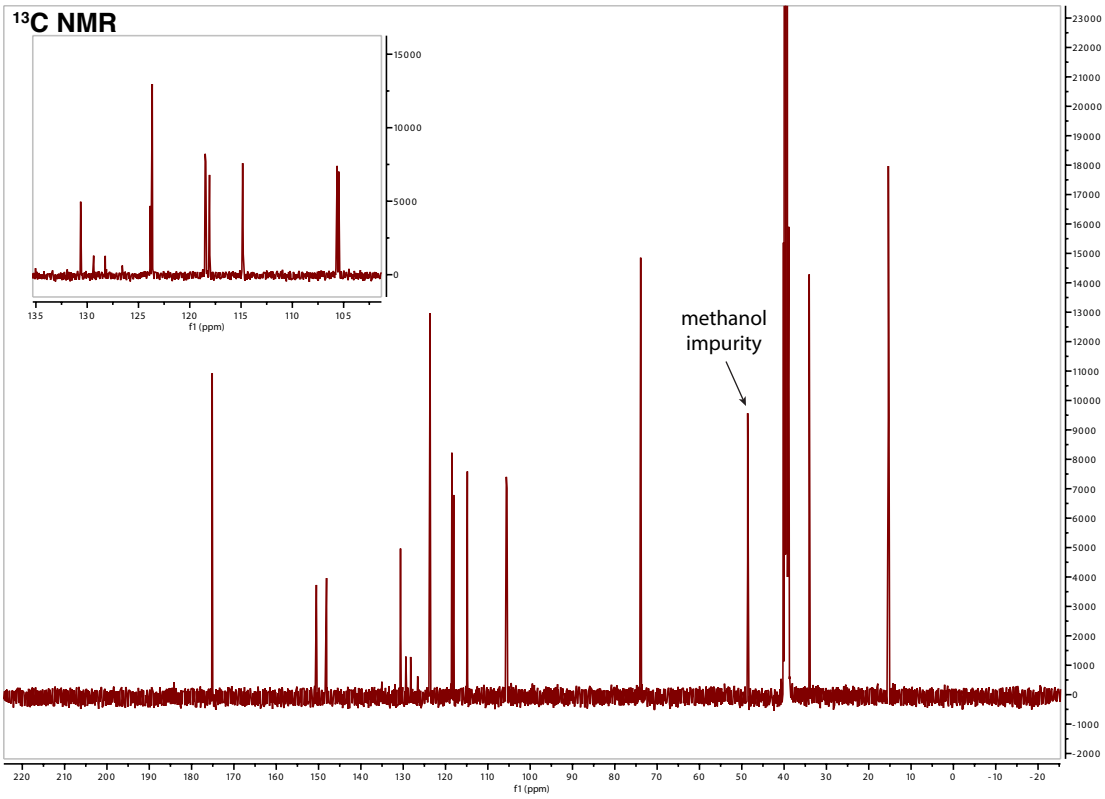
Appendix Figure B-8: ^{1}H NMR (400 MHz, DMSO-d_6) spectrum for 6-fluoro-indolmycenic acid (**6F-11**).



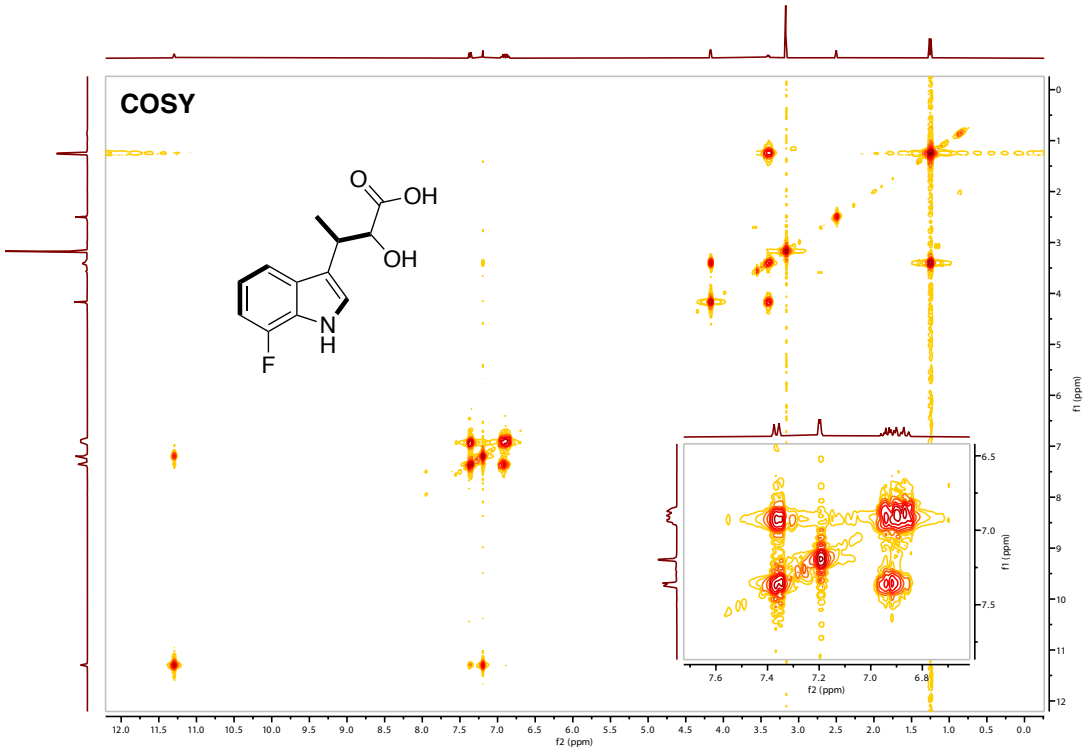
Appendix Figure B-9: ^{13}C NMR (101 MHz, DMSO-d_6) spectrum for **6F-11**.



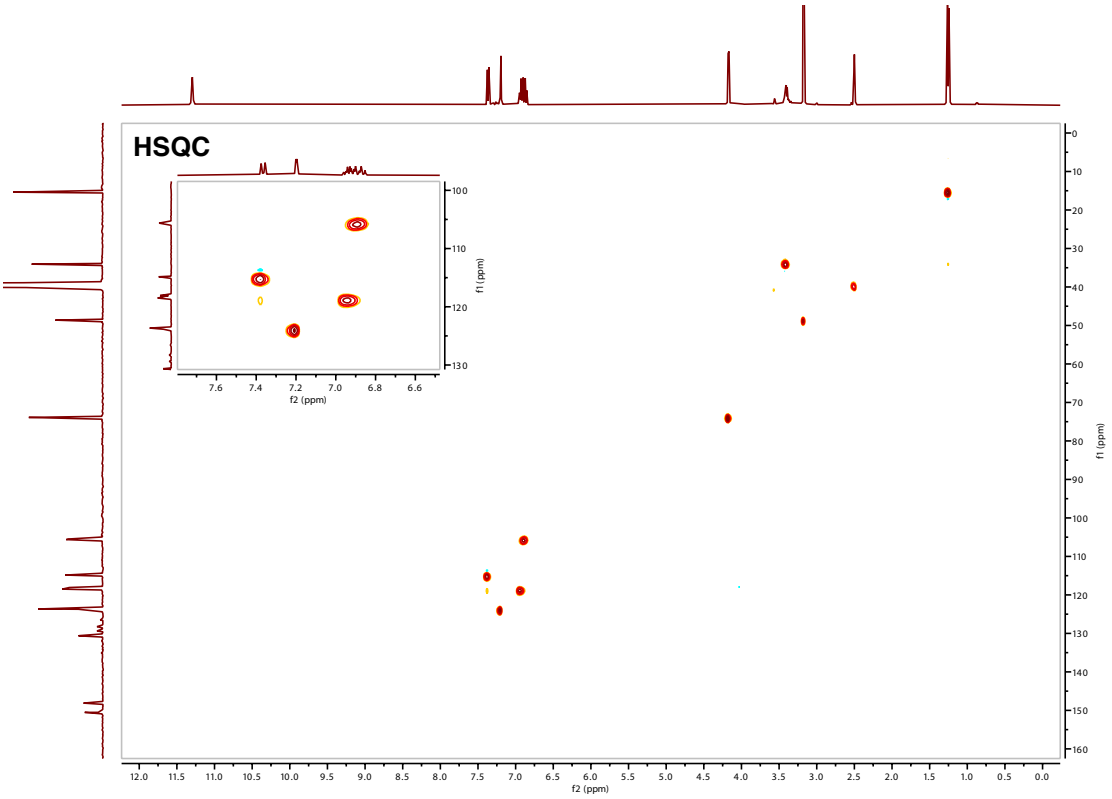
Appendix Figure B-10: ^{1}H NMR (400 MHz, DMSO-d_6) spectrum for 7-fluoro-indolmycenic acid (**7F-11**).



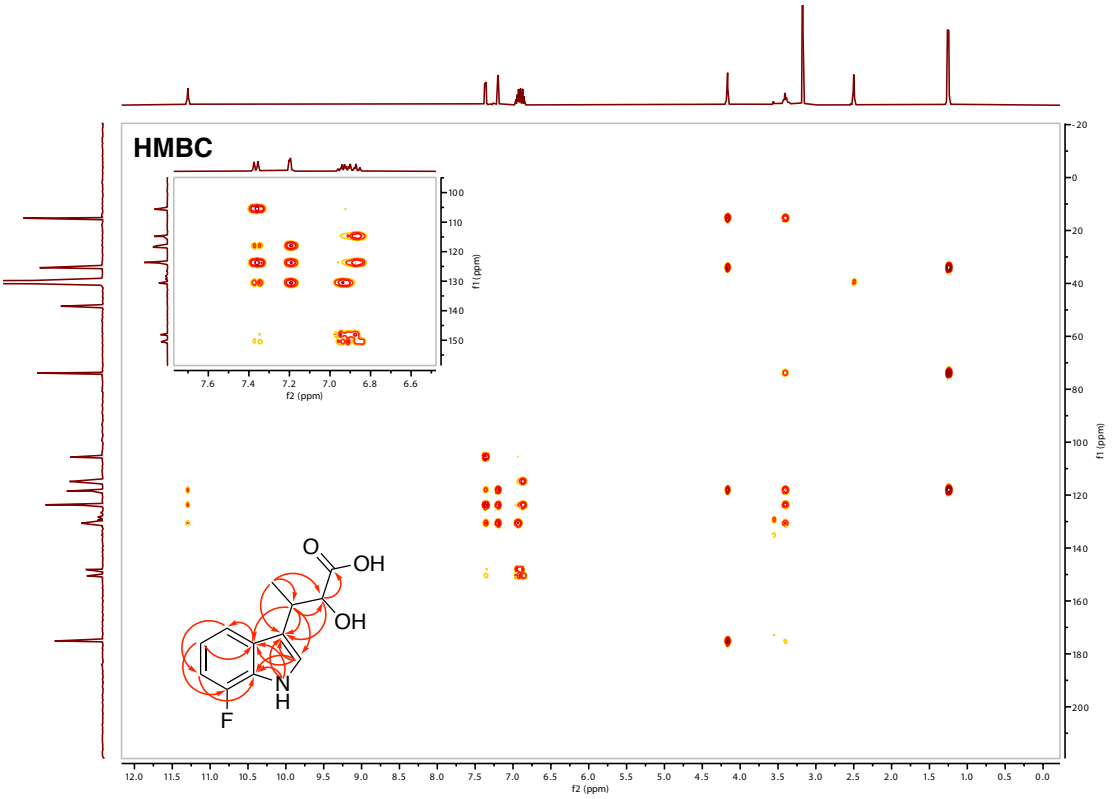
Appendix Figure B-11: ^{13}C NMR (101 MHz, DMSO-d_6) spectrum for **7F-11**.



