

Stereochemistry in Biology and Synthetic Biology

BY

TAYLOR KORNFUEHRER

B.S., The University of Texas at Austin, 2017

THESIS

Submitted as partial fulfillment of the requirements

for the degree of Master of Science in Pharmacognosy

in the Graduate College of the University of Illinois at Chicago, 2020

Chicago, Illinois

Defense Committee:

Alessandra S. Eustáquio, Chair and Advisor

Joanna E. Burdette

Nora Laslop-Vasquez

This thesis is dedicated to my mother, Sandi.

CONTRIBUTION OF AUTHORS

Chapter 1 is my own work.

Chapter 2 is adapted from an accepted article (Kornfuehrer, T. *et al.*, ChemBioChem, 2020, DOI: 10.1002/cbic.202000349) of which Taylor Kornfuehrer is the primary author and primary driver of the research. The article is currently in press at ChemBioChem. My co-authors contributed to this work in the following ways: Sean Romanowski carried out the conclusive assays with the (*S,S*) diastereomer presented in Table III, Figure 2.10, Figure 2.11. Valérie de Crécy-Lagard carried out the bioinformatic analyses presented in section 2.2.5 and contributed to writing this portion of the manuscript. Andrew Hanson was instrumental in calculations regarding physiological relevance presented in section 2.4.6. and contributed to writing this portion of the manuscript. Alessandra Eustáquio was my primary mentor for this project, provided helpful discussion, and contributed to writing the manuscript.

Chapter 3 is adapted from a published manuscript (Kornfuehrer, T., and Eustáquio, A.S., *Med. Chem. Commun.*, 2019, 10:1256-1272, DOI: 0.1039/C9MD00141G) of which Taylor Kornfuehrer is the primary author.

Chapter 4 is primarily my own unpublished work. It is part of a collaborative project with the Burkhardt lab at UC San Diego and will be continued by other lab members. Talia Cheifetz, a Summer Undergraduate Research Fellow under my guidance, carried out construction of pTC01 and the first round of protein purification described in section 4.2.2.2. All protocols for her work were devised by me.

TABLE OF CONTENTS

<u>CHAPTER</u>	<u>PAGE</u>
1. INTRODUCTION	1
1.1 Stereochemistry and Principles of Molecular Asymmetry.....	2
1.2 The Importance of Molecular Asymmetry in Biological Systems.....	4
1.3 References Cited	6
2. MAINTENANCE OF COFACTOR HOMOCHIRALITY IN BACTERIAL CELLS: ASSIGNMENT OF FUNCTION TO A STEREOSELECTIVE HYDROLASE.....	8
2.1 Introduction.....	9
2.2 Results and Discussion.....	12
2.2.1 Analysis of diastereomer composition in commercial standards of <i>S</i> -adenosyl-L-methionine.....	12
2.2.2 Cloning and purification of recombinant StDUF62.....	13
2.2.3 Preliminary assay with mixed diastereomers.....	15
2.2.4 Determination of kinetic constants for hydrolysis of isolated diastereomers.....	16
2.2.5 Bioinformatic analysis.....	25
2.3 Conclusions.....	28
2.4 Experimental Methods	29
2.4.1 Chemicals.....	29
2.4.2 High pressure liquid chromatography.....	29
2.4.2.1 <i>S</i> -adenosyl-L-methionine diastereomer analysis, isolation, and buffer exchange	29
2.4.2.2 Measuring adenosine formation.....	34
2.4.3 Strains, cultivation conditions, and genomic DNA isolation	34
2.4.4 Cloning and purification of recombinant StDUF62	35
2.4.5 <i>in vitro</i> assays	36
2.4.6 Calculation favoring physiological relevance of observed kinetic constants.....	36
2.4.7 Bioinformatic analyses.....	37
2.5 References Cited.....	38
3. MECHANISMS OF POLYKETIDE BIOSYNTHESIS AND STRUCTURE DIVERSIFICATION VIA SYNTHASE ENGINEERING..	43
3.1 Introduction.....	44
3.2 Rationale for Polyketide Synthetic Biology.....	46
3.2.1 Polyketide synthase architecture	47
3.2.2 Mechanisms of polyketide biosynthesis.....	49
3.2.2.1 Loading.....	49
3.2.2.2 Extension.....	51
3.2.2.3 Offloading	53
3.2.3 Polyketide structure diversification via synthase engineering.....	54
3.3 Monomer Diversification	55
3.3.1 Acyltransferase engineering.....	55
3.3.1.1 Domain exchange.....	56
3.3.1.2 Site-directed mutagenesis.....	57
3.3.1.3 Inactivation and cross-complementation.....	58
3.3.2 Precursor-directed biosynthesis and monomer biosynthesis.....	59
3.4 Control of Oxidation State and Stereochemistry.....	60
3.4.1 Ketoreductase engineering.....	62

TABLE OF CONTENTS (continued)

<u>CHAPTER</u>		<u>PAGE</u>
	3.4.1.1 Site-directed mutagenesis.....	62
	3.4.1.2 Domain exchange.....	63
	3.4.2 Dehydratase engineering.....	65
	3.4.2.1 Site-directed mutagenesis.....	65
	3.4.2.2 Domain exchange.....	66
	3.4.3 Enoylreductase engineering.....	67
	3.4.3.1 Site-directed mutagenesis.....	67
	3.4.3.2 Domain exchange.....	67
3.5	Engineering Cyclized or Linear Products..	69
3.6	The Present and Future of Polyketide Structure Diversification.....	71
	3.6.1 The continued role of natural products research.....	71
	3.6.2 The emergent landscape of polyketide synthase engineering.....	72
	3.6.2.1 A developing synthetic biology toolkit.....	72
	3.6.2.2 Synthase structure and function.....	74
	3.6.2.3 Evolution of polyketide synthases.....	75
	3.6.3 Bridging natural products and medicinal chemistry research.....	77
3.7	Conclusions.....	78
3.8	References Cited.....	79
4.	SYNTHETIC BIOLOGY EFFORTS TO GENERATE STEREOCHEMICAL ANALOGUES OF POLYKETIDE NATURAL PRODUCTS.....	87
4.1	Introduction.....	88
	4.1.1 Targeting the spliceosome for treatment of cancer.....	88
	4.1.2 Using synthetic biology to produce non-natural polyketide analogues with promising therapeutic properties.....	89
4.2	Results and Discussion.....	93
	4.2.1 Propagation of <i>Streptomyces</i> sp. FERM BP-7812 and analysis of phenotypically distinct isolates.....	93
	4.2.1.1 Analysis of pladienolide B production in isolates of <i>Streptomyces</i> sp. FERM BP-7812.....	97
	4.2.1.2 Laboratory domestication of <i>Streptomyces</i> sp. FERM BP-7812.....	102
	4.2.2 Efforts to use synthetic biology for production of 17S-FD-895.....	105
	4.2.2.1 <i>in silico</i> analysis of donor reductive loops.....	105
	4.2.2.2 Cloning and purification of recombinant PldMod3... ..	113
4.3	Conclusions.....	117
4.4	Future Directions.....	118
4.5	Experimental Methods.....	120
	4.5.1 Strain propagation and isolation of distinct phenotypes	120
	4.5.2 Routine strain cultivation.. ..	121
	4.5.3 Genomic DNA isolation and 16S analysis	122
	4.5.4 IDBac	122
	4.5.5 Pladienolide B production analysis.....	122
	4.5.6 Sequence analysis of B2-type ketoreductase donor candidates	125
	4.5.7 Construction of pTC01.....	125
	4.5.8 Purification of recombinant PldMod3	126
4.6	References Cited	127

TABLE OF CONTENTS (continued)

<u>CHAPTER</u>	<u>PAGE</u>
APPENDICES.....	131
Appendix A: Herboxidiene	131
Appendix B: <i>Micromonospora</i> sp. B006.....	143
Appendix C: Wiley Permission for Reproduction of Chapter 2	155
Appendix D: RSC Permission for Reproduction of Chapter 3.	157

LIST OF TABLES

<u>TABLE</u>	<u>PAGE</u>
I	KINETICS DATA FOR (<i>R,S</i>)- <i>S</i> -ADENOSYL-L-METHIONINE.....17-18
II	ADENOSINE FORMATION DURING ONE HOUR INCUBATION OF DUF62 WITH (<i>S,S</i>)- <i>S</i> -ADENOSYL-L-METHIONINE AT 30°C20
III	ADENOSINE FORMATION DURING ONE HOUR INCUBATION OF DUF62 WITH (<i>S,S</i>)- <i>S</i> -ADENOSYL-L-METHIONINE AT 37°C22
IV	APPARENT KINETIC CONSTANTS OF STDUF6224
V	<i>S</i> -ADENOSYL-L-METHIONINE DIASTEREOMER CALIBRATION CURVE.....32
VI	CONCENTRATION OF POOLED (<i>R,S</i>)- <i>S</i> -ADENOSYL-L-METHIONINE33
VII	CONCENTRATION OF POOLED (<i>S,S</i>)- <i>S</i> -ADENOSYL-L-METHIONINE33
VIII	COMPARISON OF PLADIENOLIDE B TITERS BETWEEN PHENOTYPE GROUPS99
IX	COMPARISON OF PLADIENOLIDE B TITERS BETWEEN BLACK ISOLATES.....100
X	B2-TYPE KETOREDUCTASE DONOR CANDIDATES.....106
XI	ASSEMBLY LINE CONTEXT OF DONOR B2-TYPE KR DOMAINS.....110
XII	SOLVENT SCHEMES USED FOR PLADIENOLIDE B ANALYSES.....123
XIII	PLADIENOLIDE B STANDARD INJECTIONS124
XIV	REACTION SETUP AND THERMOCYCLER PROGRAMMING FOR AMPLIFICATION OF <i>PLDMOD3</i> FROM GENOMIC DNA126

LIST OF FIGURES

<u>FIGURE</u>		<u>PAGE</u>
	<u>Chapter 1</u>	
1.1	Enantiomers of bromochlorofluoromethane.....	2
1.2	Stereochemical relationships of the tetrose saccharides.....	3
	<u>Chapter 2</u>	
2.1	The biological cofactor <i>S</i> -adenosyl-L-methionine.....	9
2.2	<i>S</i> -adenosyl-L-methionine biosynthesis, racemization, and hypothesis of DUF62 function	11
2.3	Diastereomeric composition of commercial <i>S</i> -adenosyl-L-methionine preparations	12
2.4	Cloning <i>duf62</i> into expression vector pHis8	13
2.5	Purification of recombinant StDUF62	14
2.6	StDUF62 assay with a <i>S</i> -adenosyl-L-methionine racemate indicates preference for (<i>R,S</i>) diastereomer.....	15
2.7	Identification of outlier data point via box plot analysis.....	16
2.8	Kinetic constants for (<i>R,S</i>)- <i>S</i> -adenosyl-L-methionine determined via linear regression.....	19
2.9	Testing the activity of StDUF62 with (<i>S,S</i>)- <i>S</i> -adenosyl-L-methionine at 30°C.....	20
2.10	Extended incubation of (<i>S,S</i>)- <i>S</i> -adenosyl-L-methionine at 30 °C	21
2.11	Testing the activity of StDUF62 with (<i>S,S</i>)- <i>S</i> -adenosyl-L-methionine at 30°C.....	23
2.12	Phylogenetic distribution of DUF62 in bacteria and archaea.....	26
2.13	Predicted co-regulation of DUF62 encoding genes and energy-coupled factor transporter MtsABC encoding genes	27
2.14	<i>S</i> -adenosyl-L-methionine diastereomer isolation workflow.....	30
2.15	<i>S</i> -adenosyl-L-methionine diastereomer calibration curve.....	32
2.16	Adenosine calibration curve.....	34
	<u>Chapter 3</u>	
3.1	Chemical structures of clinically-relevant polyketides.	45
3.2	Organization of the 6-deoxyerythronolide B synthase.....	48
3.3	Domain architectures found in loading modules of polyketide synthases	50
3.4	Mechanisms of polyketide chain extension in a prototypical module N.	52
3.5	Mechanisms of polyketide offloading.....	53
3.6	Examples of polyketide analogues obtained via monomer diversification and the methods employed	56

LIST OF FIGURES (continued)

<u>FIGURE</u>		<u>PAGE</u>
	<u>Chapter 3 (cont)</u>	
3.7	Control of oxidation state and stereochemistry effected by domains of the reductive loop.....	61
3.8	Alteration of oxidation state in the amphotericin B backbone via ketoreductase engineering.....	63
3.9	Alteration of oxidation state in the geldanamycin backbone via dehydratase engineering and disruption of post-assembly line tailoring.....	66
3.10	Structures of avermectins B1, ivermectins B1, and milbemycins A	68
3.11	Effecting macrocyclization of the tautomycetin via thioesterase domain exchange	70
3.12	Evolutionary studies of large aminopolyol assembly lines leads to proposed redefinition of modular boundaries in polyketide synthases.....	76
	<u>Chapter 4</u>	
4.1	The three structural classes of polyketide natural product spliceosome inhibitors	89
4.2	Notable synthetic and natural analogues from the pladienolide structural class.....	90
4.3	Proposed reductive loop swap of the pladienolide polyketide synthase.....	92
4.4	Phenotypic heterogeneity of <i>Streptomyces</i> sp. FERM BP-7812 solid media	93
4.5	IDBac analysis of isolates.....	95
4.6	Genomic DNA isolation and 16S analysis for half of the cryopreserved <i>Streptomyces</i> sp. FERM BP-7812 isolates.....	96
4.7	Preliminary assessment of pladienolide B production in mixed culture of <i>Streptomyces</i> sp. FERM BP-7812	97
4.8	Comparison of pladienolide B production between isolate phenotype groups.....	98
4.9	Comparison of pladienolide B production between black isolates.....	100
4.10	Passaging study of black isolates.....	103
4.11	Phylogenetic tree and sequence alignment of donor ketoreductase domains.....	108
4.12	Phylogenetic tree of evolutionarily relevant acyl carrier protein domains from donor candidates	109
4.13	Phylogenetic tree of evolutionarily relevant ketosynthase domains from donor candidates	111
4.14	Vector map of pTC01	113
4.15	Construction of pTC01.....	114
4.16	Purification of recombinant PldMod3	115
4.17	Modified purification of recombinant PldMod3.....	116
4.18	Pladienolide B calibration curve.....	124

LIST OF ABBREVIATIONS

6-dEB	6-deoxyerythronolide B, the erythromycin aglycone
A _L	adenylation loading
ACP	acyl-carrier protein
Acu	aculeximycin
AT	acyltransferase
AT _L	loading acyltransferase
ATP	adenosine triphosphate
AUC	area under the curve
BGC	biosynthetic gene cluster
CMT	<i>C</i> -methyltransferase
CoA	coenzyme A
DE	dimerization element
DEBS	6-deoxyerythronolide B synthase
DH	dehydratase
DUF	domain of unknown function
ECF	energy-coupled factor
ER	enoylreductase
Ery	erythromycin
EtOAc	ethyl acetate
GNAT _L	GCN5 <i>N</i> -acetyltransferase like
Gon	PM100117
Hls	halstoctacosanolide
HMT	homocysteine <i>S</i> -methyltransferases
HPLC	high-pressure liquid chromatography
IPOD	International Patent Organism Depository

LIST OF ABBREVIATIONS (continued)

Las	lasalocid
Lkm	lankamycin
LLHR	linear-linear homologous recombination
KR	ketoreductase
KS	ketosynthase
KS ^Q	condensation-incompetent ketosynthase
MeCN	acetonitrile
Meg	megalomicin
MeOH	methanol
Mer	meridamycin
Mls	mycolactone
Mon	monensin
Nbr	brasilinolide
NRPS	non-ribosomal peptide synthase
PCR	polymerase chain reaction
Pik	pikromycin
PKS	polyketide synthase
PldB	Pladienolide B
SAH	<i>S</i> -adenosylhomocysteine
SAM	<i>S</i> -adenosyl-L-methionine
Sln	salinomycin
TE	thioesterase
TFA	trifluoroacetic acid

SUMMARY

Stereochemistry is a broad subject that describes the three-dimensional arrangement of atoms in a molecule. In Chapter 1, basic principles of stereochemistry are introduced, particularly those relating to molecular asymmetry. The importance of molecular asymmetry within biological systems is described in terms of protein structure and function, enzymatic discrimination of substrates, and development of pharmaceutical drugs. The following chapters describe two different projects that showcase the importance of stereochemistry at the levels of both primary and secondary metabolism.