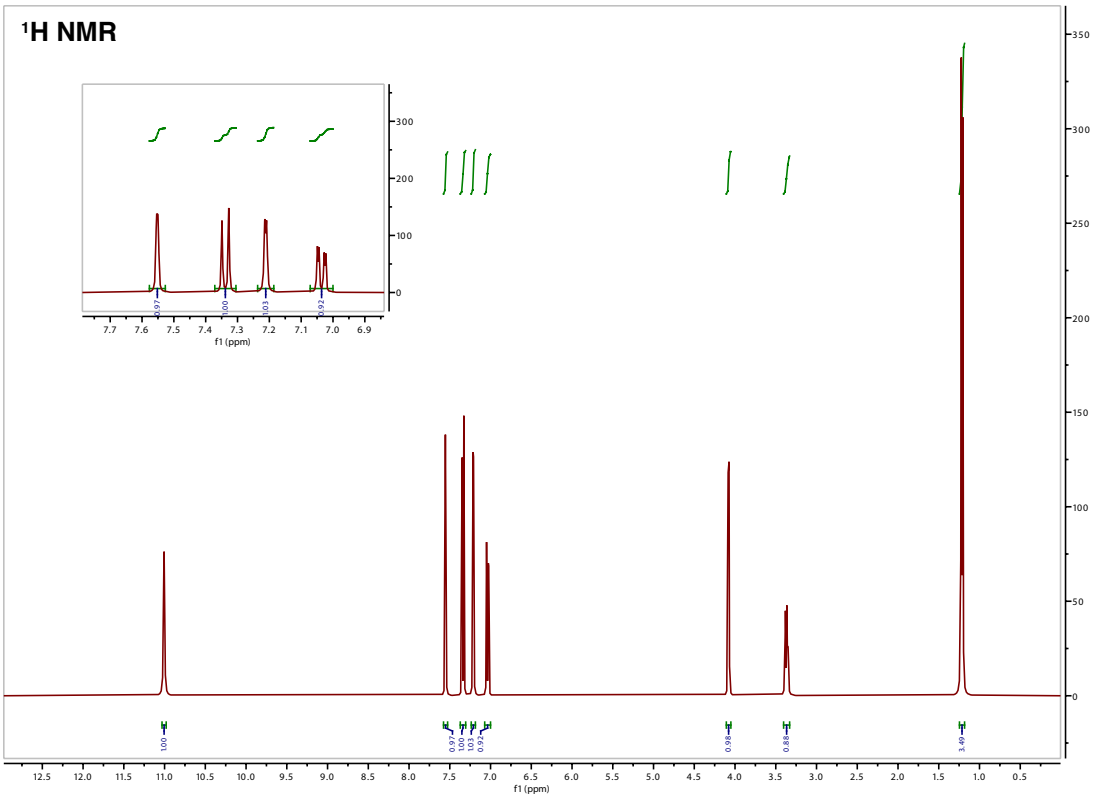
Appendix Figure B-12: COSY NMR (400 MHz, DMSO-d_6) spectrum for **7F-11**.



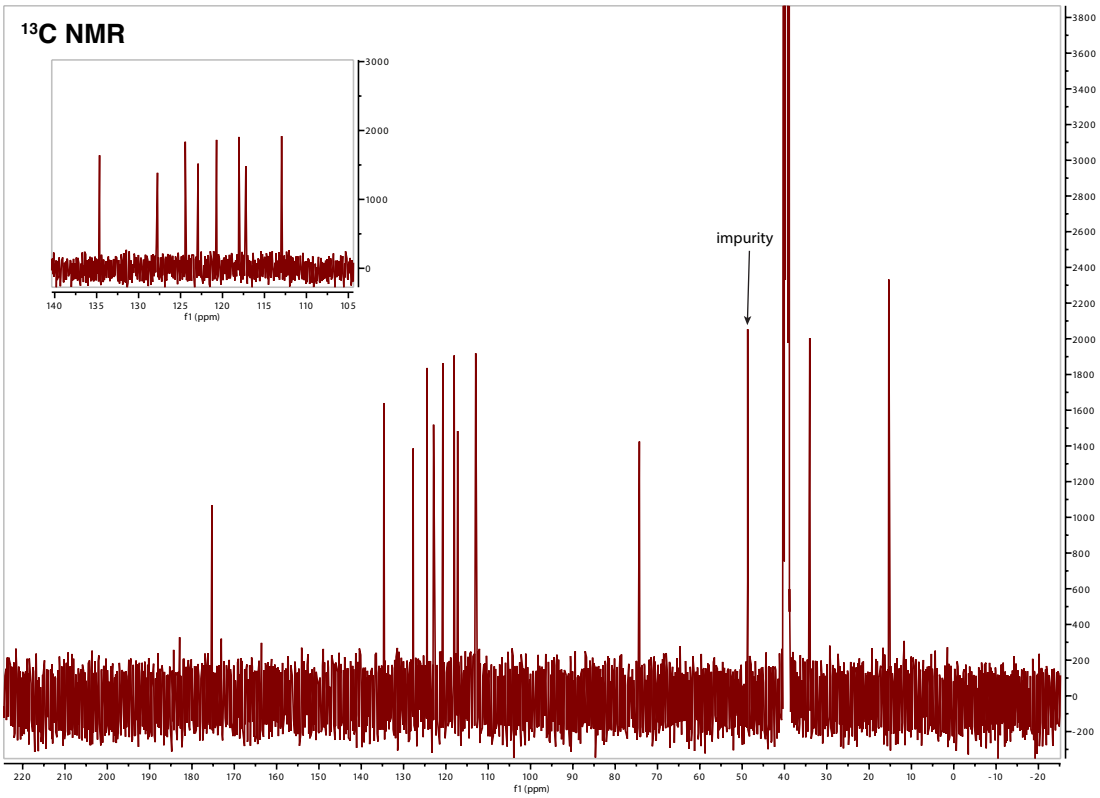
Appendix Figure B-13: HSQC NMR (400 MHz for ^{1}H NMR, 101 MHz for ^{13}C NMR, DMSO-d_6) spectrum for **7F-11**.



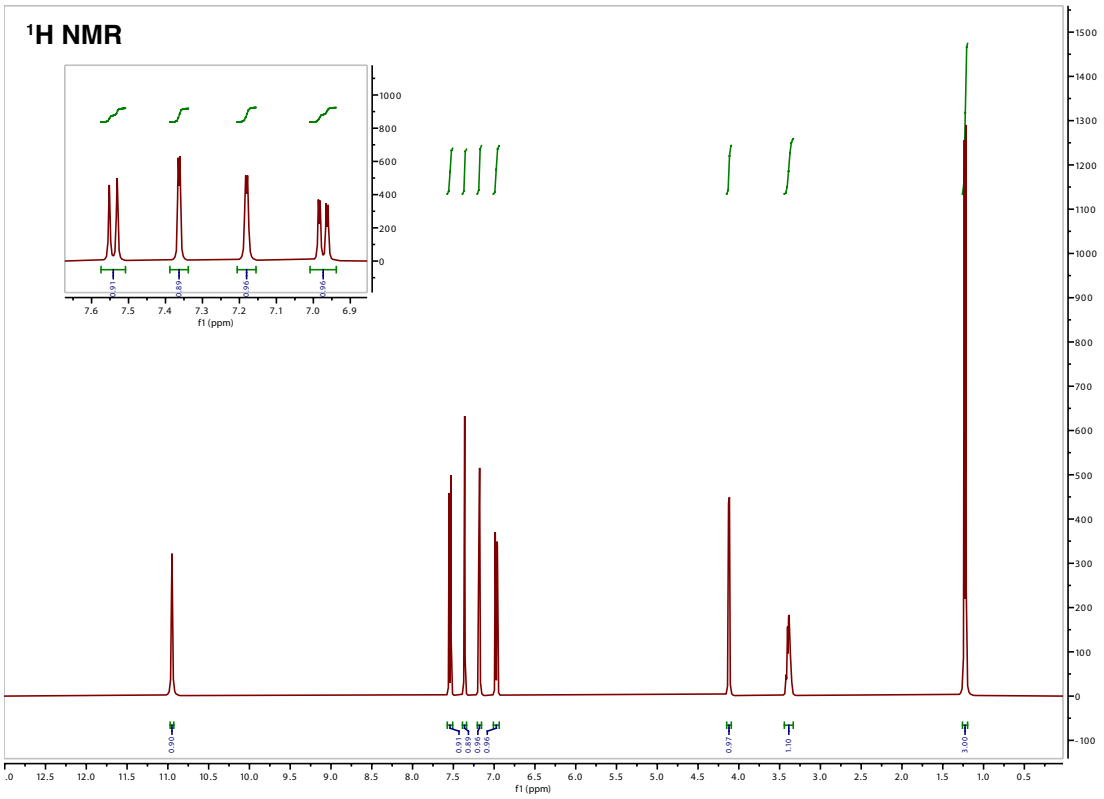
Appendix Figure B-14: HMBC NMR (400 MHz for ^{1}H NMR, 101 MHz for ^{13}C NMR, DMSO-d_6) spectrum for **7F-11**.



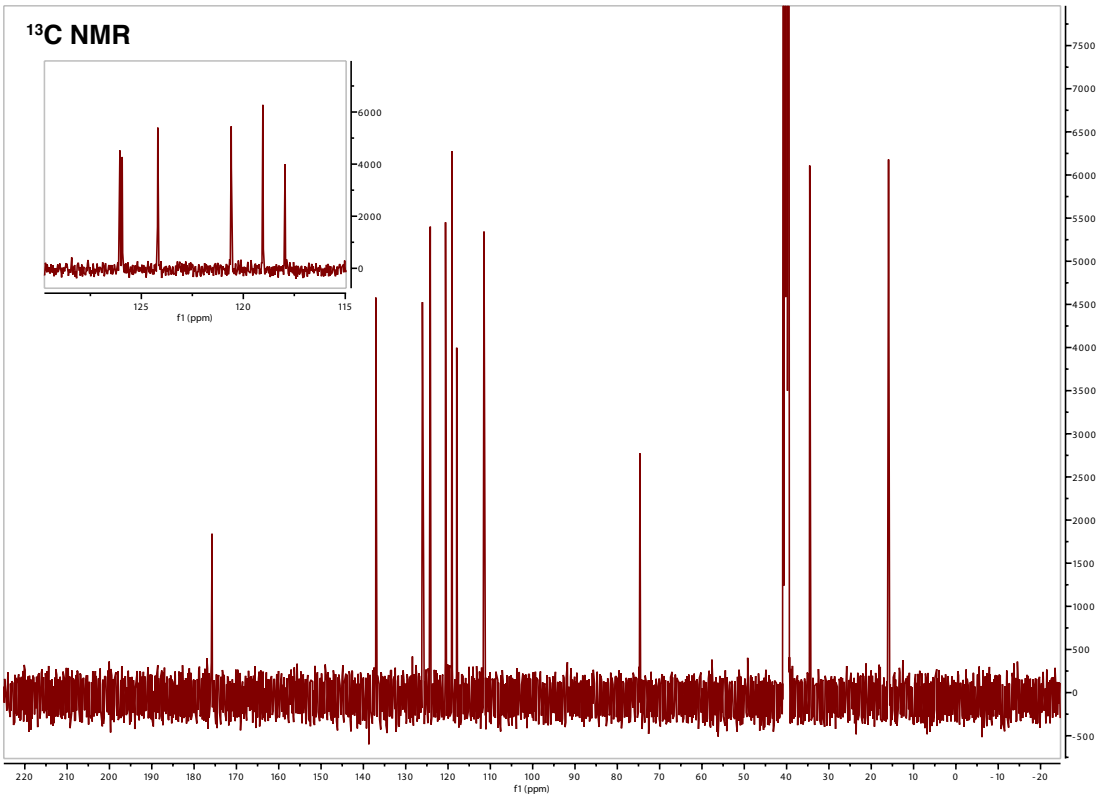
Appendix Figure B-15: ^{1}H NMR (400 MHz, DMSO-d_6) spectrum for 5-chloro-indolmycenic acid (**5Cl-11**).



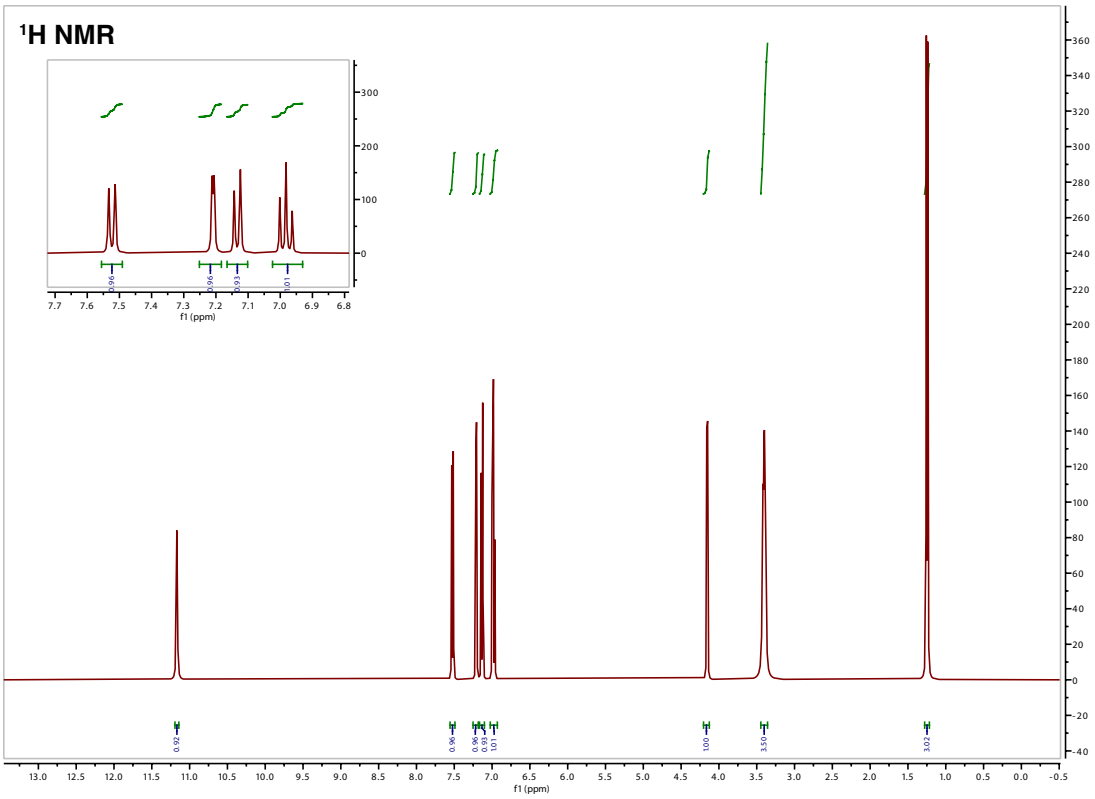
Appendix Figure B-16: ^{13}C NMR (101 MHz, DMSO-d_6) spectrum for **5Cl-11**.