Project 1: The essential biological cofactor *S*-adenosyl-L-methionine (SAM) is critical for maintenance of cellular homeostasis, but under physiological conditions it racemizes about its sulfonium center to yield an inactive diastereomeric form that is capable of competitively inhibiting SAM binding in enzymatic active sites. In Chapter 2, we tested the hypothesis that a conserved bacterial enzyme of unknown function, DUF62, specifically hydrolyzes the inactive diastereomeric form of SAM by isolating recombinant protein and assaying its function with each of the isolated SAM diastereomers. We obtained apparent kinetic constants for hydrolysis of the inactive form while no activity was shown towards the active cofactor. Thus, we achieved functional assignment of DUF62 as a stereoselective hydrolase responsible for maintenance of cofactor homochirality in the cell.

Project 2: Polyketide natural products are noted for their potent bioactivities, making them privileged scaffolds in drug discovery efforts. Polyketides feature complex stereochemistry and are biosynthesized by large, multifunctional, enzymatic assembly lines called polyketide synthases (PKSs). Chapter 3 reviews modifying a polyketide structure through PKS engineering. Beginning with an overview of PKS machinery and polyketide biosynthesis, we describe the various enzymatic domains within a PKS followed by numerous examples where domain engineering yielded modified polyketide structures. To facilitate understanding of the project discussed in Chapter 4, particular attention should be given to Section 3.4, which covers control of oxidation state and stereochemistry in a polyketide scaffold.

SUMMARY (continued)

Chapter 4 describes experiments carried out for a collaborative project with the Burkhardt lab. We seek to use synthetic biology for production of an unnatural polyketide analogue that exhibits improved therapeutic properties for treatment of acute myeloid leukemia and chronic lymphocytic leukemia. We propose performing a reductive loop swap of the third extension module in the pladienolide PKS to generate the target compound.

Our efforts were complicated by the pladienolide-producing microorganism, *Streptomyces* sp. FERM BP-7812, which was obtained from a culture collection as a heterogeneous mixture. Several phenotypically distinct isolates were cultivated. To determine which isolate could be used for our synthetic biology studies, we performed 16S analysis to assess taxonomic relatedness of the isolates and targeted metabolomics to determine the best isolate for fermentative production of pladienolide. Following deconvolution of the strain, the third extension module from the pladienolide PKS was successfully cloned and the recombinant protein was expressed and purified for use in *in vitro* extension assays. For the purposes of engineering a chimeric module that affords the desired oxidation state and stereochemistry, we identified a pool of donors and performed *in silico* analyses to determine the best candidates for the reductive loop swap.

Chapter 1

Introduction

1.1 Stereochemistry and Principles of Molecular Asymmetry

Depending on the arrangement of their atoms, two molecules with identical molecular formulas and bond connectivity may be structurally distinct. *Stereochemistry* is a broad subject that describes the three-dimensional arrangement of atoms in a molecule. One of the most intriguing stereochemical characteristics is *chirality*, or handedness. The defining feature of a chiral molecule is that it lacks a plane of symmetry and thus cannot be superimposed on its mirror image, as the left and right hands cannot be superimposed on one another. This is exemplified in the simple chiral molecule bromochlorofluoromethane. Figure 1.1 shows both possible *enantiomers* – nonsuperimposable, mirror-image forms – of bromochlorofluoromethane, which, together, form an *enantiomeric pair*.

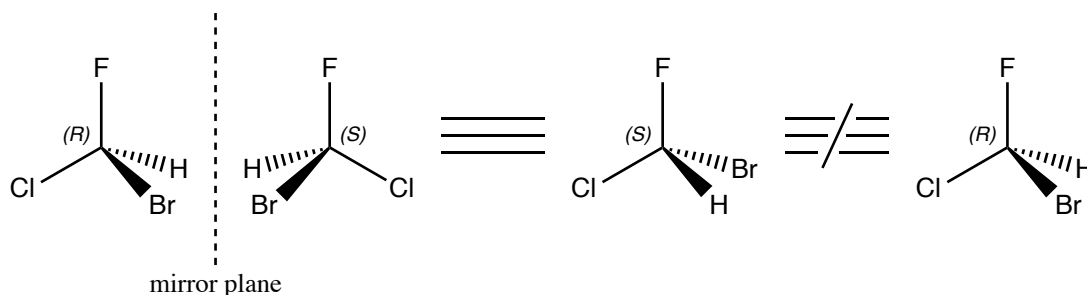


Figure 1.1: Enantiomers of bromochlorofluoromethane

Chirality is most often observed at sp^3 hybridized (tetrahedral) carbon centers with four unique substituents, as in bromochlorofluoromethane (Figure 1.1). Such a carbon may be referred to as a chiral carbon, a stereocenter, or an asymmetric carbon. Importantly, asymmetry may exist about atoms other than carbon. Though less common, stereocenters can form about nitrogen, phosphorous, and sulfur, as we will see in Chapter 2. In an achiral environment, an enantiomeric pair possesses nearly identical physio-chemical properties (boiling point, melting point, solubility, *etc.*). The sole exception is that enantiomers rotate in

opposite directions in plane-polarized light (Meierhenrich, 2008). Nevertheless, the two enantiomers are likely to possess remarkably different biological activities, indicating that biological systems form a chiral environment (Talapatra & Talapatra, 2015).

A molecule which contains more than one chiral center can form a pair of *diastereomers*, which are nonsuperimposable non-mirror images of one another. This is exemplified in Figure 1.2, which shows both enantiomeric pairs of the tetrose saccharides threose and erythrose, which are diastereomers of one another. Enantiomeric and diastereomeric relationships are indicated by red and blue arrows, respectively. Unlike enantiomers, diastereomers often exhibit markedly different physio-chemical properties (Nguyen *et al.*, 2006).

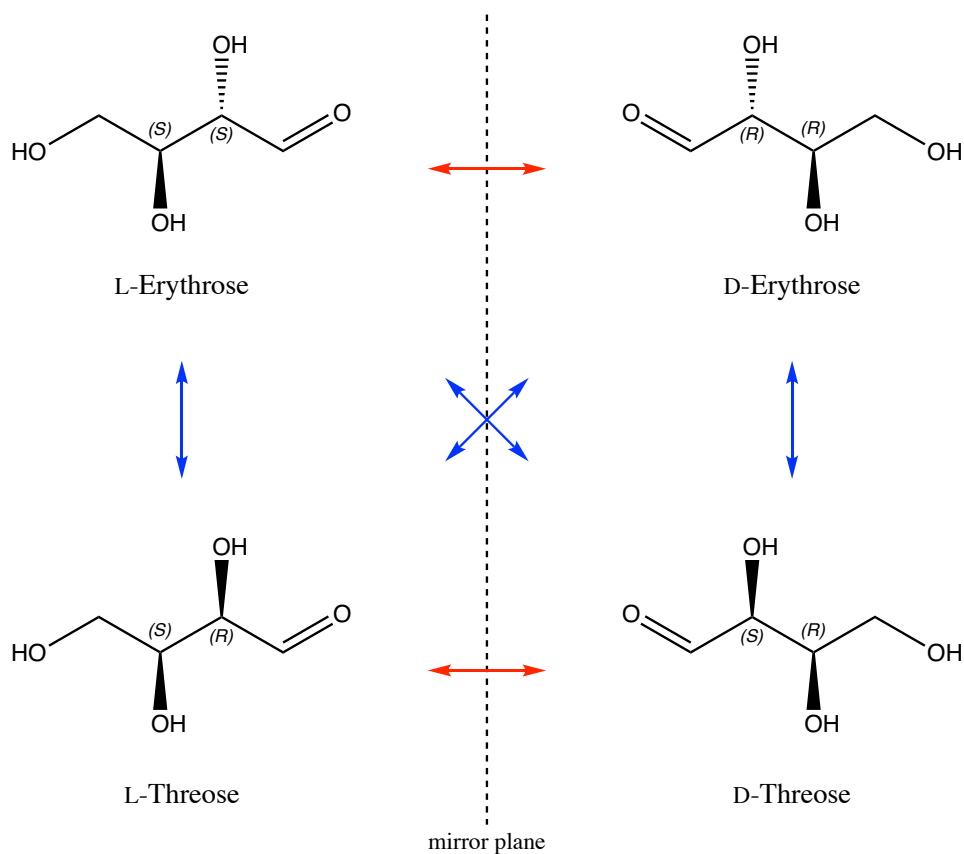


Figure 1.2: Stereochemical relationships of the tetrose saccharides.

As in Figure 1.1, the two possible configurations of substituents about an asymmetric carbon may be distinguished from one another using the *R-S* (or Cahn-Ingold-Prelog) system. The *R-S* system is most common in organic chemistry and provides the *absolute configuration* of a chiral center. Alternatively, one may describe an enantiomeric pair with the older D-L system, as in Figure 1.2. This system is more commonly used in biochemistry and provides the *relative configuration* of a chiral compound.

The stereochemical designation provided by the D-L system is always in relation to some reference compound, typically glyceraldehyde (Meierhenrich, 2008; Mathews *et al.*, 2013). Thus, the D-L system is typically less informative (*i.e.* more ambiguous) with respect to stereochemical nuance. Notice in Figure 1.2 that additional stereochemical information is provided for D- and L- threose by defining each of their two chiral centers with the appropriate *R-S* notation. Further, the D-L system does not resolve stereochemical differences between diastereomers (Talapatra & Talapatra, 2015). Still, both systems prove useful in various contexts, thus both the *R-S* and D-L system will be used herein.

1.2 The Importance of Molecular Asymmetry in Biological Systems

The first detailed observations of molecular asymmetry came in 1848 from the renowned French chemist Louis Pasteur. In an 1860 lecture on molecular asymmetry to the Chemical Society of Paris, he proclaimed:

“Thus we find introduced into physiological principles and investigations the idea of the influence of the molecular asymmetry of natural organic products, of this great character which establishes the only well marked line of demarcation that can at present be drawn between the chemistry of dead matter and the chemistry of living matter.” (Pasteur, 1964, p. 45)

Indeed, molecular asymmetry has emerged as a defining feature of living systems (Hegstrom & Kondepudi, 1990). Cells are composed of mostly chiral molecules (Inaki & Matsuno, 2016). Further, there is an overwhelming preference for one enantiomer over the other within each of the three major classes of

biomolecules – proteins, carbohydrates, and lipids – thus imbuing the cellular environment with asymmetric character (Mathews *et al.*, 2013). This phenomenon of preferred molecular handedness in living systems is referred to as *biological homochirality* (Hegstrom & Kondepundi, 1990; Blackmond, 2010). Though we are unlikely to ever arrive at a definitive conclusion, questions regarding the evolutionary origins of biological homochirality have captured the imaginations of scientists (Saghatelian, 2001; Blackmond, 2010). Additionally, much attention has been given to how biological systems maintain homochirality (Nandi, 2009).

Biological homochirality is most clearly observed in primary metabolic building blocks of proteins and polysaccharides – L-amino acids and D-sugars, respectively. The essential proteinogenic amino acids are all L-configured, though non-proteinogenic D-amino acids can be found in bacterial cell walls and other short bacterial peptides of non-ribosomal origin (Flügel, 2011). It has been shown that incorporation of D-amino acids in cellular proteins can be toxic to an organism (Ohnishi, 1962; Champney & Jensen, 1970). Further, translation of proteins with mixtures of L- and D-amino acids causes improper protein folding, preventing formation of the higher-order structures essential for their biological functions (Hegstrom & Kondepundi, 1990; Flügel, 2011).

Enzymes are the great chemists of the biological world. Given the asymmetric nature of their amino acid building blocks, it is no surprise that enzymes are also chiral. Indeed, enzyme active sites form chiral cavities, which facilitates discrimination between the two enantiomeric forms of a substrate (Inaki & Matsuno, 2016). Enzymatic action on a substrate requires that the substrate be tightly confined within the active site and positioned with proper orientation to make contact with catalytic residues (Nandi, 2009). One can imagine that the mirror image of an enzyme's natural substrate may not fit neatly into the active site, thus lowering the catalytic efficiency of any activity the enzyme may exert on that unnatural substrate.

The chiral nature of biological receptors explains the differential effects that are often observed between the two enantiomeric forms of chiral drugs. More than half of the drugs on the market are chiral compounds (Nguyen *et al.*, 2006). While many of these drugs differ primarily in potency between

enantiomers, for others the difference can be much more insidious. Drugs such as L-Dopa, a treatment for Parkinson's, must be given in enantiopure form due to the severe toxicity associated with the D-Dopa enantiomer (Nguyen *et al.*, 2006). Thus, an understanding of stereochemical nuance – at both the molecular and environmental level – is critical for medicine and the development of new drugs.

Herein, examples of stereochemistry at the level of both primary and secondary metabolism will be explored. Chapter 2 examines a bacterial enzyme of unknown function, which we show is a stereoselective hydrolase responsible for maintenance of cofactor homochirality. Chapters 3 and 4 cover polyketide natural products, which are noted for their potent biological activities and intricate stereochemistry.