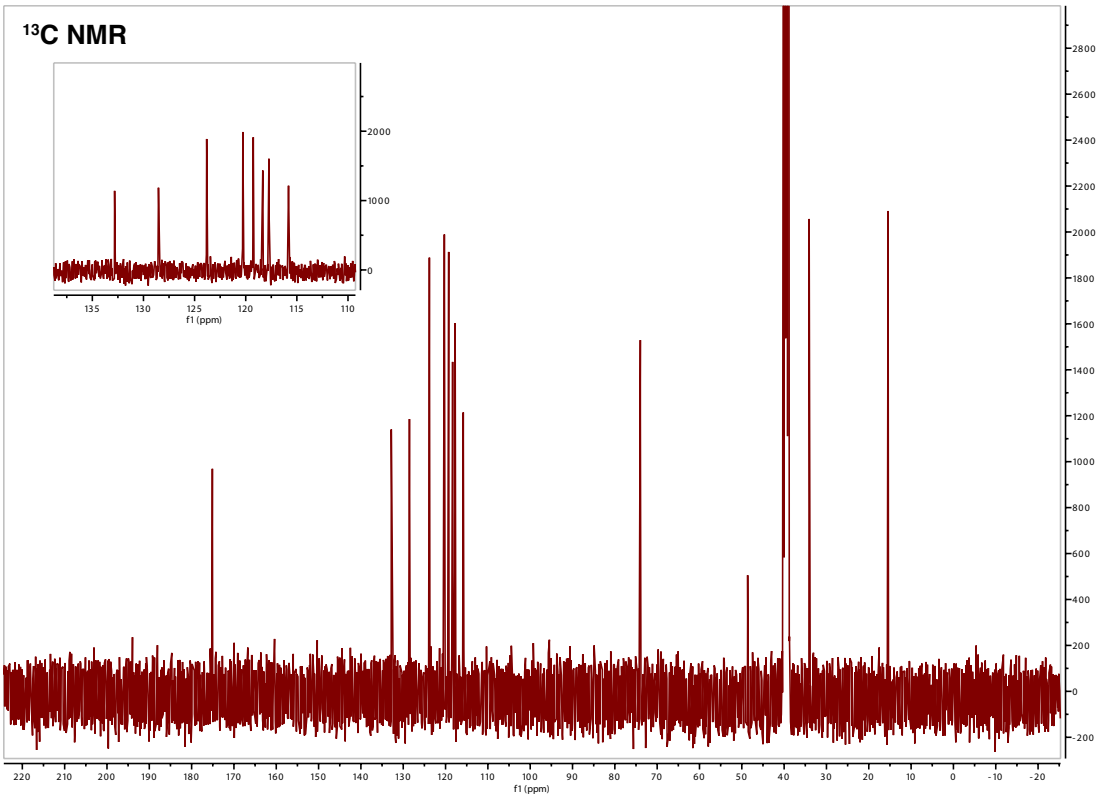
Appendix Figure B-17: ^{1}H NMR (400 MHz, DMSO-d_6) spectrum for 6-chloro-indolmycenic acid (**6Cl-11**).



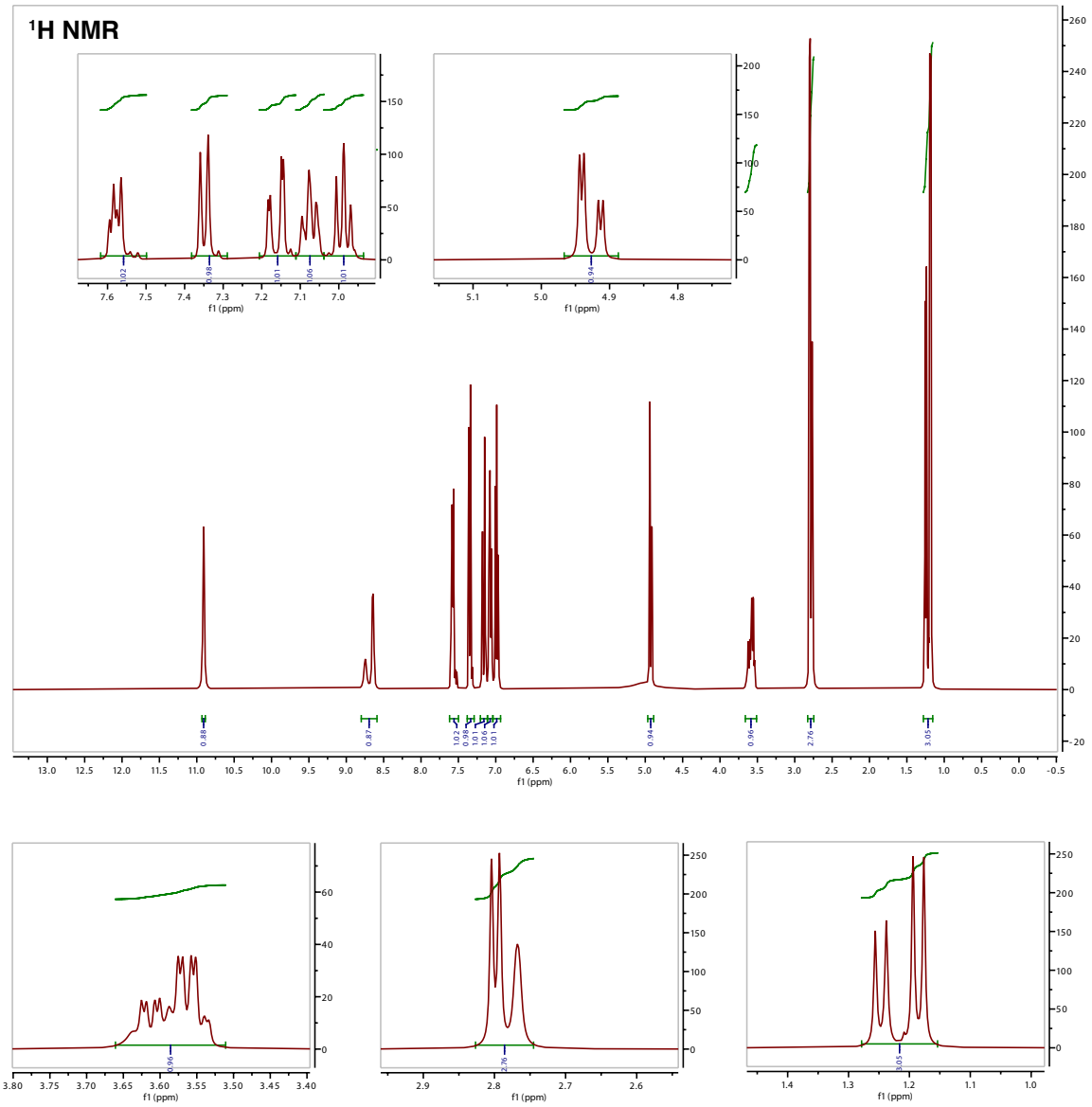
Appendix Figure B-18: ^{13}C NMR (101 MHz, DMSO-d_6) spectrum for **6Cl-11.**



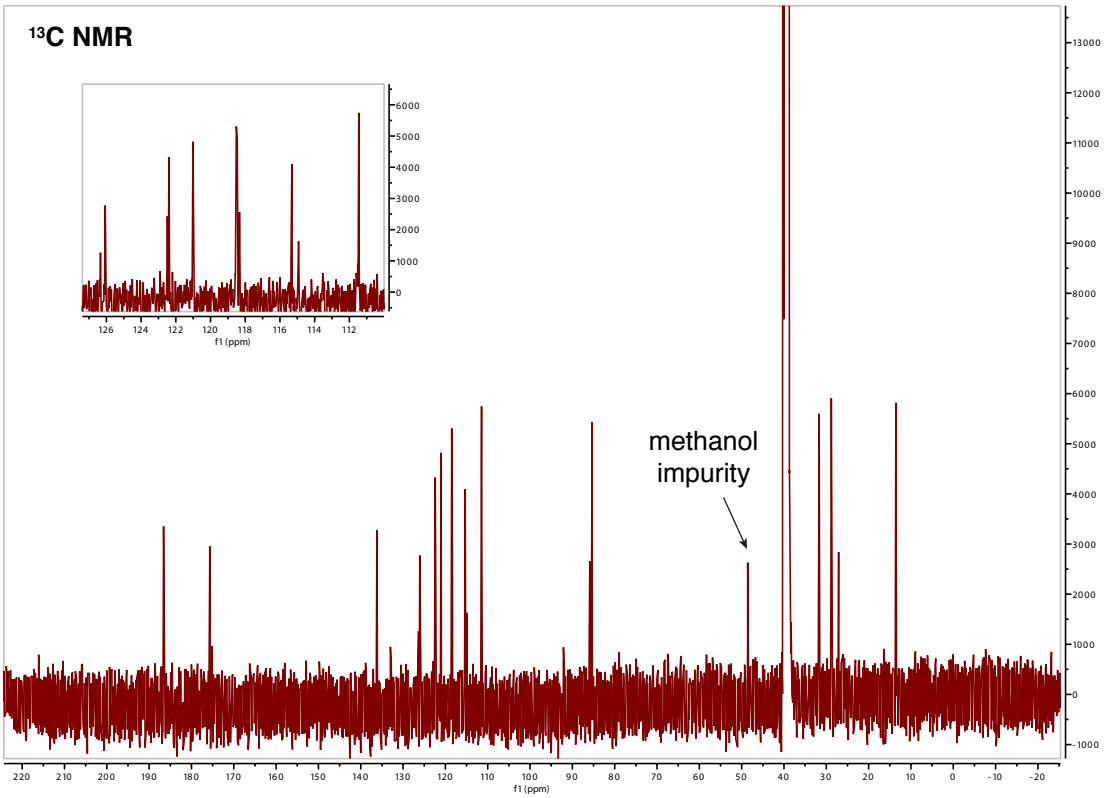
Appendix Figure B-19: ^{1}H NMR (400 MHz, DMSO-d_6) spectrum for 7-chloro-indolmycenic acid (**7Cl-11**).



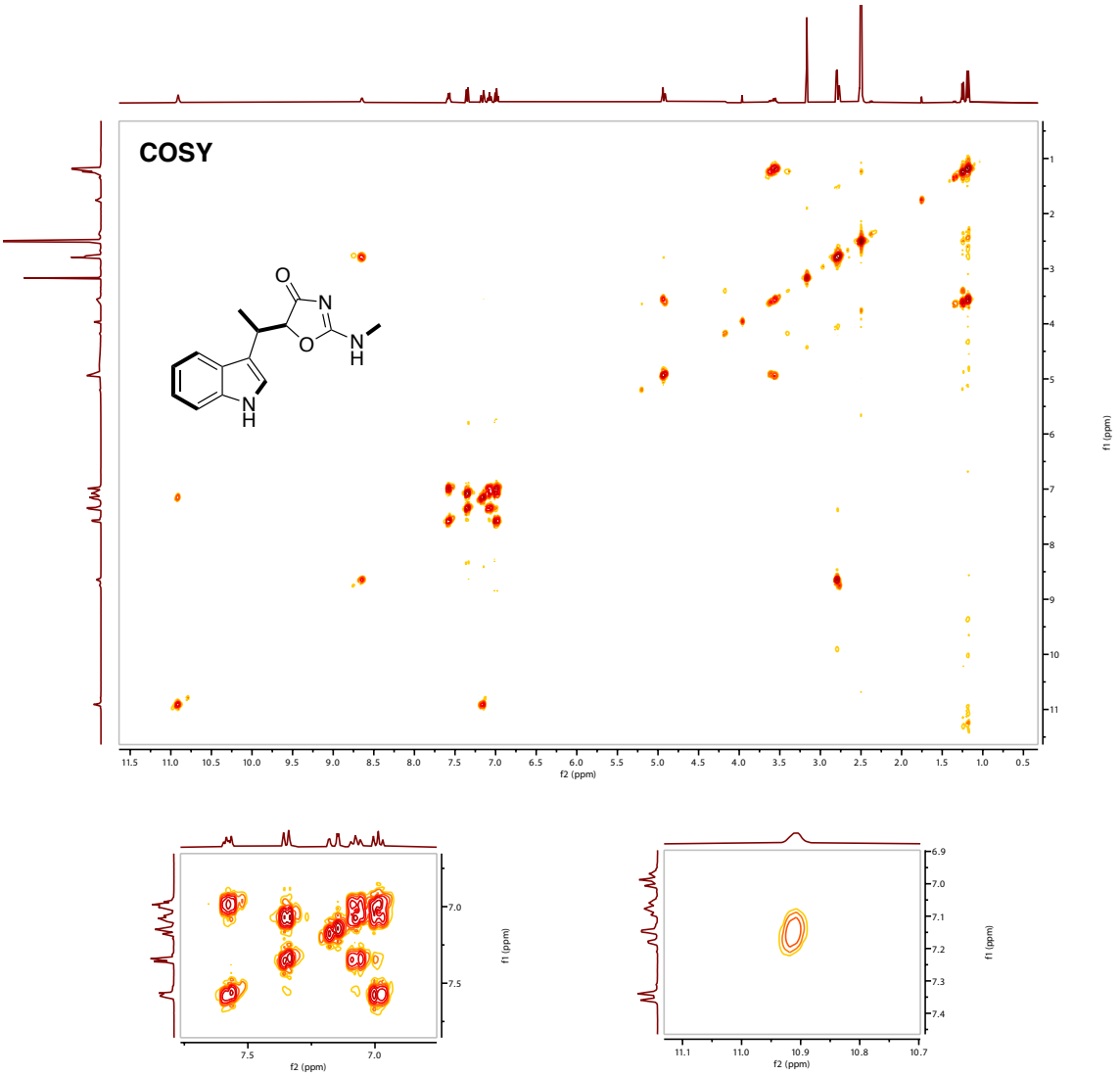
Appendix Figure B-20: ^{13}C NMR (101 MHz, DMSO-d_6) spectrum for **7Cl-11.**



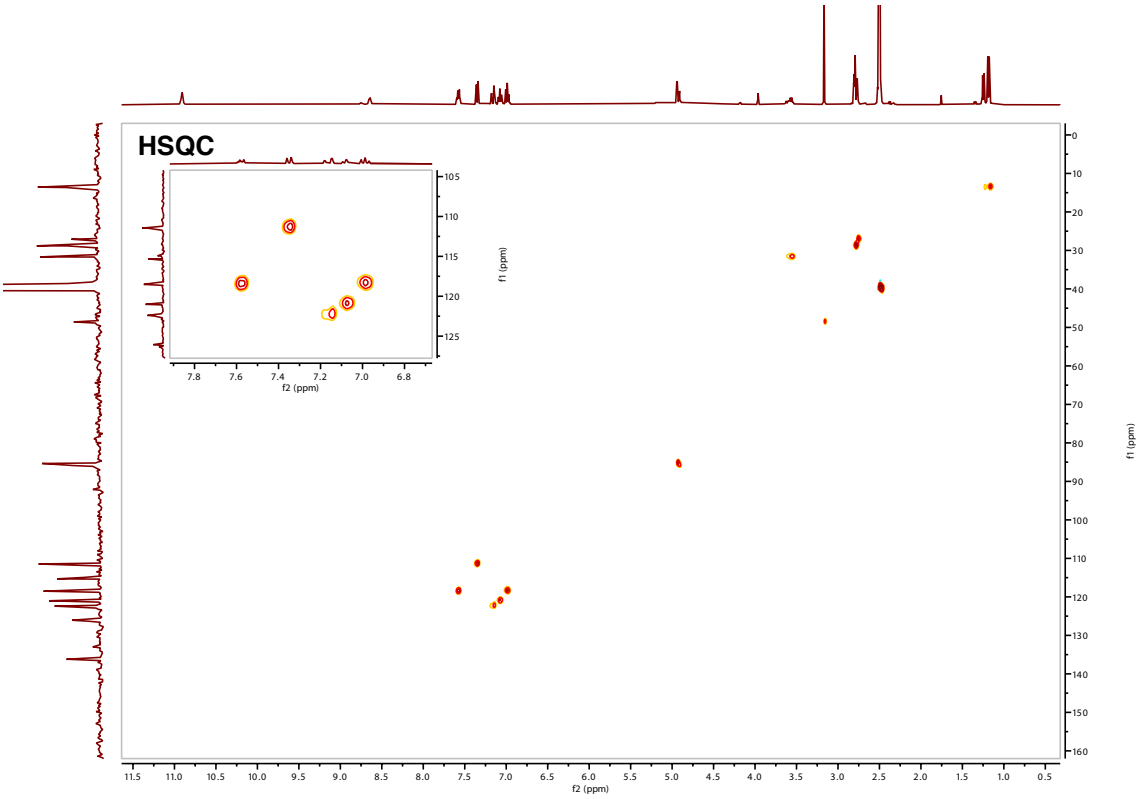
Appendix Figure B-21: ^{1}H NMR (400 MHz, DMSO-d_6) spectrum for indolmycin (**6**).



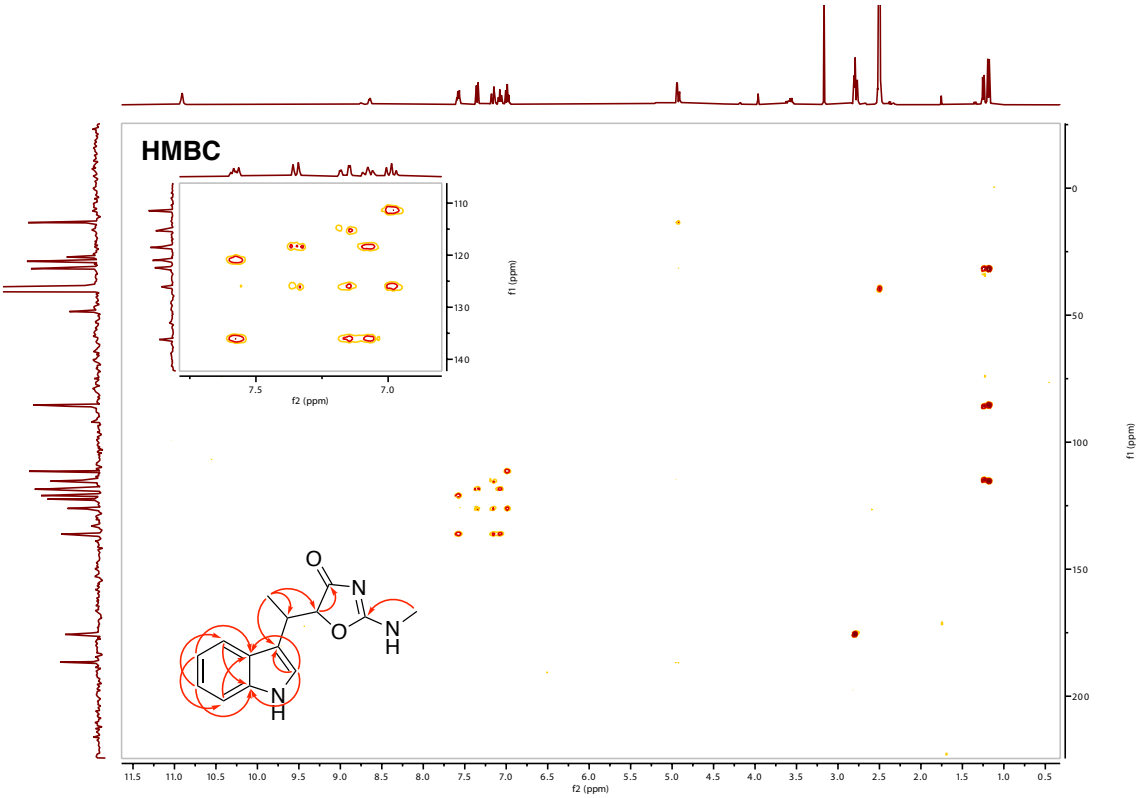
Appendix Figure B-22: ^{13}C NMR (101 MHz, DMSO-d_6) spectrum for **6**.



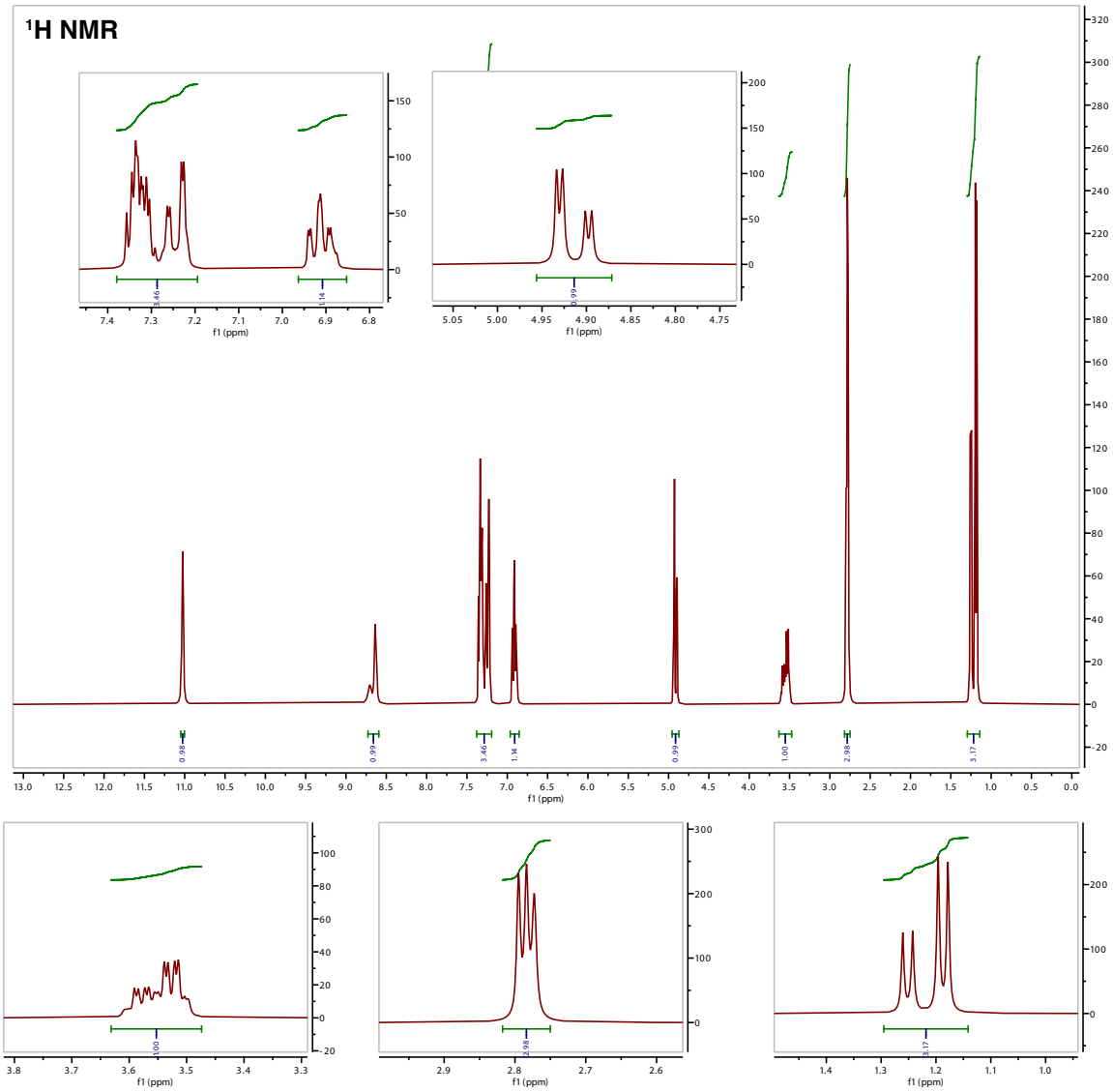
Appendix Figure B-23: COSY NMR (400 MHz, DMSO-d_6) spectrum for **6**.