1.3 References Cited

- Blackmond, D. G. (2010). The origin of biological homochirality. *Cold Spring Harbor Perspectives in Biology*, 2, a002147. doi: 10.1101/cshperspect.a002147
- Champney, W. S., & Jensen, R. A. (1970). Molecular events in the growth inhibition of *Bacillus subtilis* by D-Tyrosine. *Journal of Bacteriology*, 104, 107–116. doi: 10.1128/JB.104.1.107-116.1970.
- Flügel, R. M. (2011). *Chirality and Life: A Short Introduction to the Early Phases of Chemical Evolution*. Springer-Verlag.
- Hegstrom, R. A., & Kondepudi, D. K. (1990). The Handedness of the Universe. *Scientific American*, 262(1), 108–115.
- Inaki, M., Liu, J., & Matsuno, K. (2016). Cell chirality: Its origin and roles in left-right asymmetric development. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1710). doi: 10.1098/rstb.2015.0403
- Mathews, C.K., Van Holde, K.E., Appling, D.R., and Anthony-Cahill, S.J. (2013) *Biochemistry* (4th ed.). Toronto: Pearson.
- Meierhenrich, U. (2008). Stereochemistry for the Study of the Origin of Life. In *Advances in Astrobiology and Biogeophysics. Amino Acids and the Asymmetry of Life* (pp. 17–46). Springer, Berlin, Heidelberg. doi: 10.1007/978-3-540-76886-9
- Nandi, N. (2009). Chiral discrimination in the confined environment of biological nanospace: reactions and interactions involving amino acids and peptides. *International Reviews in Physical Chemistry*, 28(2), 111–167. doi: 10.1080/01442350902999682
- Nguyen, L. A., He, H., & Pham-Huy, C. (2006). Chiral Drugs: An Overview. *International Journal of Biomedical Science*, 2(2), 85–100.

- Ohnishi, E., Macleod, H., & Horowitz, N. H. (1962). Mutants of Neurospora in D-Amino Oxidase. *Journal of Biological Chemistry*, 237(1), 138–142.
- Pasteur, L. (1964). Researches on the Molecular Asymmetry of Natural Organic Products. In *Alembic Club Reprints: No. 14*. (Re-issue ed., pp. 26-46). Edinburgh, UK: E. & S. Livingstone. (Original work published 1860)
- Saghatelian, A., Yokobayashi, Y., Soltani, K., & Ghadiri, M. R. (2001). A chiroselective peptide replicator. *Nature*, 409(February), 797–801.
- Talapatra, S. K., & Talapatra, B. (2015). Fundamental Stereochemical Concepts and Nomenclatures. In *Chemistry of Plant Natural Products: Stereochemistry, Conformation, Biology, and Medicine* (pp. 23–201). Springer, Berlin, Heidelberg. doi: 10.1007/978-3-642-45410-3

Chapter 2

Maintenance of cofactor homochirality in bacterial cells:

Assignment of function to a stereoselective hydrolase

Adapted from ChemBioChem, An Enzyme Containing the Conserved Domain of Unknown Function DUF62 Acts as a Stereoselective (R_s, S_c)-*S*-Adenosylmethionine Hydrolase. Kornfuehrer, T., Romanowski, S.B.R., de Crécy-Lagard, V., Hanson, A.D., and Eustáquio, A.S.E. (2020)

Copyright © 2020 Wiley-VCH GmbH

[ChemBioChem, 2020 (in press), <https://doi.org/10.1002/cbic.202000349>]

2.1 Introduction

Molecular asymmetry plays an important role in biological systems, with homochirality being a “signature of life”, as exemplified by the predominance of D-sugars and L-amino acids in biological processes (Blackmond, 2010). Although carbon chirality is well recognized, asymmetry about sulfur is far less appreciated (Bentley, 2005). Pre-eminent among biomolecules with chirality about sulfur is the cofactor *S*-adenosyl-L-methionine (SAM), a sulfonium salt (Figure 2.1). Present in all living organisms, SAM can fulfill diverse biological roles (Dalhoff & Weinhold, 2008). Most significantly, SAM is the primary single carbon donor in biological transmethylation reactions (Cantoni, 1975; Martin & McMillan, 2002). The methyl acceptor substrates are diverse in nature, ranging from simple primary or secondary metabolites to complex macromolecules such as DNA, RNA and proteins (Struck *et al.*, 2012). Though methyl group donation is SAM’s primary role in the cell, SAM can also be a source of ribosyl, amino, and aminoalkyl groups, in addition to 5′-deoxyadenosyl radicals (Fontecave *et al.*, 2004).

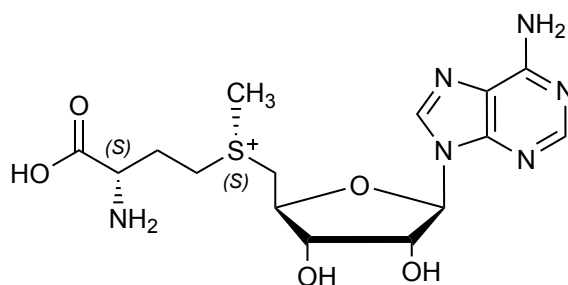


Figure 2.1: The biological cofactor *S*-adenosyl-L-methionine.

The biological cofactor SAM is synthesized in cells from adenosine triphosphate (ATP) and L-methionine (Cantoni, 1953). As shown in Figure 2.2, biological systems exclusively synthesize the (*S,S*)-SAM diastereomer (de la Haba *et al.*, 1959; Cornforth *et al.* 1977). However, biosynthesized (*S,S*)-SAM

slowly racemizes about the sulfonium center in physiological conditions to the (*R,S*)-SAM diastereomer with a reported rate constant of $1.8 \times 10^{-6} \text{ s}^{-1}$ at 37°C, pH 7.5 (Hoffman, 1986). Biological transmethylation reactions exhibit a clear preference for (*S,S*)-SAM as a substrate (de la Haba *et al.*, 1959), and there is evidence to suggest that (*R,S*)-SAM inhibits methyltransferases (Borchardt & Wu, 1976). As macromolecule methylation plays a crucial role in health and disease from humans to microorganisms (Michalak *et al.*, 2019; Kako *et al.*, 2019; Sánchez-Romero & Casadesús, 2019), buildup of (*R,S*)-SAM could have deleterious effects on cell health. Thus, it is imperative for cells to harbor mechanisms of (*R,S*)-SAM remediation (Vinci & Clarke, 2010a).

One such mechanism known for yeast, worms, plants, and flies is mediated by homocysteine *S*-methyltransferases (HMTs), which preferentially accept the (*R,S*)-SAM diastereomer as a methyl donor to generate L-methionine from homocysteine, preventing the cellular accumulation of (*R,S*)-SAM (Vinci & Clarke, 2007; Vinci & Clarke, 2010b; Bradbury *et al.*, 2014). The co-product of the HMT-catalyzed reaction, *S*-adenosylhomocysteine (SAH) – itself a methyltransferase inhibitor – can then be degraded by SAH hydrolases (Vinci & Clarke, 2007). An alternative enzyme is likely required in mammalian cells and in bacteria that do not contain HMTs; only 18% of bacteria have been reported to contain HMTs (Vinci & Clarke, 2010b; Bradbury *et al.*, 2014).

Previous work showed that Domain of Unknown Function 62 (DUF62) proteins have SAM hydrolase activity (Eustáquio *et al.*, 2008a; Deng *et al.*, 2008; Deng *et al.*, 2009). The existence of an enzyme that degrades SAM is counterintuitive as biosynthesis of SAM is metabolically costly (Wu *et al.*, 1983), and the cell is poised to make use of the reactivity of the sulfonium center in various ways (Fontecave *et al.*, 2004). Yet, these studies did not examine the differential activity of DUF62 on each of the sulfur diastereomers. Herein we tested the hypothesis that DUF62 specifically hydrolyzes (*R,S*)-SAM. Structures of the active (*S,S*) and inactive (*R,S*) forms of SAM are shown in Figure 2.2, which depicts the main role of (*S,S*)-SAM (nucleic acid, protein, and metabolite methylation) and our hypothesis that DUF62 enzymes specifically cleave (*R,S*)-SAM.

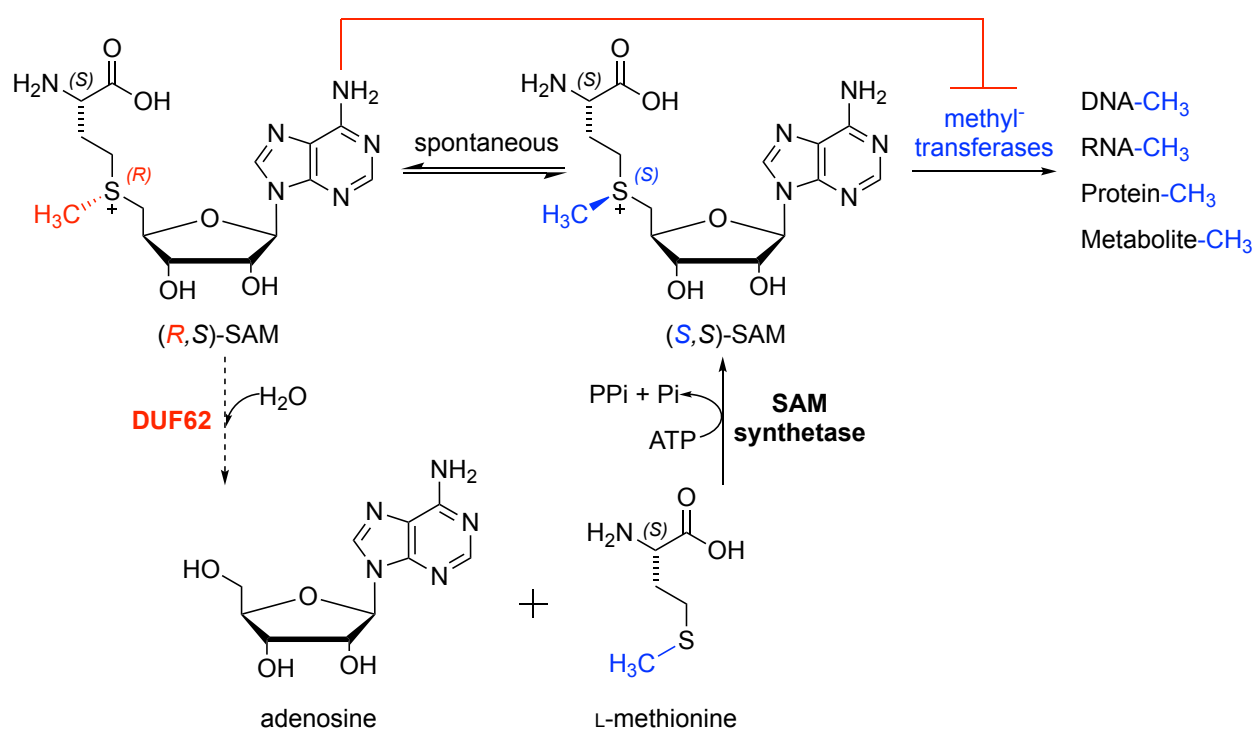


Figure 2.2: S-adenosyl-L-methionine biosynthesis, racemization, and hypothesis of DUF62 function

2.2 Results and Discussion

2.2.1 Analysis of diastereomer composition in commercial standards of *S*-adenosyl-L-methionine

Two different commercial SAM standards were obtained from BioVision Incorporated and Sigma-Aldrich. Using the SAM diastereomer *Analysis and Quantitation* method described in Section 2.4.2.1, a 1 mM standard solution was analyzed by high performance liquid chromatography (HPLC) for each commercial preparation (Figure 2.3). The BioVision standard is apparently a racemate, whereas the Sigma standard is 75:25 (*S,S*):(*R,S*). Although racemization of SAM at the sulfur center has long been described (Hoffman, 1986; Wu *et al.*, 1983), the fact that commercial preparations of SAM are of varied purity as it pertains to the active (*S,S*)-SAM diastereomer remains underappreciated.

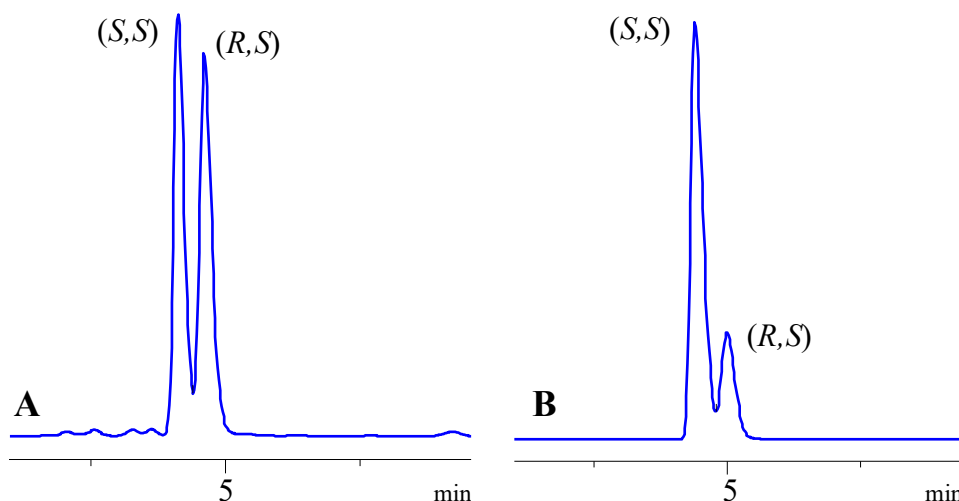


Figure 2.3: Diastereomeric composition of commercial *S*-adenosyl-L-methionine preparations. (A) BioVision (B) Sigma-Aldrich.

2.2.2 Cloning and purification of recombinant StDUF62

In order to test the hypothesis that DUF62 proteins function as stereoselective (*R,S*)-SAM hydrolases, we generated recombinant DUF62 from *Salinispora tropica* CNB-440 (StDUF62), a marine actinomycete bacterium whose genome (Udwary *et al.*, 2007) contains two DUF62 genes. One of these has been previously shown to encode a SAM-dependent chlorinase (Eustáquio *et al.*, 2008b). Moreover, no HMT genes are encoded in the genome of CNB-440.

We cloned *duf62* into the expression vector pHis8 (via traditional, restriction-ligation cloning methods to generate pHis8-St62. A vector map with relevant restriction sites is shown in Figure 2.4A. Amplification of *duf62* with addition of BamHI and XhoI restriction sites was achieved via polymerase chain reaction (PCR). Non-specific amplification was observed, and the 810 bp amplicon had to be gel-excised (Figure 2.4B). This was likely the result of high GC content (69.5%) in the *S. tropica* genome (Udwary *et al.*, 2007). Screening of obtained clones with SalI and EcoRV yielded two positive hits, visible in wells 7 and 8 of Figure 2.4C.

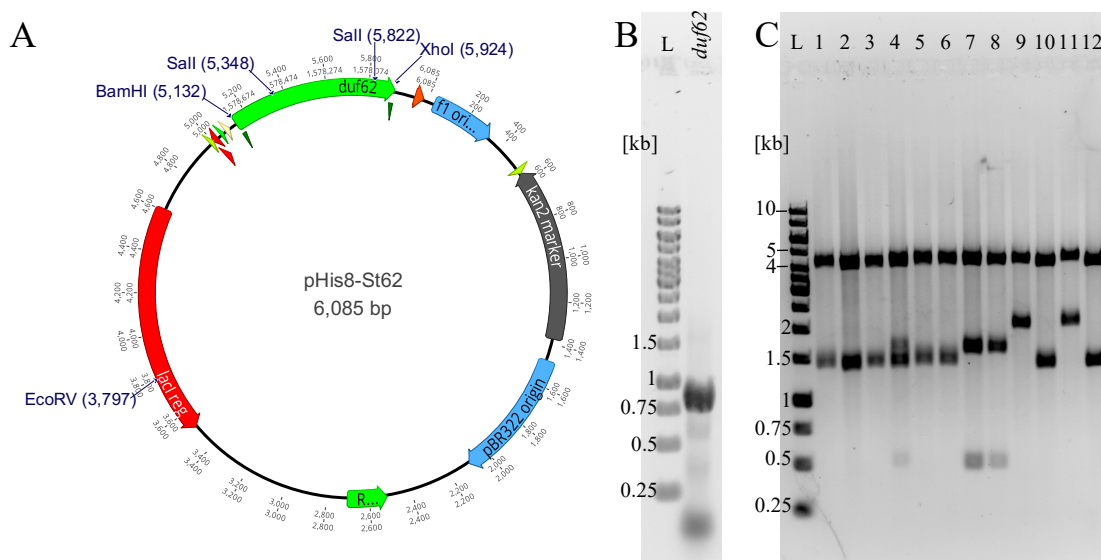


Figure 2.4: Cloning *duf62* into expression vector pHis8. (A) Vector map of pHis8-St62 (B) PCR amplification of *duf62* (C) Clone screening

Once the fidelity of the cloned construct was verified with Sanger sequencing, pHis8-St62 was electroporated into *Escherichia coli* BL21 (DE3) for protein expression and purification via Ni-NTA chromatography. The purified enzyme was analyzed by SDS-PAGE (Figure 2.5). Cells did not lyse completely upon first sonication; thus sonication was repeated a second time. In Figure 2.5, “lysate-1” refers to samples following the first round of sonication while “lysate-2” refers to samples following the second round of sonication. The purified protein was buffer exchanged (desalted) via gel filtration using disposable PD-10 columns containing Sephadex G-25M resin. Soluble (s) and insoluble (i) fractions are indicated.

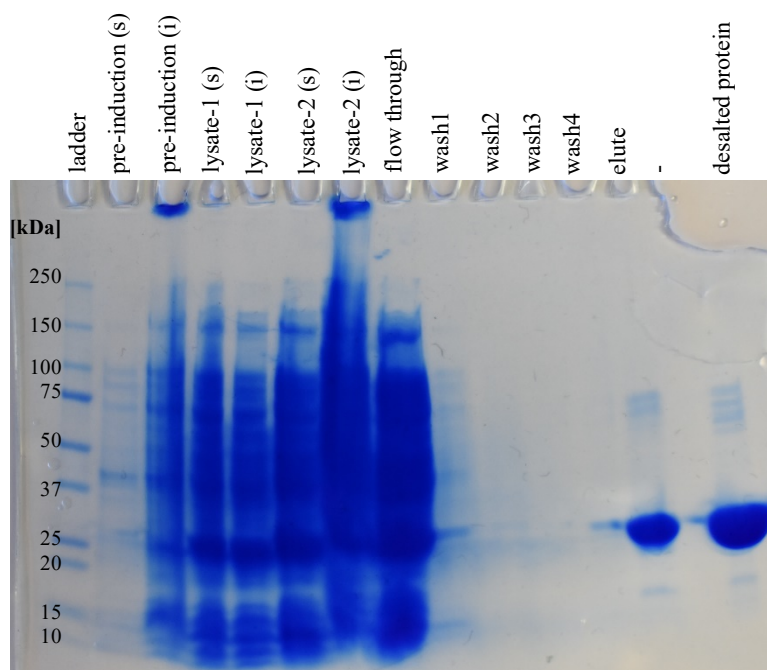


Figure 2.5: Purification of recombinant StDUF62

2.2.3 Preliminary assay with mixed diastereomers

Incubation of recombinant StDUF62 with the BioVision SAM racemate led to a decrease in the (*R,S*)-SAM peak, while the (*S,S*)-SAM peak remained relatively unchanged (Figure 2.6). The decrease in the (*R,S*)-SAM peak was concomitant with adenosine formation, identified by comparison with an authentic standard. This preliminary assay supports the hypothesis that StDUF62 acts preferentially on (*R,S*)-SAM.

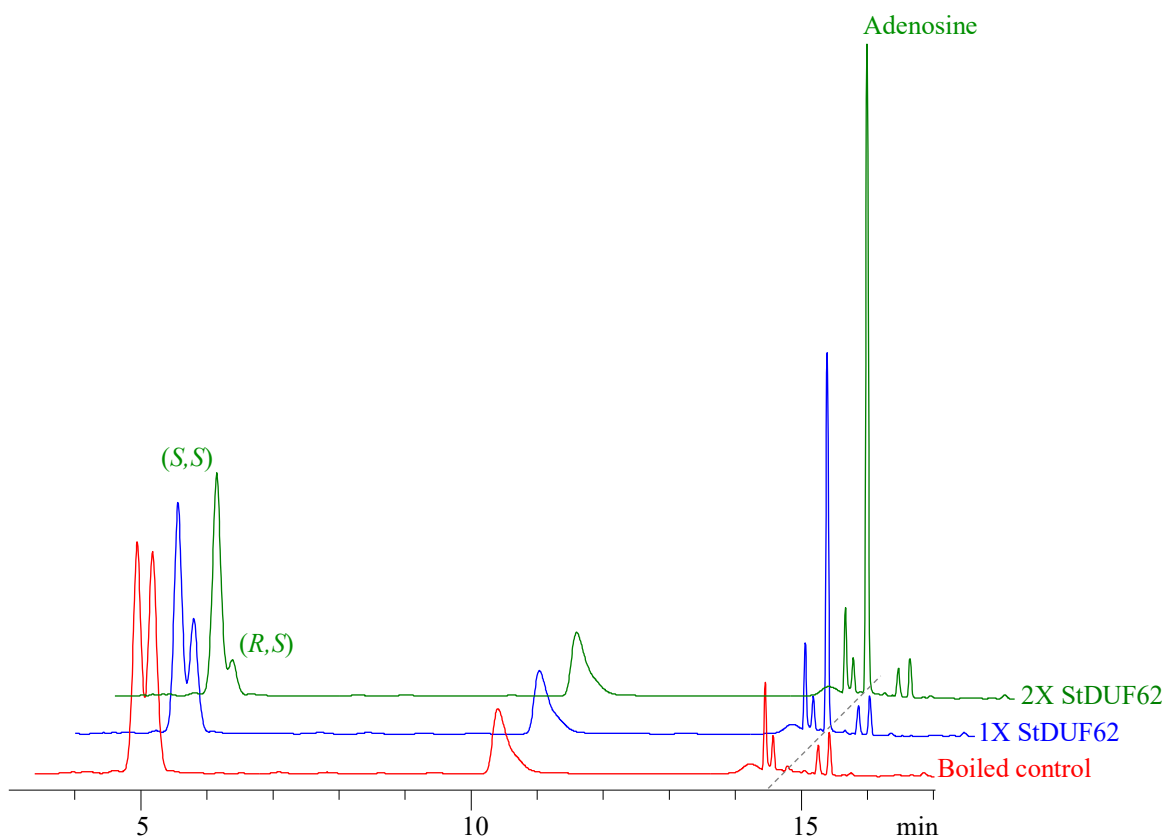


Figure 2.6: Assay of StDUF62 with *S*-adenosyl-L-methionine racemate indicates preference for the (*R,S*) diastereomer. 1mM SAM racemate was incubated with 2.5 μ M (1X) or 5 μ M StDUF62 (2X) for 1 hour at 30 $^{\circ}$ C.

2.2.4 Determination of kinetic constants for hydrolysis of isolated diastereomers

In order to test the stereoselectivity of StDUF62 further, we isolated each diastereomer from a racemic mixture as described in Section 2.4.2.1. The apparent kinetic constants we obtained for (*R,S*)-SAM using nonlinear regression are K_m of 0.63 μM and k_{cat} of $4.5 \times 10^{-3} \text{ s}^{-1}$. Kinetics assays for the (*R,S*) diastereomer involved 0-16 μM (*R,S*)-SAM incubated with 50 nM StDUF62 in 100 μL reactions at 30°C for 4 min. Data obtained from two independent experiments of triplicate assays (N=6 total) is shown in TABLE I. Previously, DUF62 proteins have been shown to co-purify with adenosine by X-ray crystallography (Deng *et al.*, 2008). Therefore, the area under the curve (AUC) for adenosine obtained with the enzyme only samples (no substrate, “blank”) is used to correct the total AUC obtained for the substrate with enzyme samples. Box plot analysis of the data using XLSTAT in Excel identified one outlier data point, indicated with a red arrow in Figure 2.7. This outlier was removed prior to nonlinear regression analysis for determination of kinetic constants (Figure 2.8).

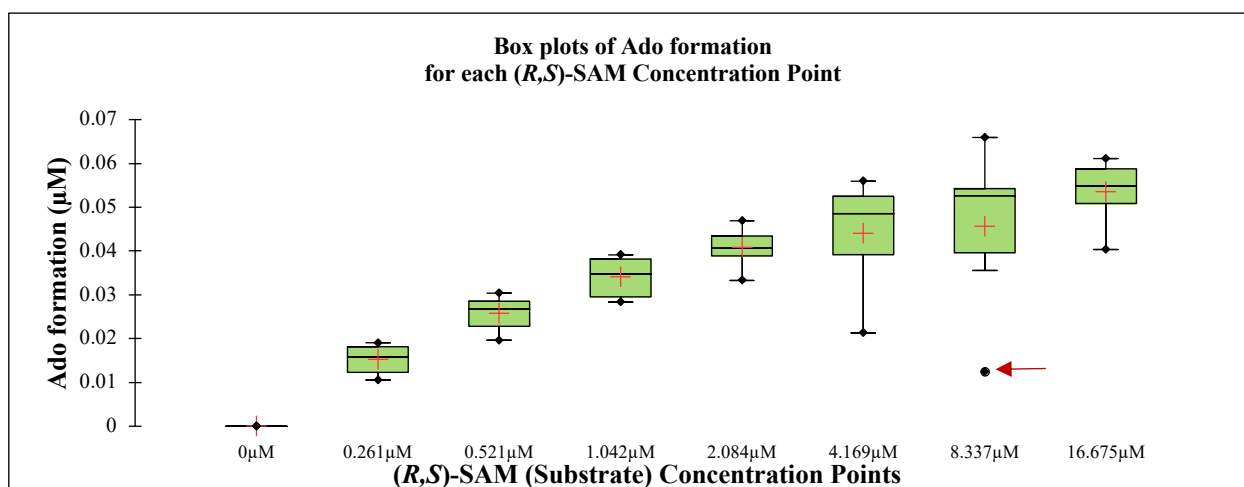


Figure 2.7: Identification of outlier data point via box plot analysis

TABLE I: RAW KINETICS DATA FOR (R,S)-S-ADENOSYL-L-METHIONINE

N	[(R,S)-SAM] (μM)	adenosine peak AUC (mAU*s)		Calculated [Ado] (μM)	% of SAM converted
		total AUC	blank corrected		
1	0.0	0.5016	0.0000	0.0000	-
	0.3	1.1799	0.6783	0.0160	6.1%
	0.5	1.7924	1.2908	0.0304	5.8%
	1.0	2.1475	1.6459	0.0387	3.7%
	2.1	2.4947	1.9930	0.0469	2.3%
	4.2	2.8784	2.3768	0.0559	1.3%
	8.3	3.3026	2.8009	0.0659	0.8%
	16.7	3.0286	2.5270	0.0595	0.4%
2	0.0	0.5794	0.0000	0.0000	-
	0.3	1.0269	0.4474	0.0105	4.0%
	0.5	1.5217	0.9423	0.0222	4.3%
	1.0	1.7904	1.2110	0.0285	2.7%
	2.1	2.2555	1.6761	0.0394	1.9%
	4.2	1.4850	0.9056	0.0213	0.5%
	8.3 ^a	1.1056	0.5261	0.0124	0.1%
	16.7	2.7078	2.1283	0.0501	0.3%
3	0.0	0.5246	0.0000	0.0000	-
	0.3	1.1847	0.6602	0.0155	6.0%
	0.5	1.5840	1.0594	0.0249	4.8%
	1.0	1.7280	1.2034	0.0283	2.7%
	2.1	1.9386	1.4140	0.0333	1.6%
	4.2	2.1057	1.5811	0.0372	0.9%
	8.3	2.0357	1.5111	0.0356	0.4%
	16.7	2.2388	1.7143	0.0403	0.2%

^a Outlier data point that was excluded from calculation of kinetic constants.

TABLE I: RAW KINETICS DATA FOR (R,S)-S-ADENOSYL-L-METHIONINE (CONT.)