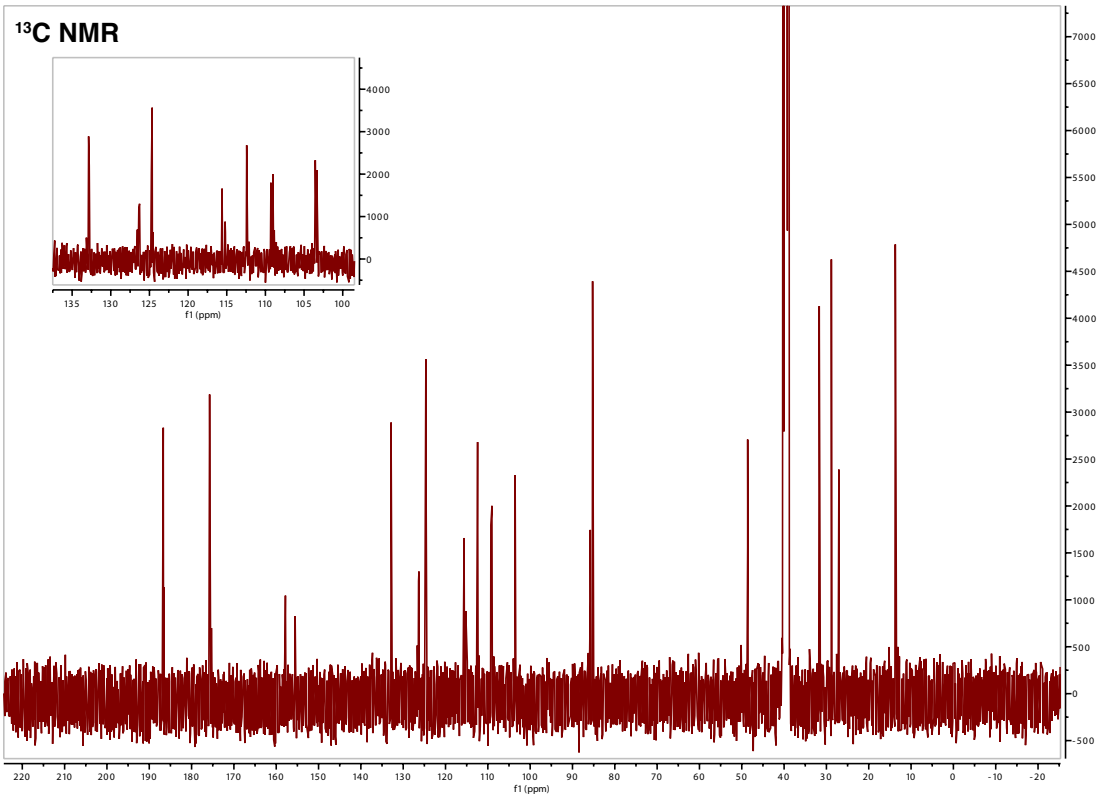
Appendix Figure B-24: HSQC NMR (400 MHz for ^{1}H NMR, 101 MHz for ^{13}C NMR, DMSO-d_6) spectrum for **6**.



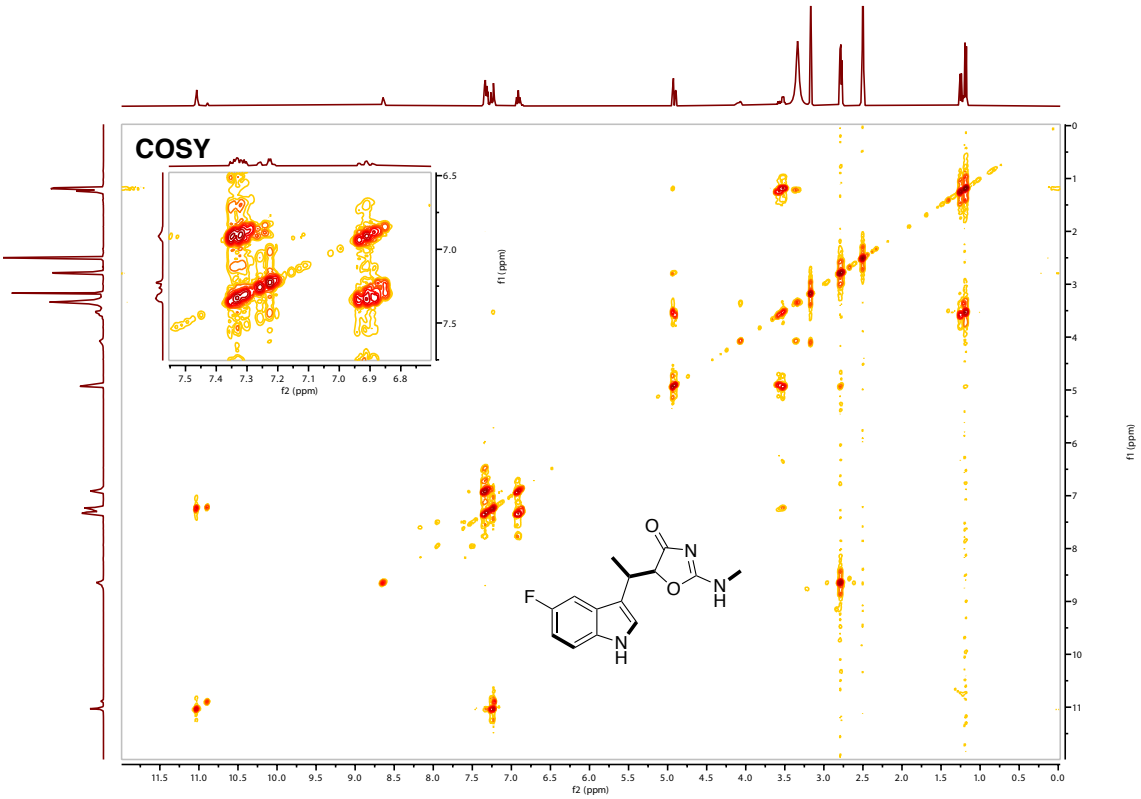
Appendix Figure B-25: HMBC NMR (400 MHz for ^{1}H NMR, 101 MHz for ^{13}C NMR, DMSO-d_6) spectrum for **6**.



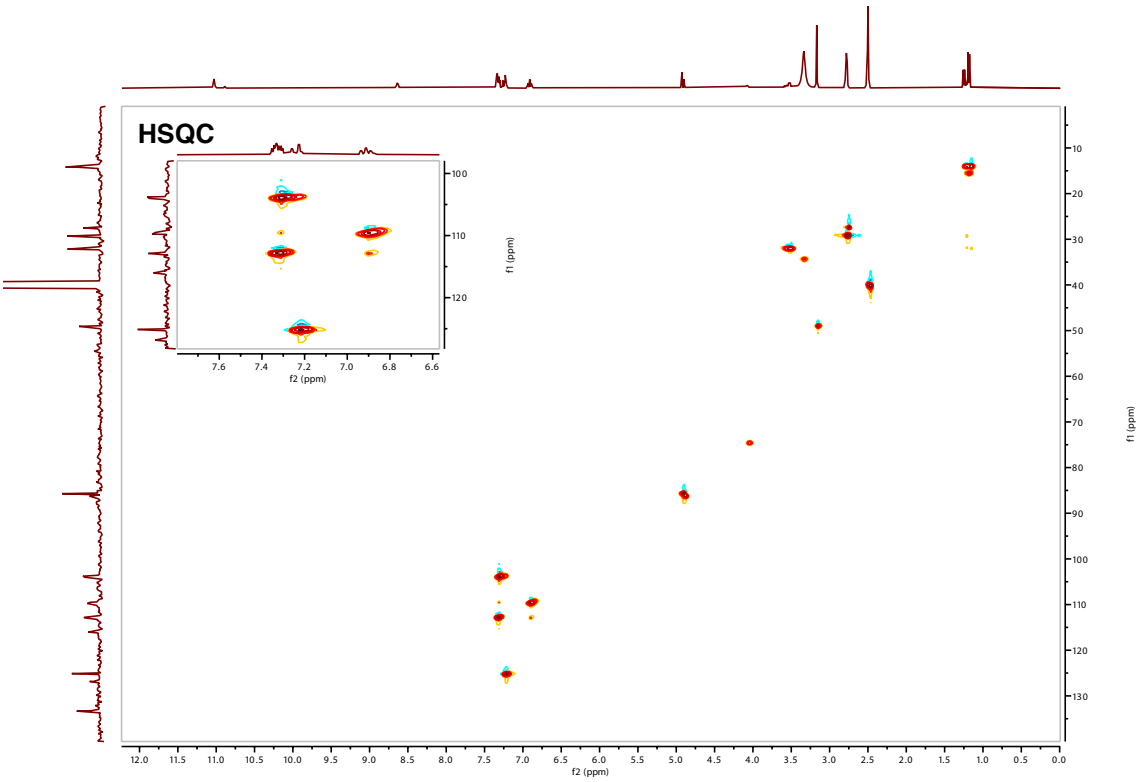
Appendix Figure B-26: ^{1}H NMR (400 MHz, DMSO-d_6) spectrum for 5-fluoro-indolmycin (**5F-6**).



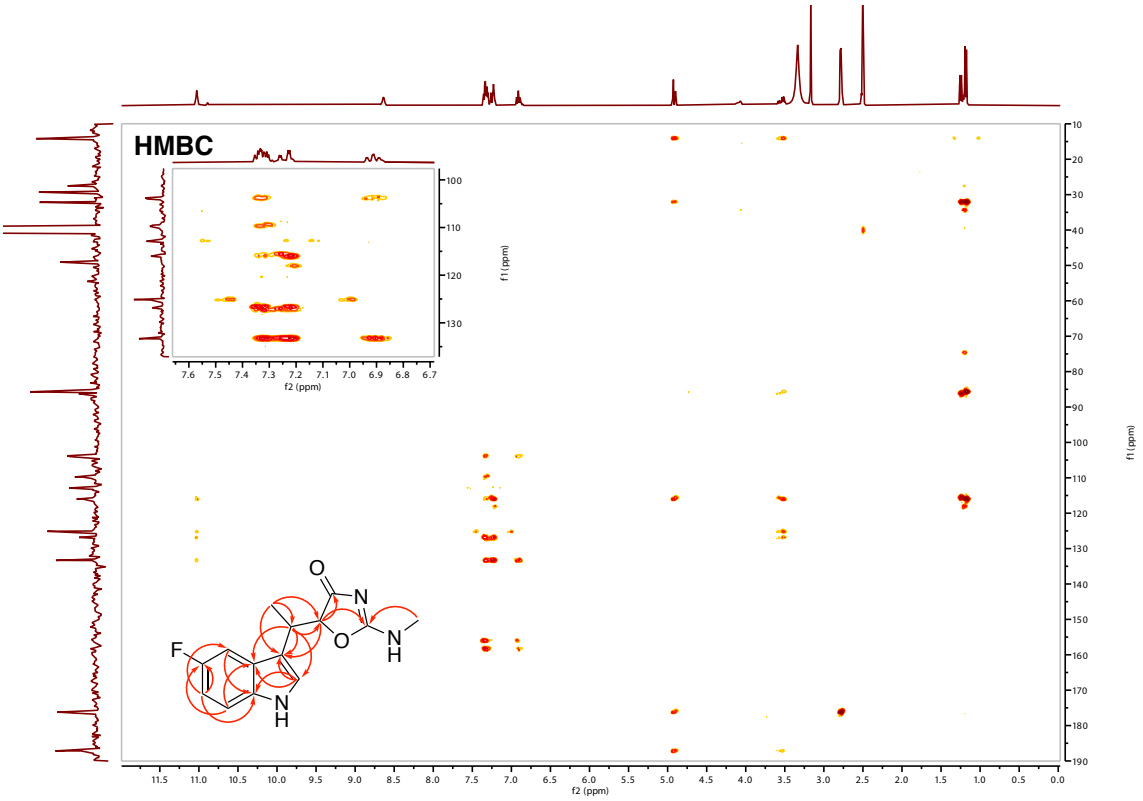
Appendix Figure B-27: ^{13}C NMR (101 MHz, DMSO-d_6) spectrum for **5F-6**.