N	[(R,S)-SAM] (μM)	adenosine peak AUC (mAU*s)		Calculated [Ado] (μM)	% of SAM converted
		total AUC	blank corrected		
4	0.0	0.6400	0.0000	0.0000	-
	0.3	1.1198	0.4798	0.0113	4.3%
	0.5	1.4743	0.8344	0.0196	3.8%
	1.0	2.1963	1.5563	0.0366	3.5%
	2.1	2.2878	1.6479	0.0388	1.9%
	4.2	2.8614	2.2214	0.0523	1.3%
	8.3	2.9567	2.3167	0.0545	0.7%
	16.7	3.0546	2.4146	0.0568	0.3%
5	0.0	0.5295	0.0000	0.0000	-
	0.3	1.3314	0.8019	0.0189	7.2%
	0.5	1.7420	1.2125	0.0285	5.5%
	1.0	2.1915	1.6620	0.0391	3.8%
	2.1	2.3085	1.7790	0.0419	2.0%
	4.2	2.7614	2.2319	0.0525	1.3%
	8.3	2.7252	2.1957	0.0517	0.6%
	16.7	3.1259	2.5964	0.0611	0.4%
6	0.0	0.3829	0.0000	0.0000	-
	0.3	1.1929	0.8100	0.0191	7.3%
	0.5	1.5981	1.2152	0.0286	5.5%
	1.0	1.7808	1.3979	0.0329	3.2%
	2.1	2.2538	1.8709	0.0440	2.1%
	4.2	2.2889	1.9060	0.0449	1.1%
	8.3	2.6570	2.2741	0.0535	0.6%
	16.7	2.6321	2.2492	0.0529	0.3%

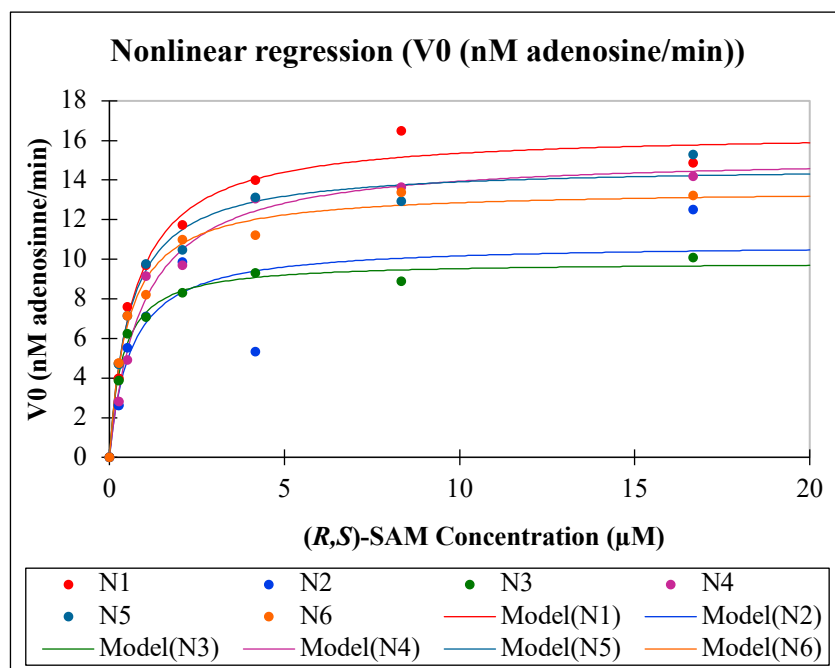


Figure 2.8: Kinetic constants for (R,S)-S-adenosyl-L-methionine determined via linear regression

For qualitative determination of StDUF62 activity on (S,S)-SAM, 10 μM (S,S)-SAM was incubated with 0.5 μM StDUF62 at 30°C for 1 hour. We observed 52 nM adenosine formation, corresponding to only 0.5% of (S,S)-SAM in the assay (Figure 2.9, TABLE II). Over the one-hour incubation, 65 nM of (R,S)-SAM would theoretically be formed from 10 μM (S,S)-SAM using the reported k_R of $1.8 \times 10^{-6} \text{ s}^{-1}$, which was determined at 37°C (Hoffman, 1986). According to these calculations, the observed product formation may be explained by racemization of (S,S)-SAM followed by degradation of newly formed (R,S)-SAM by StDUF62. Here and throughout, “E(b)” is boiled enzyme only (control); “S₀/S” is substrate only (control); “S+E(b)” is substrate with boiled enzyme (control); and “S+E” is substrate with enzyme (experimental).

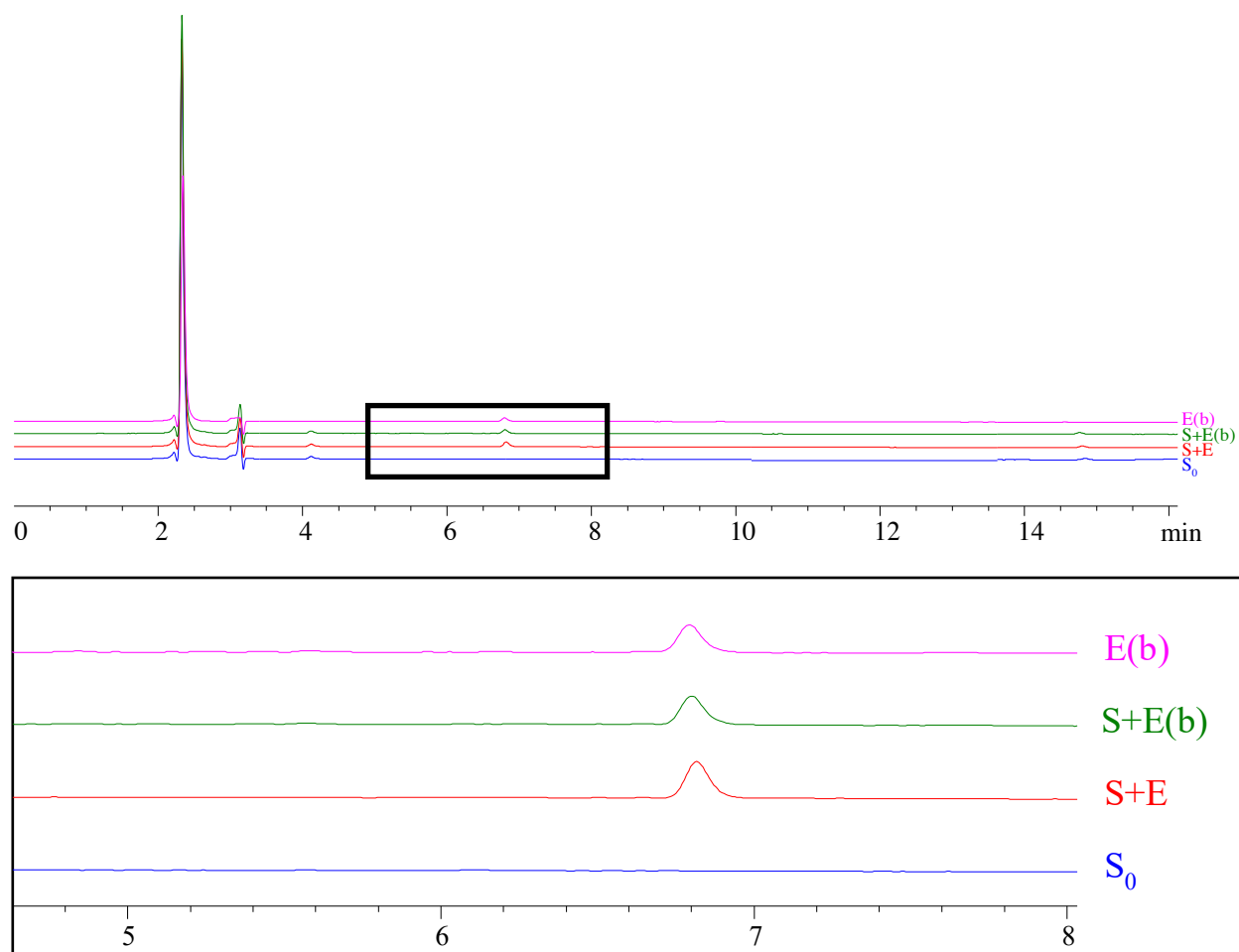


Figure 2.9: Testing the activity of StDUF62 with (S,S)-S-adenosyl-L-methionine at 30°C

TABLE II: ADENOSINE FORMATION DURING ONE HOUR INCUBATION OF DUF62 WITH (S,S)- S-ADENOSYL-L-METHIONINE AT 30°C

	Adenosine Formation (μM)			Average (μM)	Std Dev (μM)	[Adenosine] Corrected (μM)
	N=1	N=2	N=3			
S_0	0.00000	0.00000	0.00000	0.00000	0.00000	0
$S + E$	0.22279	0.22890	0.22752	0.22640	0.00321	0.05158
$S + E(b)$	0.17488	0.17507	0.17450	0.17482	0.00029	0
$E(b)$	0.16486	0.17812	0.17447	0.17248	0.00685	0

To test this, we incubated 10 μM (*S,S*)-SAM in assay buffer in absence of enzyme for 1 hour at 30°C. However, we were not able to detect (*R,S*)-SAM using the SAM diastereomer *Analysis and Quantification* method, as 52 nM (*R,S*)-SAM falls below our detection limit. We then tested an extended incubation time with 10 μM (*S,S*)-SAM incubated in phosphate buffer (pH 7.9) at 30 °C for 24 hours (Figure 2.10). Aliquots taken at the indicated time (5, 12, or 24 hours) were analyzed by HPLC using the *SAM Analysis and Quantitation* method.

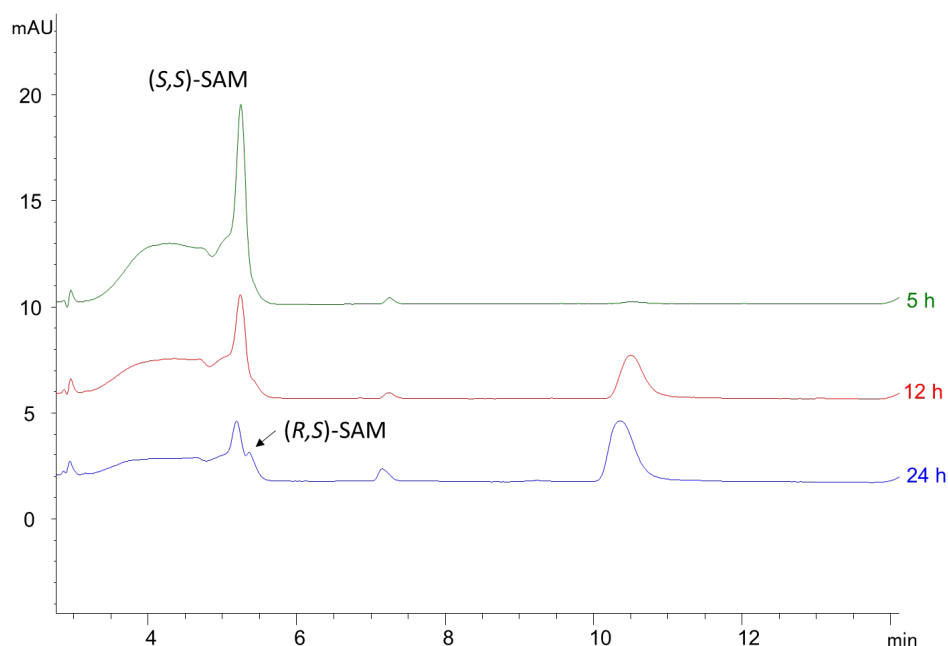


Figure 2.10: Extended incubation of (*S,S*)-*S*-adenosyl-L-methionine at 30 °C.

In addition to epimerization to (*R,S*)-SAM, two other pathways for (*S,S*)-SAM degradation have been reported, *i.e.* cleavage to 5'-methylthioadenosine and homoserine lactone, and hydrolysis to adenine and *S*-pentosylmethionine (Hoffman, 1986). These competing pathways for (*S,S*)-SAM degradation at physiological pH, and tailing of the (*S,S*)-SAM peak into the (*R,S*)-SAM peak prevented accurate experimental determination of (*R,S*)-SAM formation under the StDUF62 assay conditions.

We then repeated the StDUF62 assay on (*S,S*)-SAM but this time incubated the reaction at 37 °C so that the reported recemization rate can be used more accurately. We observed formation of 56 nM adenosine (TABLE III, Figure 2.11), which is comparable to the results we obtained at 30 °C, indirectly confirming the premise that the observed adenosine stems from (*S,S*)-SAM epimerization. In any case, regardless of whether the trace adenosine amounts observed stems from epimerization of (*S,S*)-SAM under the assay conditions or actual, low activity of StDUF62 toward (*S,S*)-SAM, StDUF62 shows a clear preference for (*R,S*)-SAM and virtually no activity toward (*S,S*)-SAM.

TABLE III: ADENOSINE FORMATION DURING ONE HOUR INCUBATION OF DUF62 WITH
(*S,S*)-*S*-ADENOSYL-L-METHIONINE AT 37°C

Sample	Adenosine Formed (μM)			Average (μM)	Std Dev (μM)	Adenosine (μM) corrected
	N=1	N=2	N=3			
S + E	0.22358	0.22594	0.22358	0.22437	0.00136	0.05648
E	0.16710	0.16710	0.16945	0.16788	0.00136	0.00000

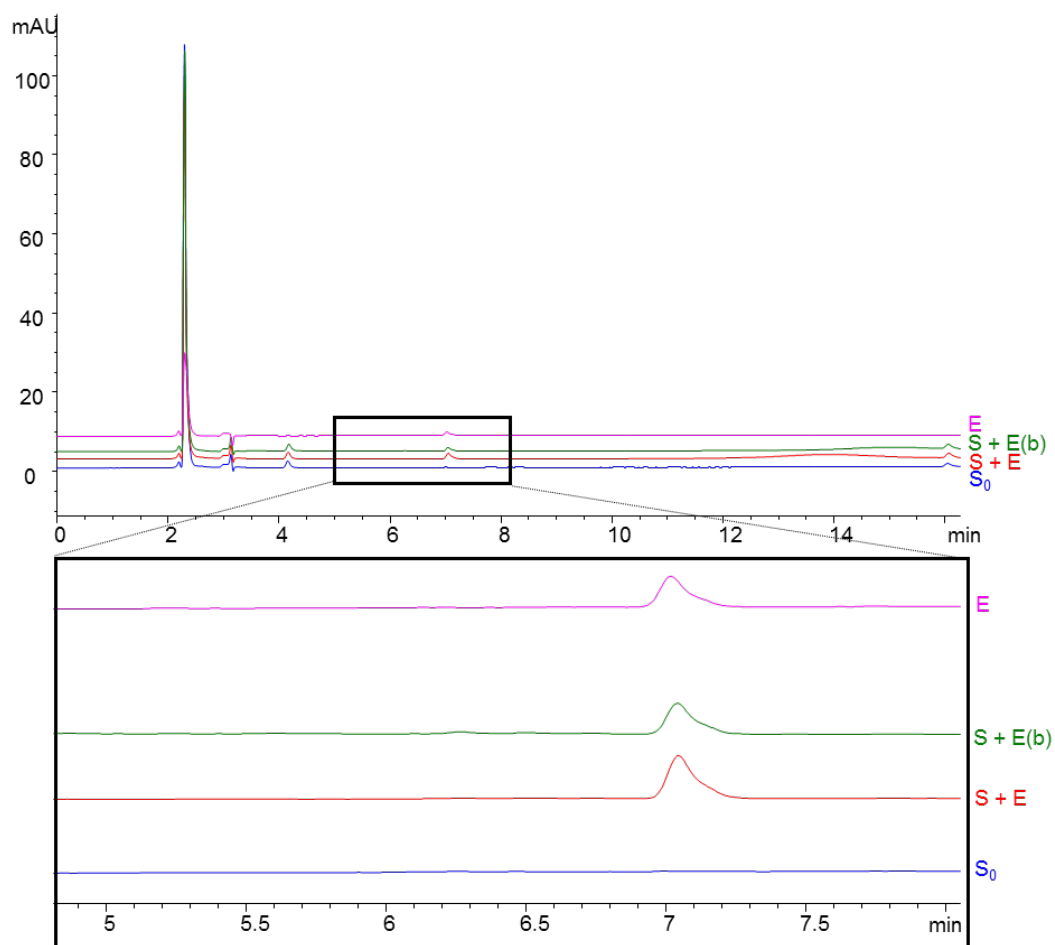


Figure 2.11: Testing the activity of StDUF62 with (S,S)-S-adenosyl-L-methionine at 37°C

The observed kinetic constants for StDUF62 (TABLE IV) appear physiologically relevant and are in agreement with a role in preventing accumulation of (*R,S*)-SAM in cells. Average values of the six (*R,S*)-SAM kinetics data sets are shown \pm standard deviation in TABLE IV. The StDUF62 very slow (*R,S*)-SAM turnover rate of $4.5 \times 10^{-3} \text{ s}^{-1}$ is comparable to the reported turnover of halogenase paralogs (Dong *et al.*, 2004; Zhu *et al.*, 2007; Eustáquio *et al.*, 2008b). In previous studies, DUF62 enzymes were tested with halides but showed no relevant activity (Eustáquio *et al.*, 2008a; Deng *et al.*, 2008). Moreover, SAM analogues SAH and 5'-deoxy-5'-methylthioadenosine did not function as substrates either (Eustáquio *et al.*, 2008a), and SAH was shown to function as an inhibitor (Deng *et al.*, 2008).