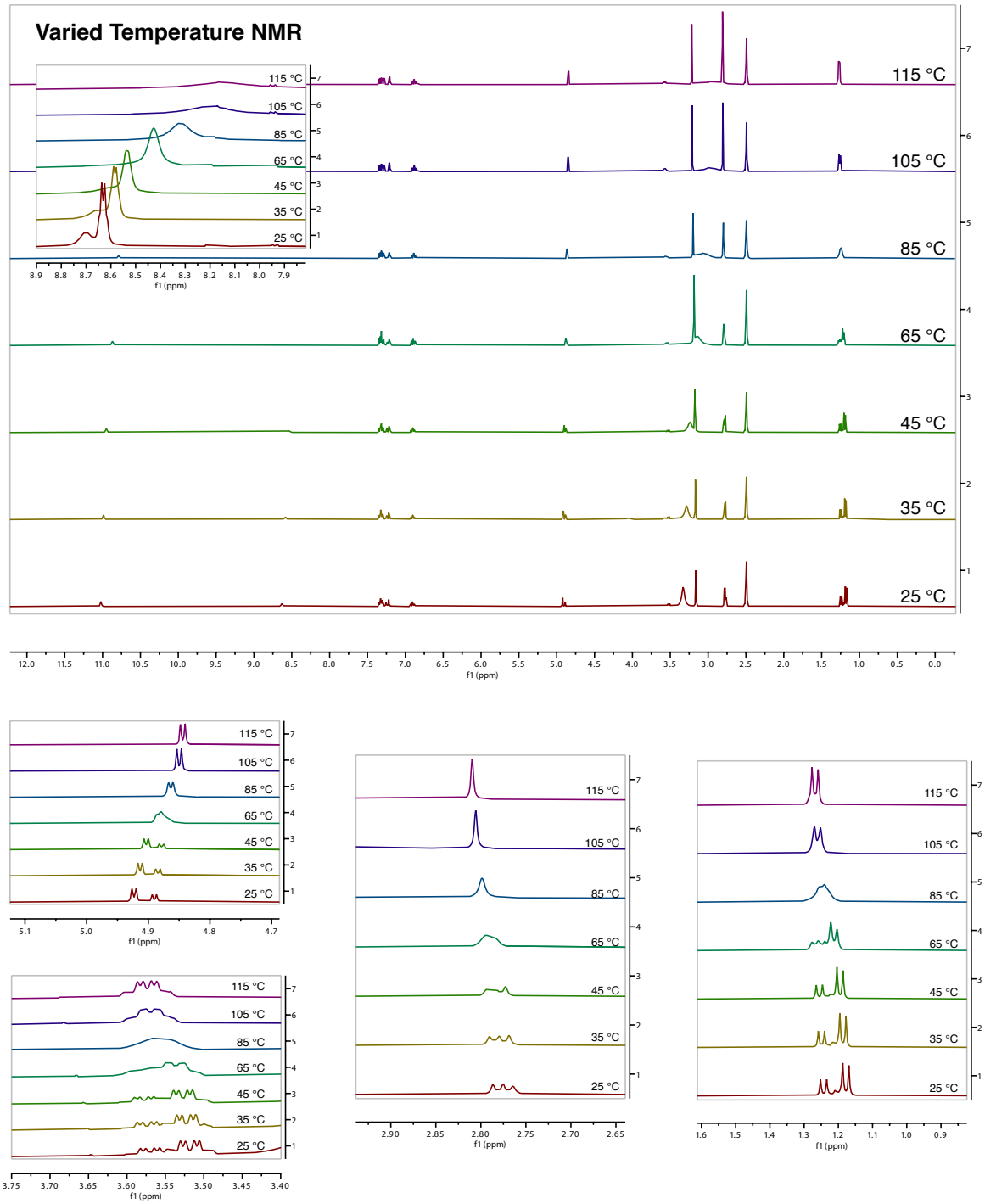
Appendix Figure B-28: COSY NMR (400 MHz for ^{1}H NMR, DMSO-d_6) spectrum for **5F-6**.



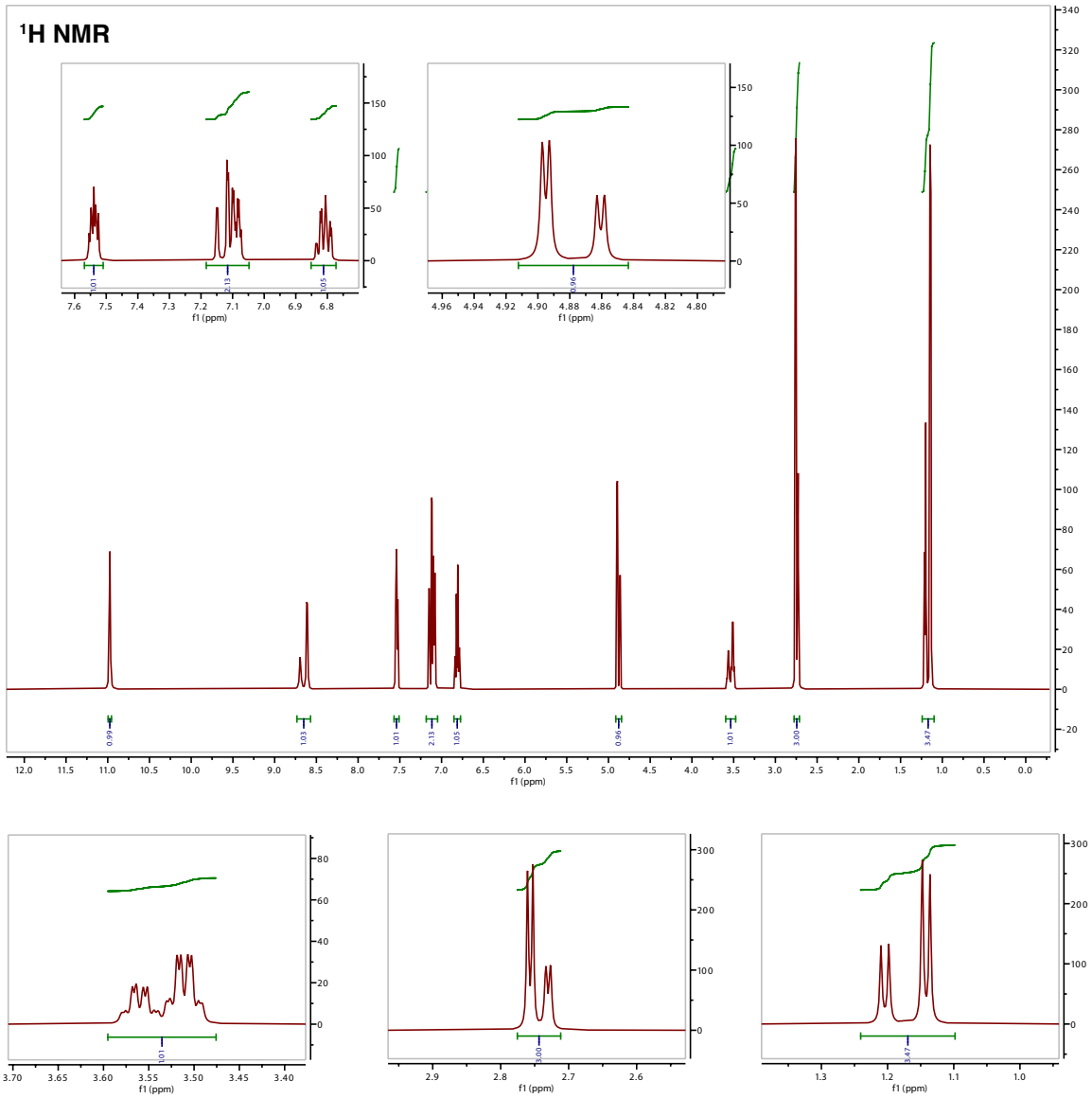
Appendix Figure B-29: HSQC NMR (400 MHz for ^{1}H NMR, 101 MHz for ^{13}C NMR, DMSO-d_6) spectrum for **5F-6**.



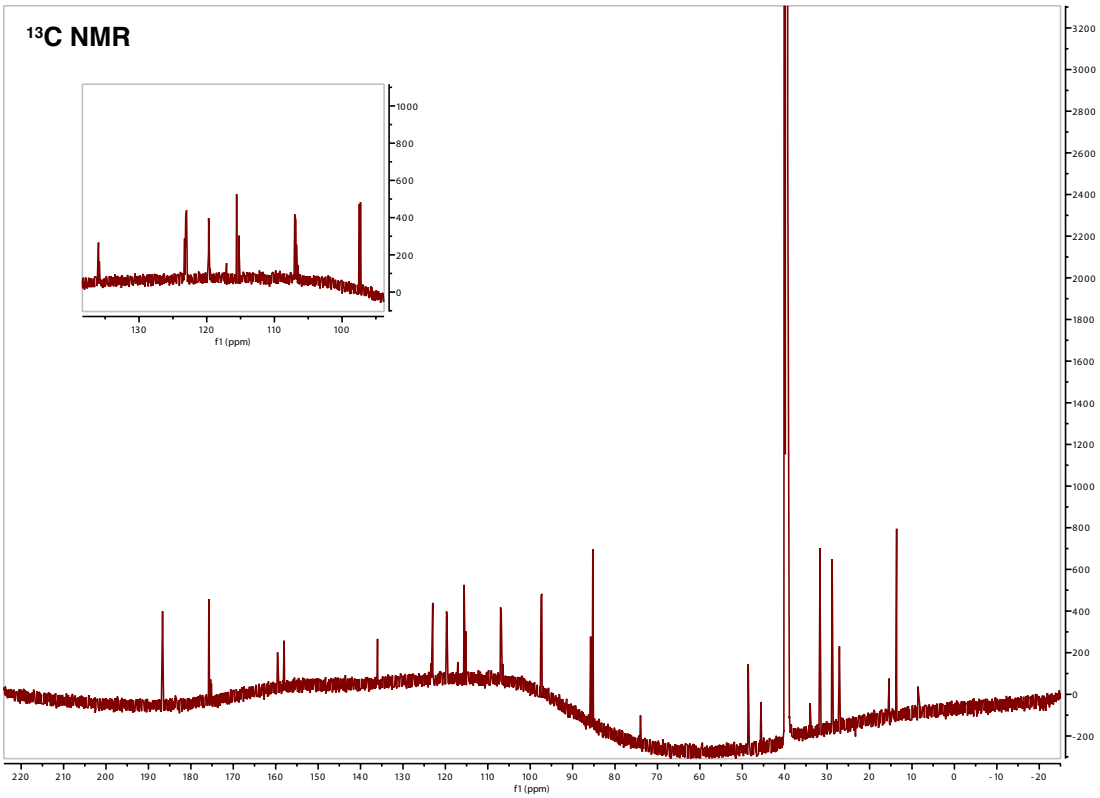
Appendix Figure B-30: HMBC NMR (400 MHz for ^{1}H NMR, 101 MHz for ^{13}C NMR, DMSO-d_6) spectrum for **5F-6**.



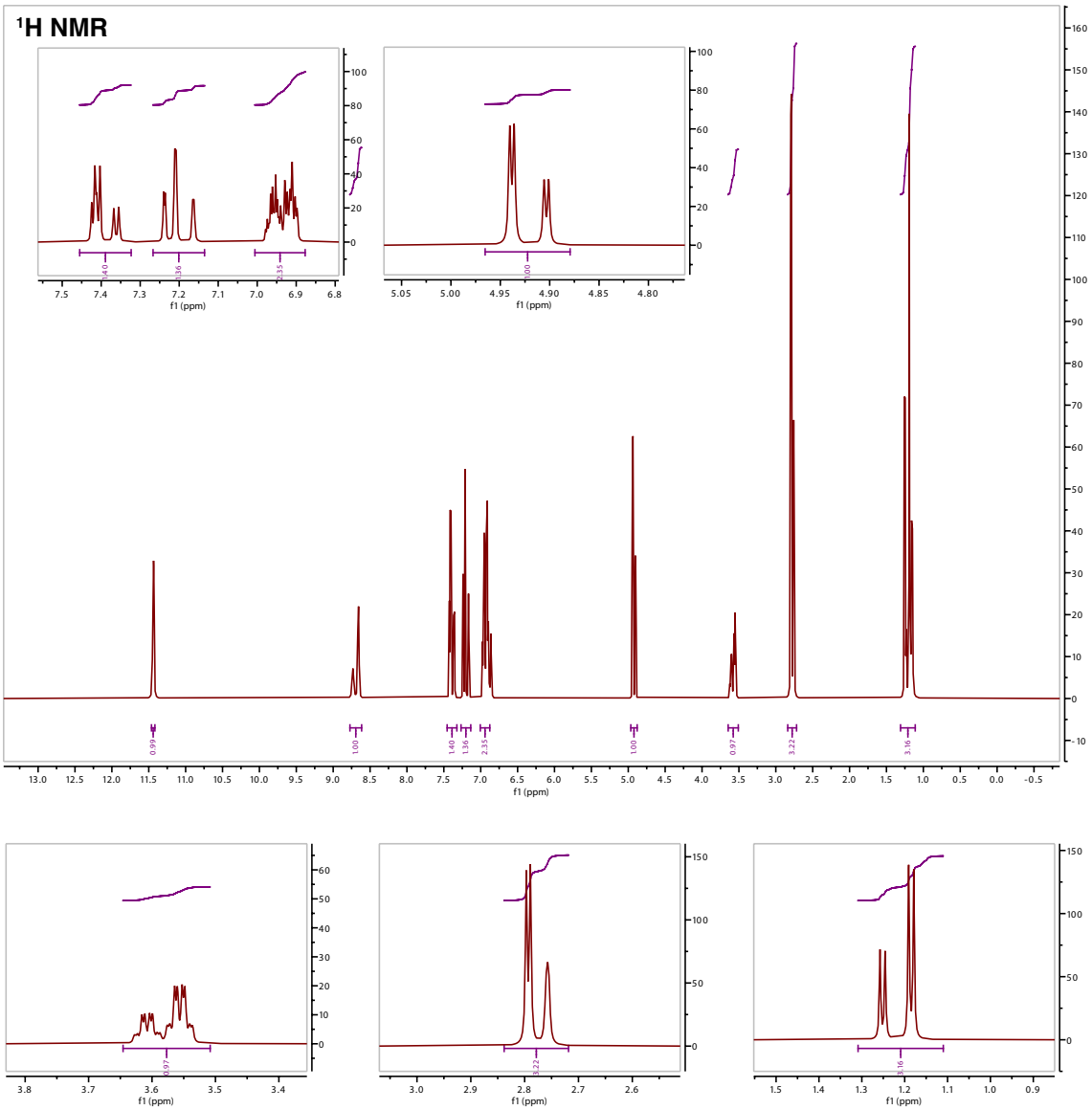
Appendix Figure B-31: Varied temperature NMR (400 MHz, DMSO-d_6) spectra for **5F-6**.



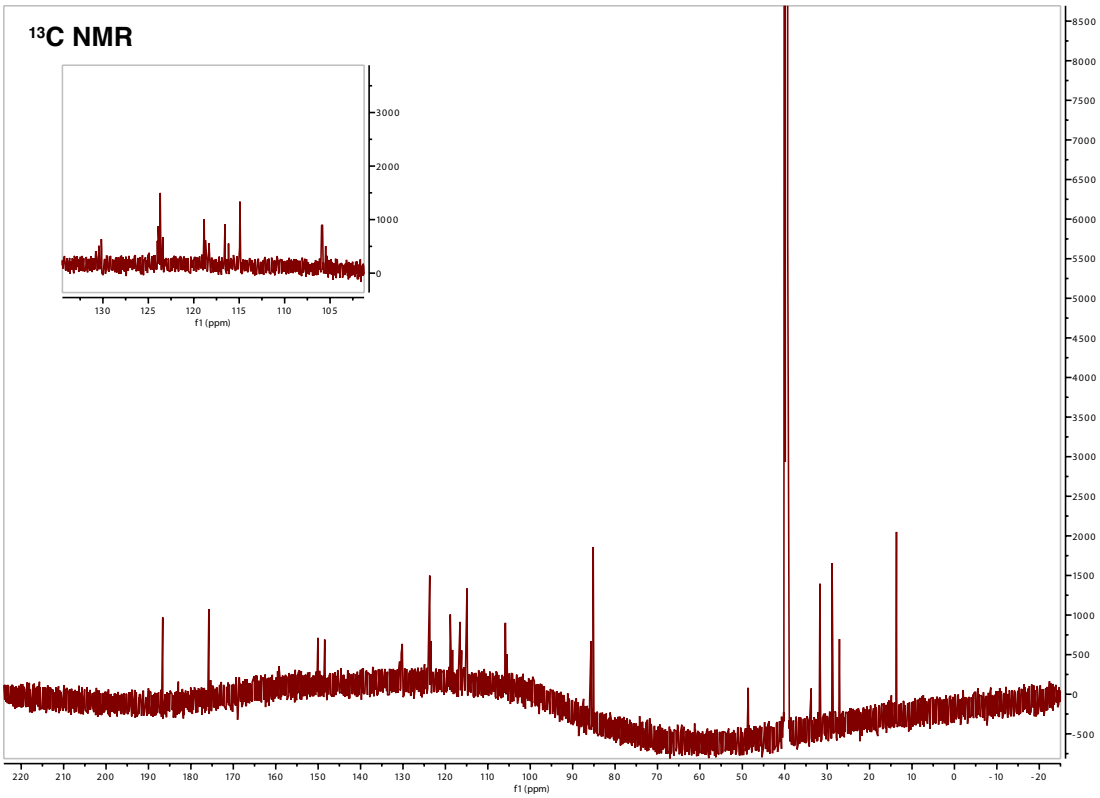
Appendix Figure B-32: ^{1}H NMR (600 MHz, DMSO-d_6) spectrum for 6-fluoro-indolmycin (**6F-6**).



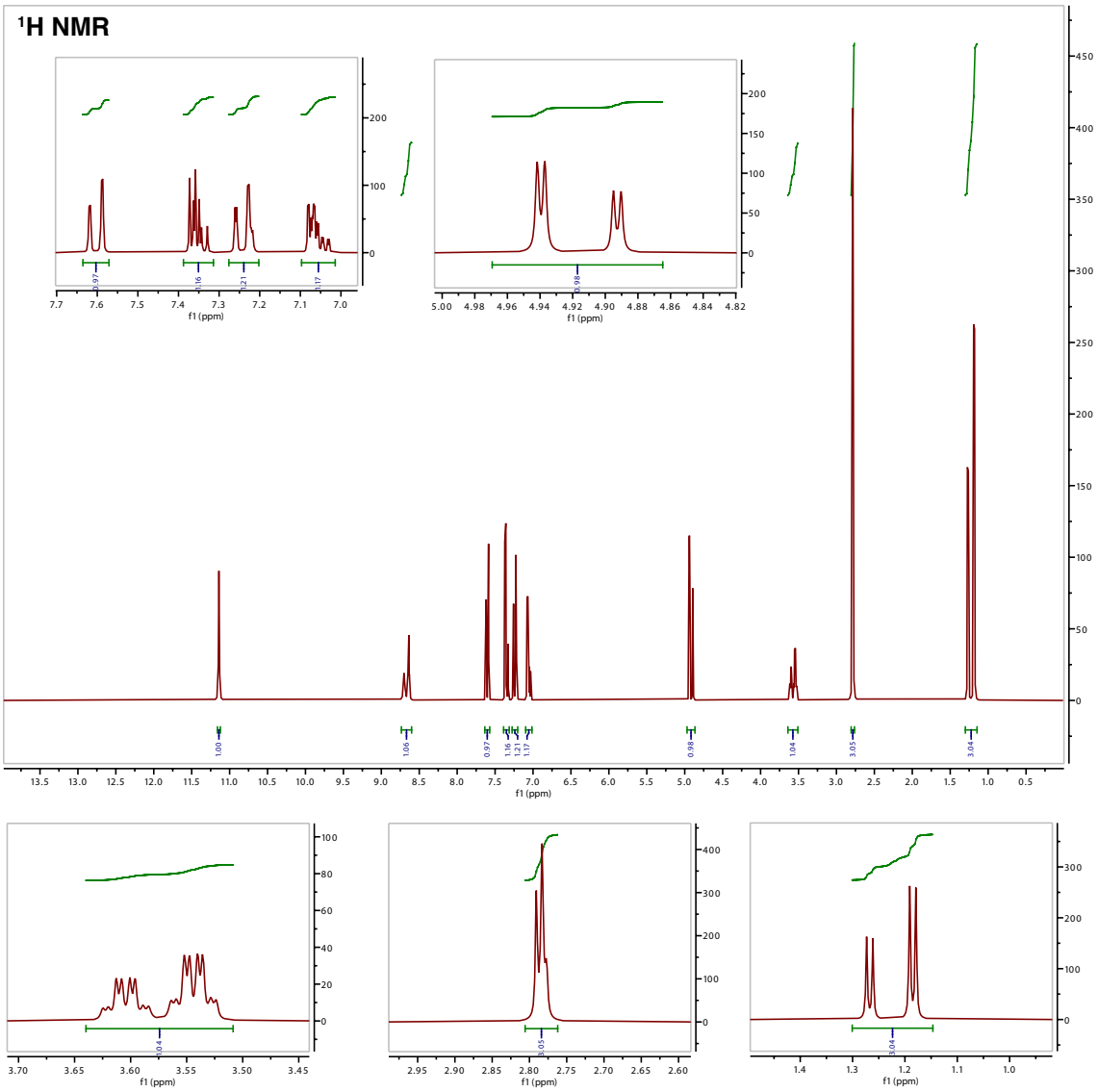
Appendix Figure B-33: ^{13}C NMR (151 MHz, DMSO-d_6) spectrum for **6F-6**.



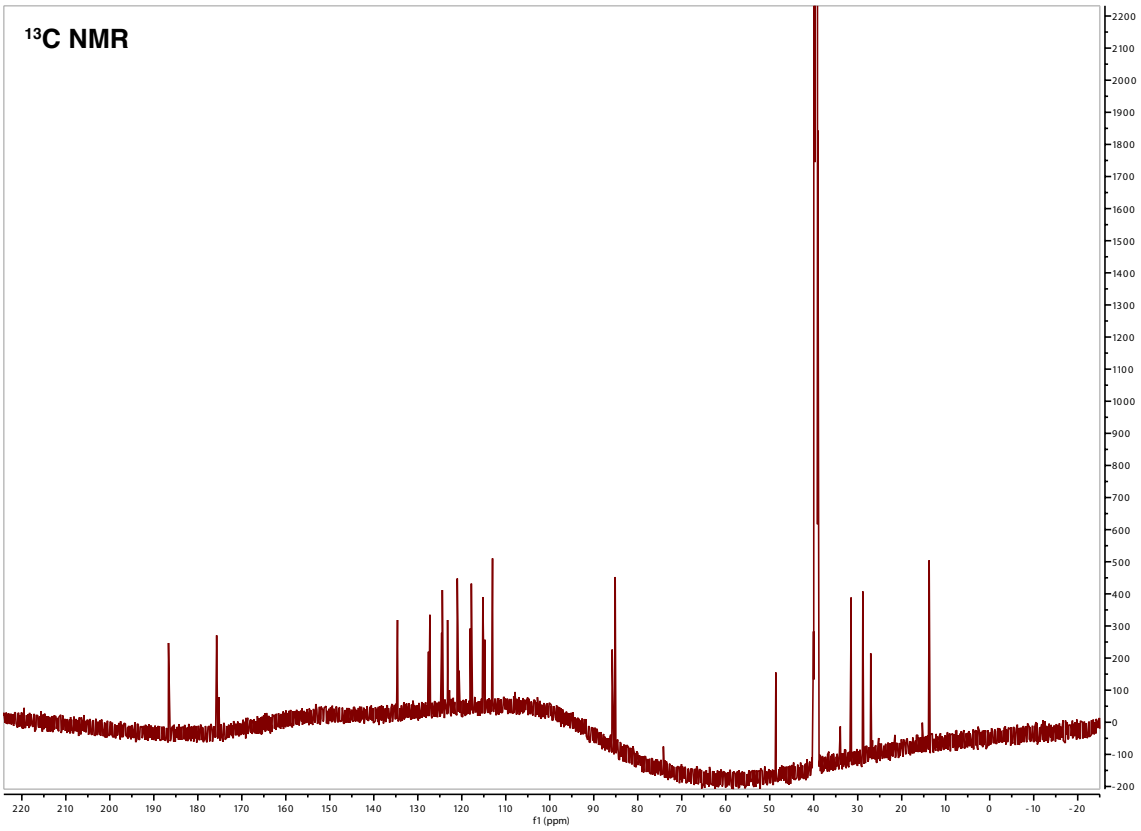
Appendix Figure B-34: ^{1}H NMR (600 MHz, DMSO-d_6) spectrum for 7-fluoro-indolmycin (**7F-6**).