Finally, a calculation of the rate of (*R,S*)-SAM formation based on the reported racemization rate of $1.8 \times 10^{-6} \text{ s}^{-1}$ (Hoffman, 1986), and an intracellular (*S,S*)-SAM concentration of 180 μM (Bennett *et al.*, 2009; Halliday *et al.*, 2010), favors the possibility that even a low concentration of StDUF62 ($<0.001\%$ of total protein) could deal with the *in vivo* rate of (*S,S*)-SAM racemization (see calculation in section 2.4.6). Future studies such as comparative analysis of StDUF62 deletion mutant and wild-type strains will be necessary to determine whether StDUF62 indeed prevents accumulation of (*R,S*)-SAM in cells.

TABLE IV: APPARENT KINETIC CONSTANTS OF STDUF62

Substrate	K_m (μM)	k_{cat} (s^{-1})	Relative activity (%)	k_{cat}/K_m ($\mu\text{M}^{-1} \text{ s}^{-1}$)
(<i>R,S</i>)-SAM	0.63 ± 0.19	$4.5 \times 10^{-3} \pm 0.9 \times 10^{-3}$	100	7.1×10^{-3}
(<i>S,S</i>)-SAM	---	---	---	---

2.2.5 Bioinformatic analysis

The previously reported mechanism to prevent the cellular accumulation of (*R,S*)-SAM levels is mediated by HMTs (Vinci & Clarke, 2007; Vinci & Clarke 2010a). Previous analyses (Bradbury *et al.*, 2014) using the PubSEED database and subsystem tools had shown that 18% (280/1,535) of the bacterial genomes analyzed but none (0/65) of the archaeal genomes encoded an HMT. We first performed a DUF62 distribution analysis using the same PubSEED database and tools and found DUF62 homologs in 15% (223/1,535) of bacterial and in 56% (37/65) of archaeal genomes.

Because the PubSEED genome set includes only genomes released up to 2014, we also used the AnnoTree tools to obtain more current data from a larger genome set. The AnnoTree analysis using the PF01887 PFAM family, which contains DUF62 and a handful of fluorinase/chlorinase enzymes (Eustáquio *et al.*, 2008b; Dong *et al.*, 2004) showed that 21% (5,052/23,458) of bacterial and 42% (531/1,248) of archaeal genomes encoded DUF62 proteins, distributed over the whole bacterial and archaeal phylogenetic trees (Figure 2.12A and Figure 2.12B, respectively). As the PFAM family for HMT (PF02574) is not isofunctional, we used KOG family K00547 that is better annotated with only HMT proteins (EC 2.1.1.10) as query. This analysis showed that HMT proteins were encoded by 27% (6,392/23,458) of the bacterial and by 5.9% (74/1,248) of the archaeal genomes. In addition, 56% (13,084/23,458) of the bacteria and 52% (655/1,248) of the archaea analyzed had neither HMT nor DUF62. From the genomes that do encode these proteins, there is a partial inverse relationship between the presence of DUF62 homologs and of HMT, except that 10.3% (1,073/10,374) of the bacterial and 2.4% (14/593) of the archaeal genomes encode both HMT and DUF62, respectively (Figure 2.12C). To the best of our knowledge DUF62 homologs are not found in eukaryotes. Two eukaryotic hits are found in the PF01887 family, but further BLAST analysis strongly suggests that these are bacterial contaminations.

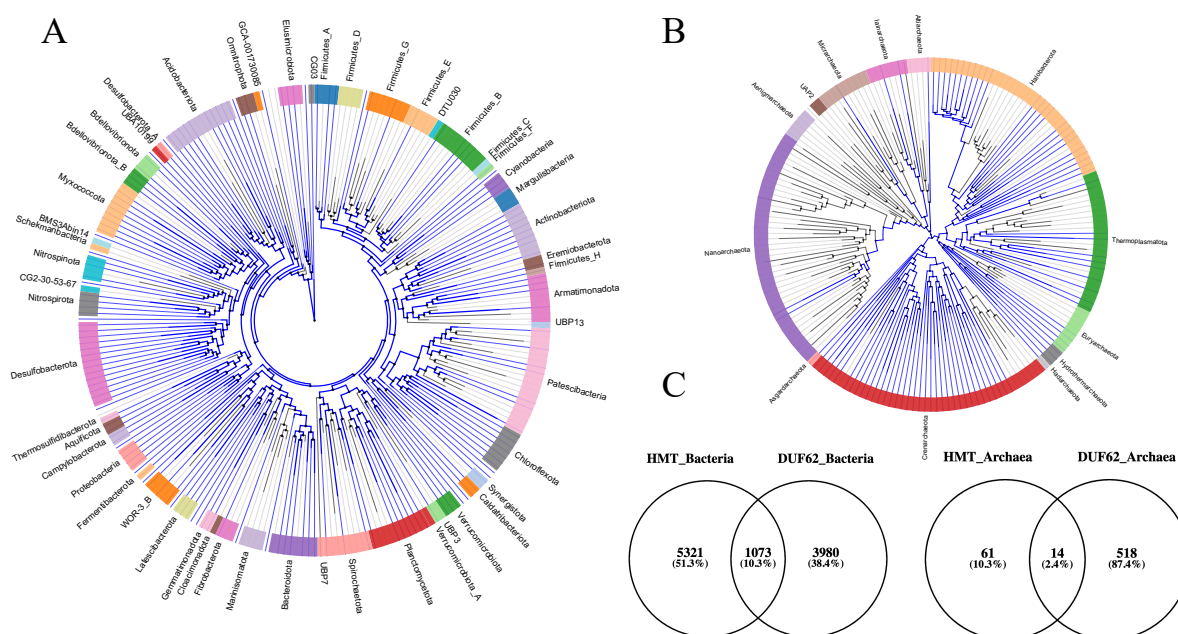


Figure 2.12: Phylogenetic distribution of DUF62 in bacteria and archaea. Results of an AnnoTree query using PFAM family PF01887. Branches are highlighted in blue for phyla that harbor members of the DUF62 family in the Bacteria (A) and Archaea (B) domains. The trees were generated directly in AnnoTree. (<http://annotree.uwaterloo.ca/app/#/?qtype=pfam&qstring=PF01887&eval=0.00001>). (C) Venn diagrams showing the number of genomes that encode DUF62 and HMT homologs in Bacteria (left) and in Archaea (right).

Comparative genomics and physical clustering analysis can provide insight into how genes may relate to biological processes (Osterman & Overbeek, 2003). Although half of the DUF62 genes in the PubSEED database do not show conserved clustering, the other half (882/1,740 bacteria) are physically clustered with *mtsABC* genes annotated as methionine-regulated, energy-coupling factor (ECF) transporters (Rodionov *et al.*, 2004) (see AdoMetRepair_VDC SEED subsystem link provided in section 2.4.7). The probability that two genes cluster together by chance in an average bacterial genome containing ~4,500 genes is very low (0.02%); thus to observe clustering in 50% of the cases is a strong correlation. In addition, *mtsABC* and DUF62 genes are predicted to be co-transcribed and to be under the control of the methionine-sensing regulator MtaR (Shelver *et al.*, 2003) in several firmicutes (Figure 2.13). Based on the (*R,S*)-SAM

hydrolase activity of StDUF62 reported here, it is likely that this subgroup of ECF transporters is involved in salvaging (*R,S*)-SAM as a source of L-methionine.

Orthologous regulated operons containing COG1912 gene

Properties				
Regulog:	MtaR – Streptococcaceae			
Regulator type:	Transcription factor			
Regulator family:	LysR			
Regulation mode:	repressor			
Biological process:	Methionine metabolism			
Effector:	Homocysteine			
Phylum:	Firmicutes			







Operon	Position	Score	Sequence	Locus Tag of the First Gene
<i>Streptococcus gordonii</i> str. Challis substr. CH1				
 COG1912-mtsA	-30	4.5	TAAAGTTCAGGACTATA	SGO_0468
<i>Streptococcus mitis</i> B6				
 COG1912-mtsA-mtsB-mtsC	-91	4.6	TATAAGGTAAACTATA	smi_1812
<i>Streptococcus mutans</i> UA159				
 COG1912-mtsA-mtsB-mtsC	-110	5.4	TATAGGTAATAACTATA	SMU.1936c
<i>Streptococcus pneumoniae</i> TIGR4				
 COG1912-mtsA-mtsB-mtsC	-64	4.6	TATAAGGTAAACTATA	SP_0481
<i>Streptococcus thermophilus</i> CNRZ1066				
 COG1912-mtsA-mtsB-mtsC	-100	4.5	CATAGCCAAAGGCTATA	str0304
<i>Streptococcus uberis</i> 0140J				
 COG1912-mtsA-mtsB-mtsC	-106	6.3	TATAGTTTACTACTATA	SUB0314

Figure 2.13: Predicted co-regulation of DUF62 encoding genes and energy-coupled factor transporter MtsABC encoding genes. DUF62/COG1912 encoding genes and ECF transporter MtsABC encoding genes in operons predicted to be regulated by MtaR. Data extracted from RegPrecise (http://regprecise.sbpdiscovery.org:8080/WebRegPrecise/ort_operons.jsp?project_id=3620&ort_id=1289134).

2.3 Conclusions

The genomic era brought the hope of a deeper understanding of living systems. However, as the volume of genome sequence data increases, functional annotation of encoded proteins lags far behind. Even the smallest bacterial genomes still contain at least 25% of genes of unknown function (Hutchison *et al.*, 2016). A considerable number of conserved unknown genes are now known to be involved in metabolite damage control and have been referred to as “house cleaning” genes (de Crécy-Lagard *et al.*, 2018). Cofactor damage control is of particular importance due to the intrinsic reactivity of cofactors and their essential metabolic roles (de Crécy-Lagard *et al.*, 2018; Linster *et al.*, 2013).

The sulfonium salt (*S,S*)-SAM (Figure 2.1) is a ubiquitous cofactor essential for many cellular processes (Fontecave *et al.*, 2004). As the primary methyl donor in methyltransferase-catalyzed reactions, (*S,S*)-SAM helps maintain cellular homeostasis (Kako *et al.*, 2019). Cells use dynamic methylation of DNA, RNA and proteins to control the flow of genetic information (Kako *et al.*, 2019), impacting development, cell-fate determination, and adaptability from bacteria to humans (Michalak *et al.*, 2019; Sánchez-Romero & Casadesús, 2019). An imbalance in macromolecule methylation has been implicated in cancer and aging in humans (Michalak *et al.*, 2019), and in phenotypic variations such as decreased virulence and upregulation of stress responses in bacteria (Sánchez-Romero & Casadesús, 2019).

Maintenance of (*S,S*)-SAM homochirality is important for cellular health given that the (*R,S*)-SAM diastereomer is largely inactive as a methyl donor and can function as an inhibitor of methyltransferases (de la Haba *et al.*, 1959; Borchardt & Wu, 1976). We showed here that StDUF62 is a stereoselective (*R,S*)-SAM hydrolase.

2.4 **Experimental Methods**

2.4.1 **Chemicals**

Ammonium formate, ammonium acetate, formic acid, glacial acetic acid, and hydrochloric acid were obtained from Fisher. Trifluoroacetic acid (TFA) was purchased from Alfa Aesar. Isopropyl β -D-1-thiogalactopyranoside (IPTG), potassium phosphate monobasic, potassium phosphate dibasic, HPLC grade water, HPLC grade acetonitrile (MeCN), adenosine, and *S*-(5'-Adenosyl)-L-methionine *p*-toluenesulfonate salt were purchased from Sigma-Aldrich. *S*-Adenosyl-L-methionine disulfate tosylate was purchased from BioVision Incorporated.

2.4.2 **High pressure liquid chromatography**

Diastereomer isolation and enzymatic assay analysis were performed using an Agilent 1260 Infinity HPLC system with a Kinetex 5 μ m C18 column (100Å, 250 \times 4.6mm, Phenomenex). A SecurityGuard ULTRA Column Guard with a C18 (4.6mm ID) cartridge was used for enzymatic assay analysis only.

2.4.2.1 ***S*-adenosyl-L-methionine diastereomer analysis, isolation, and buffer exchange**

Diastereomer isolation was carried out using a previously described (Zhang and Klinman, 2015) system of methods, which utilizes three different buffers under isocratic conditions for SAM diastereomer *Analysis and Quantitation* (50 mM ammonium formate buffer, pH 4.0), *Preparative Isolation* (50 mM ammonium acetate buffer + 1.0% TFA, pH 5.4), and *Buffer Exchange* (10 mM phosphate buffer, pH 7.9). A workflow schematic is shown in Figure 2.14.

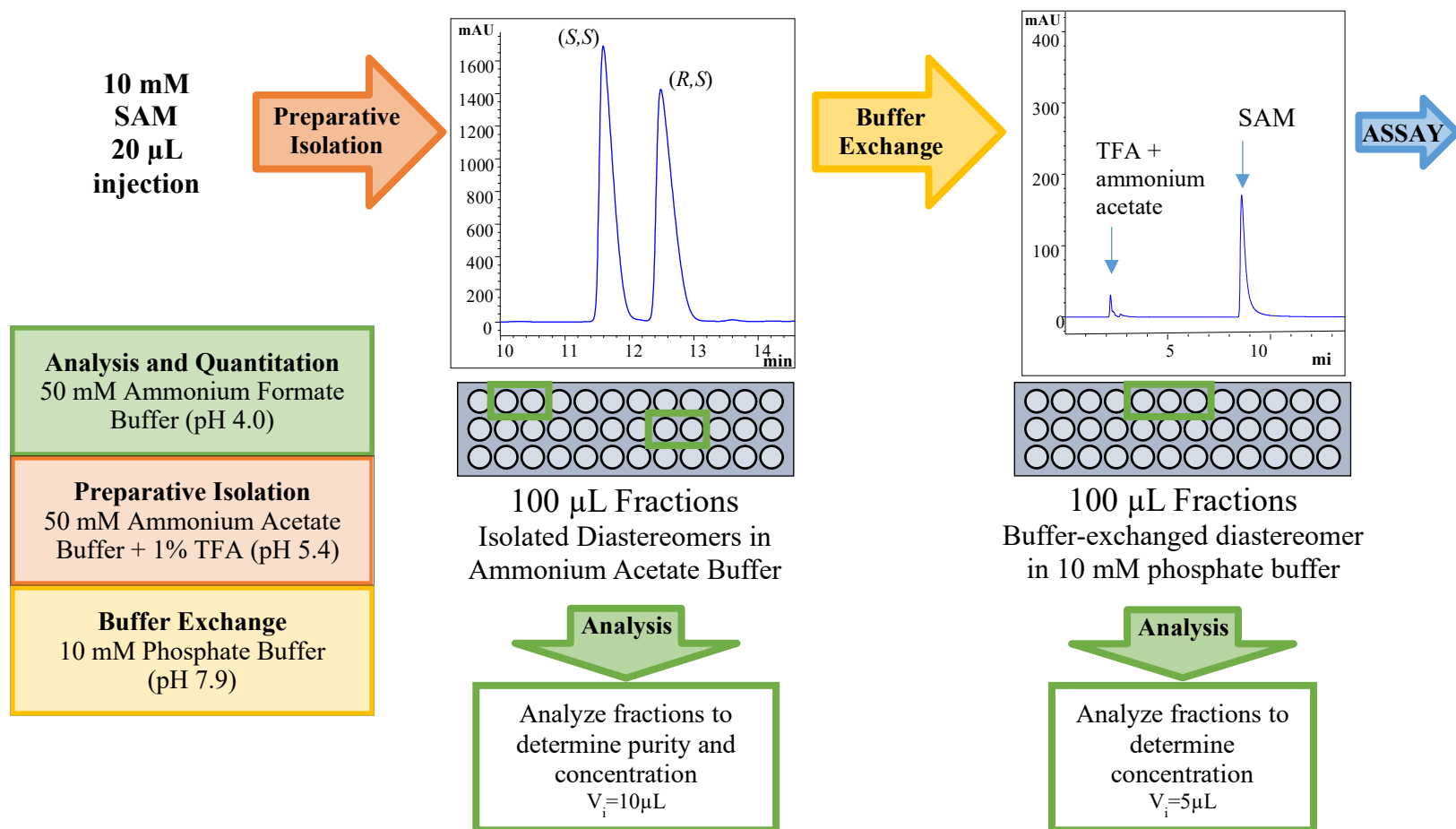


Figure 2.14: S-adenosyl-L-methionine diastereomer isolation workflow

First, 20 μL of 10 mM SAM racemate (BioVision) was injected and run using the *Preparative Isolation* method. Fractions (100 μL) were collected over the course of the eluting diastereomer peaks, which achieved baseline separation. Next, 10 μL of each fraction was analyzed via the *Analysis and Quantitation* method to assess purity and concentration according to a SAM calibration curve that was generated using a racemic standard (TABLE V, Figure 2.15). Those fractions containing pure diastereomer in high concentration were pooled for buffer exchange. Then, 100 μL of isolated diastereomer in ammonium acetate buffer with TFA was injected and run using the *Buffer Exchange* method. Fractions (100 μL) were collected over the course of the eluting peak, yielding isolated SAM diastereomer in phosphate buffer. Finally, 5 μL of each fraction was analyzed using the *Analysis and Quantitation* method to determine concentration according to the SAM calibration curve.

In order for all replicates of assays to include the same concentration points, the disparate tubes of buffer-exchanged SAM were pooled into three different tubes of low, medium, and high concentration (*R,S*)-SAM or into two different tubes of low and high concentration (*S,S*)-SAM. Then, 10 μL from each tube was then analyzed to determine concentration of the pooled sample. Each sample was analyzed in triplicate. Diastereomer concentration was calculated using the SAM calibration curve shown in Figure 2.15. The calculated concentration was then volume corrected as the calibration curve was generated from 50 μL injections. Representative results obtained are shown in TABLE VI and TABLE VII.

TABLE V: *S*-ADENOSYL-L-METHIONINE DIASTEREOMER CALIBRATION CURVE

Total SAM Concentration (μM)	Diastereomer Concentration (μM)	AUC (mAU*s)	
		(<i>S,S</i>)	(<i>R,S</i>)
500	250	6701.4331	7134.9043
100	50	1354.6785	1428.7741
50	25	682.1271	716.9954
10	5	129.1181	138.3402
5	2.5	67.0044	72.0780
2	1	26.1227	28.3722
1	0.5	12.8262	14.4223

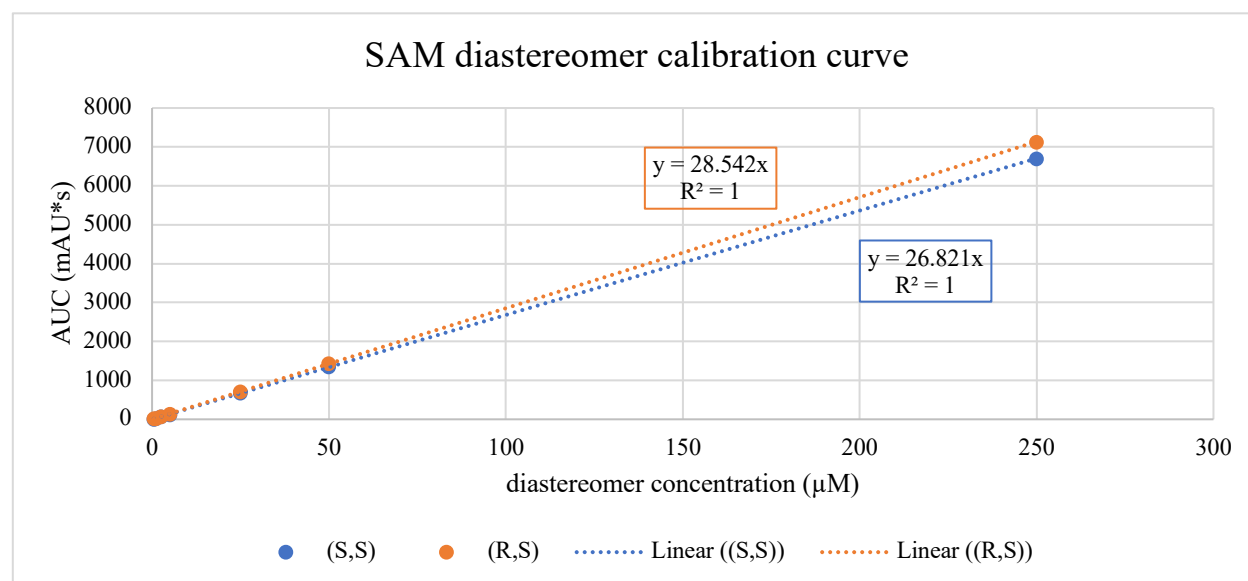
Figure 2.15: *S*-adenosyl-L-methionine diastereomer calibration curve

TABLE VI: CONCENTRATION OF POOLED (R,S)-S-ADENOSYL-L-METHIONINE

(R,S)-SAM conc. group	Injection Volume (μL)	AUC (mAU*s)		Diastereomer Concentration (μM)		Injection Volume Corrected (μM)		Average (μM)	Std. Dev (μM)
		(S,S)	(R,S)	(S,S)	(R,S)	(S,S)	(R,S)	(R,S)	(R,S)
low1	10	0	74.76997	0	2.6	0	13.1	13.1	0.03
low2	10	0	74.46971	0	2.6	0	13.0		
low3	10	0	74.8137	0	2.6	0	13.1		
med1	10	0	140.4971	0	4.9	0	24.6	24.6	0.05
med2	10	0	140.5639	0	4.9	0	24.6		
med3	10	0	140.0887	0	4.9	0	24.5		
high1	10	0	211.6289	0	7.4	0	37.1	37.1	0.13
high2	10	0	210.7271	0	7.4	0	36.9		
high3	10	0	212.2159	0	7.4	0	37.2		

TABLE VII: CONCENTRATION OF POOLED (S,S)-S-ADENOSYL-L-METHIONINE

(S,S)-SAM conc. group	Injection Volume (μL)	AUC (mAU*s)		Diastereomer Concentration (μM)		Injection Volume Corrected (μM)		Average (μM)	Std. Dev (μM)
		(S,S)	(R,S)	(S,S)	(R,S)	(S,S)	(R,S)	(S,S)	(S,S)
low1	10	116.3753	0	4.3	0	21.7	0	21.6	0.05
low2	10	115.8236	0	4.3	0	21.6	0		
low3	10	116.1118	0	4.3	0	21.6	0		
high1	10	229.7706	0	8.6	0	42.8	0	42.5	0.35
high2	10	227.3993	0	8.5	0	42.4	0		
high3	10	226.0085	0	8.4	0	42.1	0		

2.4.2.2 Measuring adenosine formation

Analysis of enzymatic assays was carried out as previously described (Schaffrath *et al.*, 2005). In short, adenosine formation was analyzed using an *Adenosine Quantification* method (50 mM KH_2PO_4 :MeCN in a linear gradient from 95:5 to 80:20 over 20 minutes). An adenosine calibration curve was generated with a range 0-50 μM adenosine. Three sets of calibrant solutions were prepared via independent serial dilutions of adenosine stock solution in 50 mM phosphate buffer (pH 7.9) and analyzed ($V_i = 50 \mu\text{L}$) using the above described *Adenosine Quantification* HPLC method (Figure 2.16).

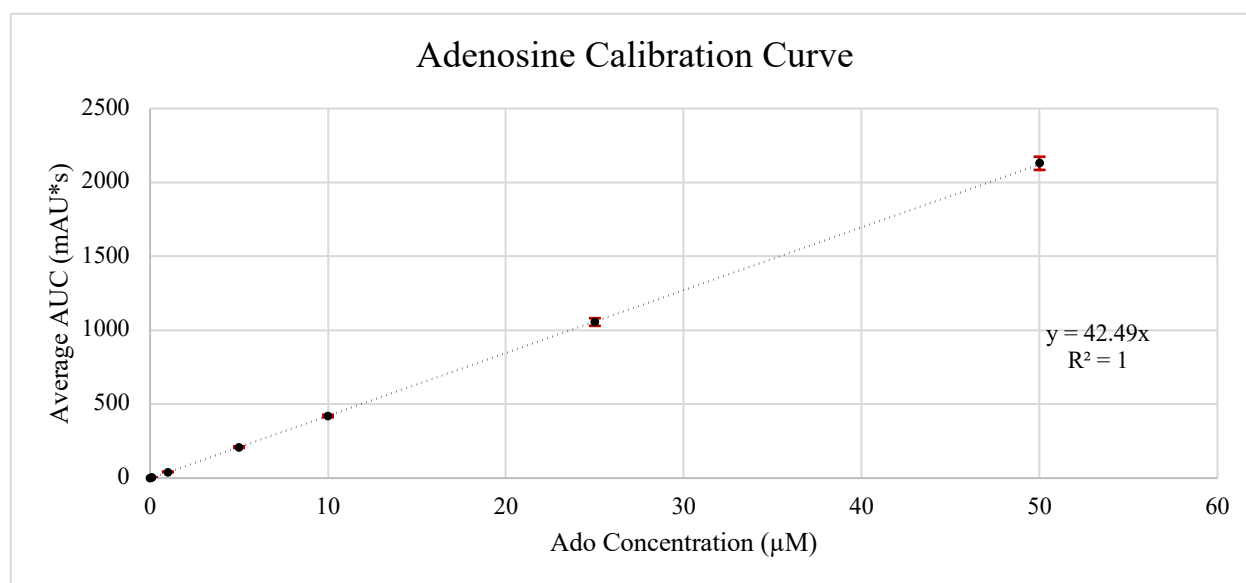


Figure 2.16: Adenosine calibration curve.

2.4.3 Strains, cultivation conditions, and genomic DNA isolation

E. coli DH5 α was used for cloning and cultured in LB broth or agar. *E. coli* BL21(DE3) was used for recombinant protein production after cultivation in Terrific Broth. For genomic DNA isolation, a 50 mL culture of *Salinispora tropica* CNB-440 (Udwary *et al.*, 2007) was grown in a 250 mL flask for 3 days at

30°C and 180 rpm in A1 Marine medium (10g soluble starch, 4 g yeast extract, 2 g peptone, and 36 g instant ocean per liter). Total genomic DNA was isolated using the Sigma GenElute Bacterial Genomic DNA Kit.

2.4.4 Cloning and purification of recombinant StDUF62

The *duf62* gene from *S. tropica* CNB-440 (GenBank accession: WP_011905304) was amplified by PCR from total genomic DNA using the forward primer 5'-CTA TAA GGA TCC ATG GCG CCG ACG CCC TGG ATC-3' (BamHI site underlined) and the reverse primer 5'-TAT ATT CTC GAG TCA ACC GGC GGT GAC GCG CAG-3' (XhoI site underlined). Following digestion with BamHI and XhoI (NEB), the amplicon was ligated into pHis8 (Jez *et al.*, 2000), yielding vector pHis8-St62. pHis8-St62 was constructed using *E. coli* DH5 α Chemically Competent Cells (Invitrogen) and kanamycin (50 μ g/ml) for selection of mutants.

pHis8-St62 was introduced into *E. coli* BL21(DE3) (NEB) by transformation. A 1 L culture was grown in Terrific Broth [24 g yeast extract, 12 g tryptone, 4 mL glycerol, 12.54 g K₂HPO₄, 2.31 g KH₂PO₄ per 1 liter water] containing kanamycin (50 μ g/ml) at 37 °C and 200 rpm, until OD₆₀₀=0.8, at which point gene expression was induced by the addition of IPTG (500 μ M final concentration). Incubation was resumed at a decreased temperature of 20 °C for 18 hours. Cells were harvested by centrifugation and N-terminal His₈-tagged StDUF62 was purified by Ni-NTA (Protino® Ni-NTA agarose, Macherey-Nagel) affinity chromatography. Purification was carried out as previously described (Eustáquio *et al.*, 2008b) but with the following modifications: addition of 1mg/mL lysozyme to the lysis buffer and retention of His₈ tag. The yield was 3 mg per liter of culture. Buffer exchange to 50 mM phosphate buffer (pH 7.9) was performed via gel filtration chromatography using disposable PD-10 columns (GE Healthcare) containing Sephadex G-25M resin. The purified enzyme was analysed by SDS-PAGE using Mini-PROTEAN TGX Precast Gel, 4-20% (Bio-Rad) in Tris/Glycin/SDS Buffer (Bio-Rad). Protein concentration was determined *via* the method of Bradford using the Bio-Rad Protein Assay kit.