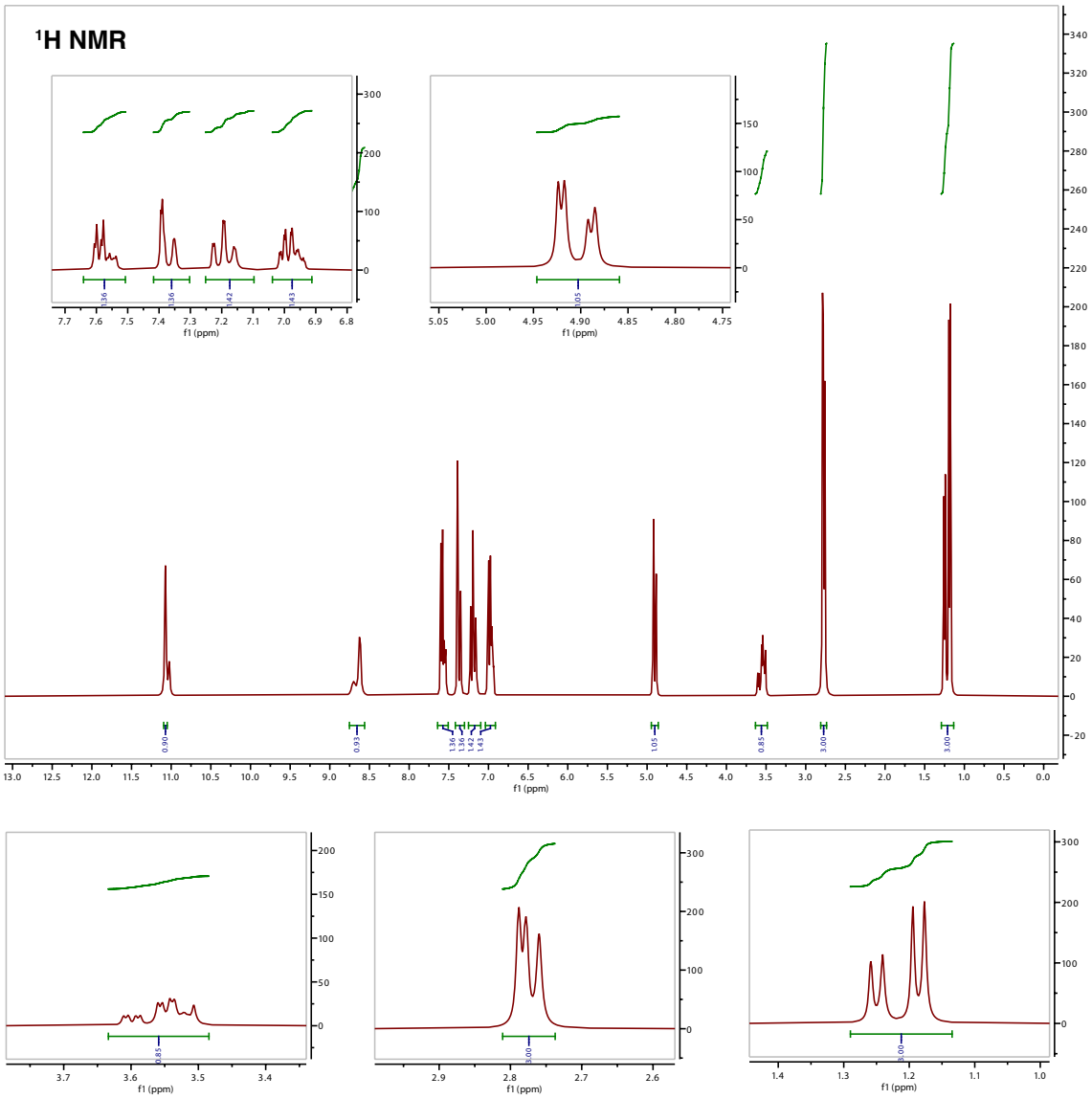
Appendix Figure B-35: ^{13}C NMR (151 MHz, DMSO-d_6) spectrum for **7F-6**.



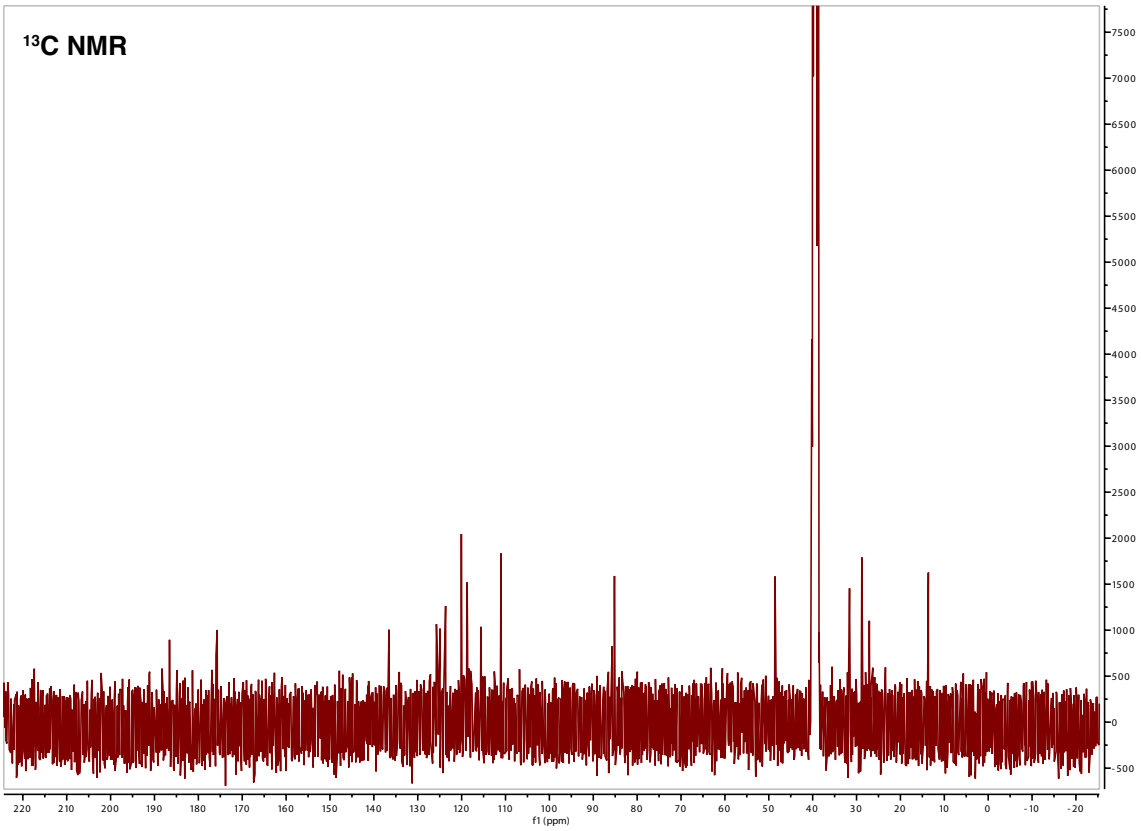
Appendix Figure B-36: ^{1}H NMR (600 MHz, DMSO-d_6) spectrum for 5-chloro-indolmycin (**5Cl-6**).



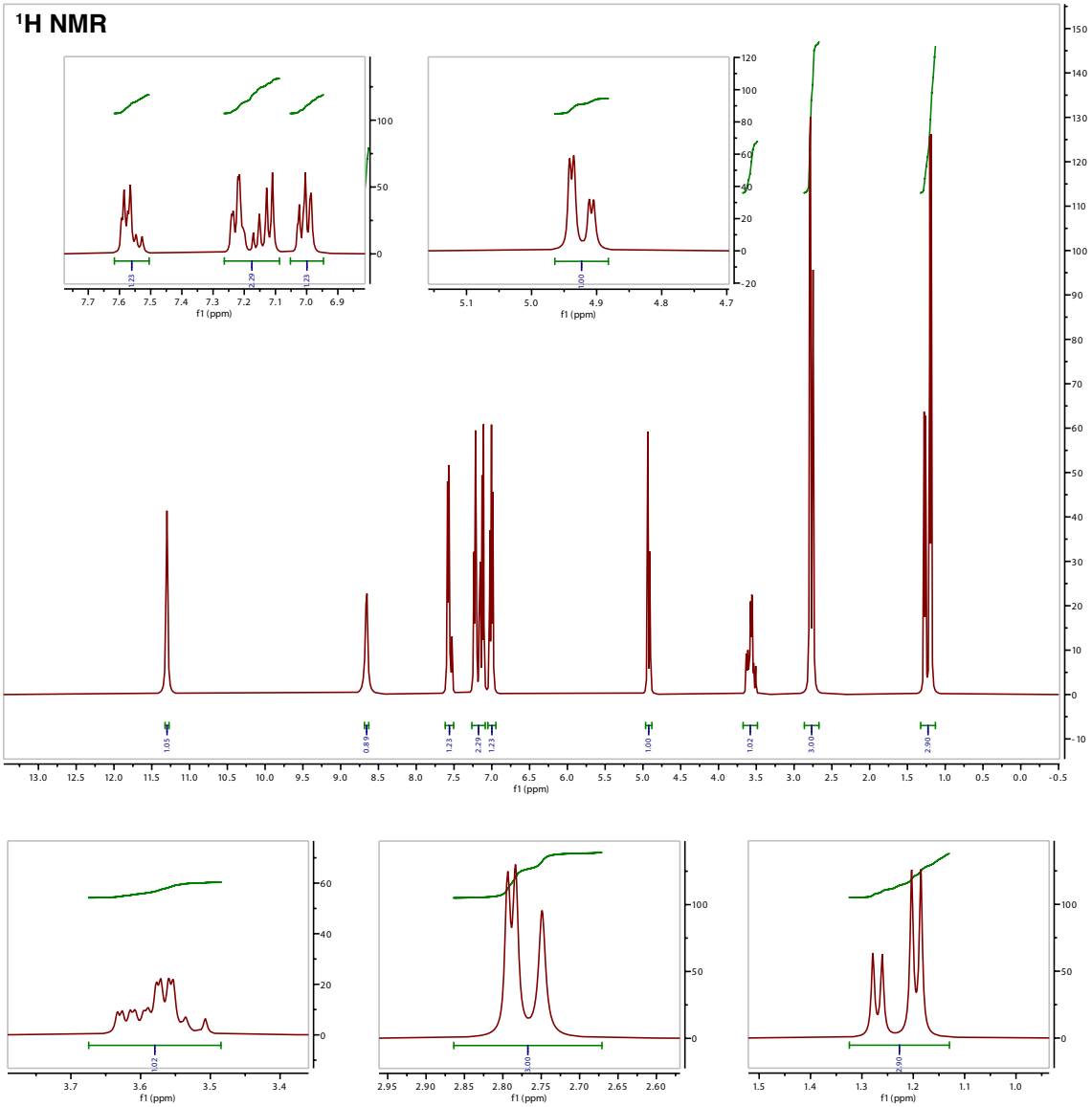
Appendix Figure B-37: ^{13}C NMR (151 MHz, DMSO-d_6) spectrum for **5Cl-6**.



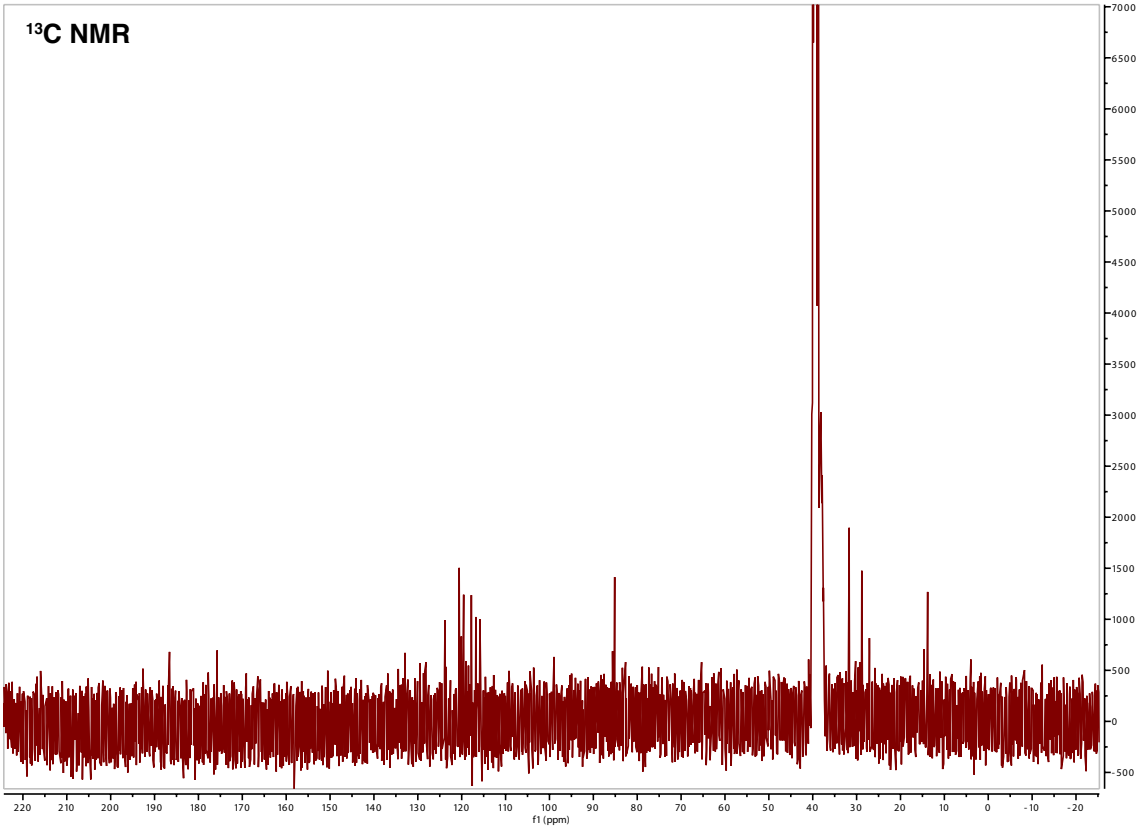
Appendix Figure B-38: ^{1}H NMR (400 MHz, DMSO-d_6) spectrum for 6-chloro-indolmycin (**6Cl-6**).



Appendix Figure B-39: ^{13}C NMR (101 MHz, DMSO-d_6) spectrum for **6Cl-6**.



Appendix Figure B-40: ^{1}H NMR (400 MHz, DMSO-d_6) spectrum for 7-chloro-indolmycin (**7Cl-6**).



Appendix Figure B-41: ^{13}C NMR (101 MHz, DMSO-d_6) spectrum for **7Cl-6**.