2.4.5 in vitro assays

All assays were carried out as 100 μL reactions in 50 mM phosphate buffer (pH 7.9). Preliminary assays were carried out by incubating SAM racemate (1 mM) or isolated SAM diastereomer (0.5-45 μM) with StDUF62 (50-5000 nM) at 30°C for 1 hour. Kinetic constants were determined for StDUF62 with (*R,S*)-SAM by incubating 0-16 μM (*R,S*)-SAM with 50 nM StDUF62 at 30°C for 4 min. Reactions were quenched by freezing in a dry ice/ethanol bath. While thawing on ice, 10 μL 1 N HCl was added to each tube to denature the protein. Protein was precipitated out of the reaction mixture by centrifugation at 10,000 rcf and 2°C for 10 min. 60 μL of assay supernatant was removed for analysis ($V_i = 55 \mu\text{L}$).

Assays of StDUF62 with (*S,S*)-SAM involved both analysis of adenosine formation as well as analysis of SAM diastereomer content. First, enzymatic assays containing 10 μM (*S,S*)-SAM and 0.5 μM StDUF62 were incubated at 30°C for 1 hour and analyzed using the above described *Adenosine Quantification* method to measure adenosine formation ($V_i = 55 \mu\text{L}$). In an effort to determine whether the observed adenosine formation was due to enzyme activity on (*S,S*)-SAM or residual (*R,S*)-SAM contaminants, the isolated (*S,S*)-SAM stock was reanalyzed with the above described *Analysis and Quantitation* method ($V_i = 100 \mu\text{L}$). As no contaminating (*R,S*)-SAM was observed in the stock solution, a 10 μM (*S,S*)-SAM solution (containing no enzyme) was incubated at 30°C for 1 hour and analyzed with the above described *Analysis and Quantitation* method ($V_i = 100 \mu\text{L}$) to determine if 50 nM (*R,S*)-SAM was formed via racemization during the 1 hour incubation. Again, no (*R,S*)-SAM was observed. Notably, these samples also included a large (*S,S*)-SAM peak which exhibits tailing and would easily obscure the 50 nM (*R,S*) peak.

2.4.6 Calculation favoring physiological relevance of observed kinetic constants

Total SAM concentration in bacteria = 180 nmol/mL. Racemization rate constant = $1.8 \times 10^{-6} \text{ s}^{-1}$. Therefore, rate of (*R,S*)-SAM formation = $1.17 \text{ nmol mL}^{-1} \text{ h}^{-1}$. StDUF62 $k_{\text{cat}} = 4.5 \times 10^{-3} \text{ s}^{-1} = 16.2 \text{ h}^{-1}$, *i.e.* 1 nmol of StDUF62 hydrolyzes 16.2 nmol (*R,S*)-SAM per hour. To hydrolyze 1.17 nmol of (*R,S*)-SAM per hour therefore requires 0.072 nmol of StDUF62 which equals 1.9 μg (given MW = 26.6 kDa). Hence a

cellular StDUF62 concentration of $1.9 \mu\text{g mL}^{-1}$ is required to match the rate of (*R,S*)-SAM formation. Total protein concentration in bacteria = 300 mg mL^{-1} , *i.e.* a StDUF62 concentration of 0.0006% of total protein would suffice to deal with (*R,S*)-SAM formation. Such a modest expression level would be fairly typical of metabolite damage-control enzymes (e.g. Hüdig *et al.*, 2015).

2.4.7 **Bioinformatic analyses**

The BLAST tools (Altschul *et al.*, 1997) and resources at NCBI (<http://www.ncbi.nlm.nih.gov/>; NCBI resource coordinators, 2016) were used routinely. Multiple sequence alignments were built using Multalin (Corpet, 1988). Protein family analysis were performed using the Pfam database tools (El-Gebali *et al.*, 2019) and the KEGG orthology tool (Kanehisa *et al.*, 2016).

Analysis of phylogenetic distribution and physical clustering was performed in the SEED database (Overbeek *et al.*, 2014). Results are available in the “AdoMetRepair VDC” subsystem on the public SEED server¹. A representative genome set of ~1500 genomes were chosen based on phylogenetic diversity as previously described (Niehaus *et al.*, 2015). Additional phylogenetic distribution analyses were performed using the AnnoTree tool (<http://annotree.uwaterloo.ca/>; Mendler *et al.*, 2019) using the Release 04-RS89 (19th June 2019) of the Genome Taxonomy DataBase (<https://gtdb.ecogenomic.org/>; Chaumeil *et al.*, 2020). Physical clustering was analyzed with the SEED subsystem coloring tool or the SeedViewer Compare Regions tool (Kanehisa *et al.*, 2016).

The Venny tool was used to compare lists (<https://bioinfogp.cnb.csic.es/tools/venny/>; Oliveros, 2007-2015). Transcription factor Regulog prediction analysis were performed in RegPrecise 3 (<http://regprecise.sbpdiscovery.org:8080/WebRegPrecise/>) Novichkov, *et al.*, 2013).

¹ http://pubseed.theseed.org/SubsysEditor.cgi?page=ShowSpreadsheet&subsystem=AdoMetRepair_VDC

2.5 References Cited

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17), 3389–3402.
- Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., & Rabinowitz, J. D. (2009). Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nature Chemical Biology*, 5(8), 593–599. doi: 10.1038/nchembio.186
- Bentley, R. (2005). Role of sulfur chirality in the chemical processes of biology. *Chemical Society Reviews*, 34(7), 609–624. doi: 10.1039/b418284g
- Blackmond, D. G. (2010). The origin of biological homochirality. *Cold Spring Harbor Perspectives in Biology*, 2, a002147. doi: 10.1101/cshperspect.a002147
- Borchardt, R. T., & Wu, Y. S. (1976). Potential Inhibitors of S-Adenosylmethionine-Dependent Methyltransferases. 5. Role of the Asymmetric Sulfonium Pole in the Enzymatic Binding of S-Adenosyl-L-methionine. *Journal of Medicinal Chemistry*, 19(9), 1099–1103. doi: 10.1021/jm00231a004
- Bradbury, L. M. T., Ziemak, M. J., El Badawi-Sidhu, M., Fiehn, O., & Hanson, A. D. (2014). Plant-driven repurposing of the ancient S-adenosylmethionine repair enzyme homocysteine S-methyltransferase. *Biochemical Journal*, 463(2), 279–286. doi: 10.1042/BJ20140753
- Cantoni, G. L. (1953). S-Adenosylmethionine; a new intermediate formed enzymatically from L-methionine and adenosinetriphosphate. *The Journal of Biological Chemistry*, 204(1), 403–416.
- Cantoni, G. L. (1975). Biological Methylation: Selected Aspects. *Annual Review of Biochemistry*, 44, 435–451.
- Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2020). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, 36(6), 1925–1927. doi: 10.1093/bioinformatics/btz848
- Cornforth, J. W., Reichard, S. A., Talalay, P., Carrell, H. L., & Glusker, J. P. (1977). Determination of the Absolute Configuration at the Sulfonium Center of S-Adenosylmethionine. Correlation with the Absolute Configuration of the Diastereomeric S-Carboxymethyl-(S)-methionine Salts. *Journal of the American Chemical Society*, 99(22), 7292–7300. doi: 10.1021/ja00464a032
- Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research*, 16(22), 10881–10890. doi: 10.1093/nar/16.22.10881
- Dalhoff, C., & Weinhold, E. (2008). S-Adenosyl-L-methionine and Related Compounds. In P. Herdewijn (Ed.), *Modified Nucleosides: in Biochemistry, Biotechnology and Medicine* (pp. 223–247). Weinheim, Germany: WILEY-VCH Verlag GmbH & Co. KGaA.

- de Crécy-Lagard, V., Haas, D., & Hanson, A. D. (2018). Newly-discovered enzymes that function in metabolite damage-control. *Current Opinion in Chemical Biology*, 47C, 101–108. doi: 10.1016/j.cbpa.2018.09.014
- de la Haba, G., Jamieson, G. A., Mudd, S. H., & Richards, H. H. (1959). S-Adenosylmethionine: The Relation of Configuration at the Sulfonium Center to Enzymatic Reactivity. *Journal of the American Chemical Society*, 81, 3975–3980. doi: 10.1021/ja01524a039
- Deng, H., Botting, C. H., Hamilton, J. T. G., Russell, R. J. M., & O'Hagan, D. (2008). S-adenosyl-L-methionine:Hydroxide adenosyltransferase: A SAM enzyme. *Angewandte Chemie - International Edition*, 47(29), 5357–5361. doi: 10.1002/anie.200800794
- Deng, H., McMahon, S. A., Eustáquio, A. S., Moore, B. S., Naismith, J. H., & O'Hagan, D. (2009). Mechanistic insights into water activation in SAM hydroxide adenosyltransferase (duf-62). *ChemBioChem*, 10(15), 2455–2459. doi: 10.1002/cbic.200900369
- Dong, C., Huang, F., Deng, H., Schaffrath, C., Spencer, J. B., O'Hagan, D., & Naismith, J. H. (2004). Crystal structure and mechanism of a bacterial fluorinating enzyme. *Nature*, 427(6974), 561–565. doi: 10.1038/nature02280
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., & Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research*, 47, D427–D432. doi: 10.1093/nar/gky995
- Eustáquio, A. S., Härle, J., Noel, J. P., & Moore, B. S. (2008a). S-Adenosyl-L-Methionine Hydrolase (Adenosine-Forming), a Conserved Bacterial and Archeal Protein Related to SAM-Dependent Halogenases. *ChemBioChem*, 9(14), 2215–2219. doi: 10.1002/cbic.200800341
- Eustáquio, A. S., Pojer, F., Noel, J. P., & Moore, B. S. (2008b). Discovery and characterization of a marine bacterial SAM-dependent chlorinase. *Nature Chemical Biology*, 4(1), 69–74. doi: 10.1038/nchembio.2007.56
- Fontecave, M., Atta, M., & Mulliez, E. (2004). S-adenosylmethionine: Nothing goes to waste. *Trends in Biochemical Sciences*, 29(5), 243–249. doi: 10.1016/j.tibs.2004.03.007
- Halliday, N. M., Hardie, K. R., Williams, P., Winzer, K., & Barrett, D. A. (2010). Quantitative liquid chromatography-tandem mass spectrometry profiling of activated methyl cycle metabolites involved in LuxS-dependent quorum sensing in Escherichia coli. *Analytical Biochemistry*, 403, 20–29. doi: 10.1016/j.ab.2010.04.021
- Hanson, A. D., Pribat, A., Waller, J. C., & de Crécy-Lagard, V. (2010). “Unknown” proteins and “orphans” enzymes: The missing half of the engineering part list - And how to find it. *Biochemical Journal*, 425(1), 1–11. doi: 10.1042/BJ20091328
- Hoffman, J. L. (1986). Chromatographic Analysis of the Chiral and Covalent Instability of S-Adenosyl-L-methionine. *Biochemistry*, 25(15), 4444–4449. doi: 10.1021/bi00363a041

- Hüdig, M., Maier, A., Scherrers, I., Seidel, L., Jansen, E. E. W., Mettler-Altmann, T., Engqvist, M. K. M., & Maurino, V. G. (2015). Plants Possess a Cyclic Mitochondrial Metabolic Pathway similar to the Mammalian Metabolic Repair Mechanism Involving Malate Dehydrogenase and L-2-Hydroxyglutarate Dehydrogenase. *Plant & Cell Physiology*, 56(9), 1820–1830. doi: 10.1093/pcp/pcv108
- Hutchison III, C. A., Chuang, R.-Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., Gill, J., Kannan, K., Karas, B. J., Ma, L., Pelletier, J. F., Qi, Z.-Q., Richter, R. A., Strychalski, E. A., Sun, L., Suzuki, Y., Tsvetanova, B., Wise, K. S., Smith, H. O., Glass, J. I., Merryman, C., Gibson, D. G., & Venter, J. C. (2016). Design and synthesis of a minimal bacterial genome. *Science*, 351(6280). doi: 10.1126/science.aad6253
- Kako, K., Kim, J.-D., & Fukamizu, A. (2019). Emerging impacts of biological methylation on genetic information. *Journal of Biochemistry*, 165(1), 9–18. doi: 10.1093/jb/mvy075
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44, D457–D462. doi: 10.1093/nar/gkv1070
- Linster, C. L., Van Schaftingen, E., & Hanson, A. D. (2013). Metabolite damage and its repair or pre-emption. *Nature Chemical Biology*, 9(2), 72–80. doi: 10.1038/nchembio.1141
- Martin, J. L., & McMillan, F. M. (2002). SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Current Opinion in Structural Biology*, 12, 783–793. doi: 10.1016/S0959-440X(02)00391-3
- Mendler, K., Chen, H., Parks, D. H., Lobb, B., Hug, L. A., & Doxey, A. C. (2019). Annotree: Visualization and exploration of a functionally annotated microbial tree of life. *Nucleic Acids Research*, 47(9), 4442–4448. doi: 10.1093/nar/gkz246
- Michalak, E. M., Burr, M. L., Bannister, A. J., & Dawson, M. A. (2019). The roles of DNA, RNA and histone methylation in ageing and cancer. *Nature Reviews Molecular Cell Biology*, 20(10), 573–589. doi: 10.1038/s41580-019-0143-1
- NCBI Resource Coordinators (2016). Database resources of the National Center for Biotechnology Information. *Nucleic acids research*, 44(D1), D7–D19. doi: <https://doi.org/10.1093/nar/gkv1290>
- Niehaus, T. D., Gerdes, S., Hodge-Hanson, K., Zhukov, A., Cooper, A. J. L., ElBadawi-Sidhu, M., Fiehn, O., Downs, D. M., & Hanson, A. D. (2015). Genomic and experimental evidence for multiple metabolic functions in the RidA/YjgF/YER057c/UK114 (Rid) protein family. *BMC Genomics*, 16(382), 1–14. doi: 10.1186/s12864-015-1584-3
- Novichkov, P. S., Kazakov, A. E., Ravcheev, D. A., Leyn, S. A., Kovaleva, G. Y., Sutormin, R. A., Kazanov, M. D., Riehl, W., Arkin, A. P., Dubchak, I., & Rodionov, D. A. (2013). RegPrecise 3.0 - A resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genomics*, 14(745), 1–12. doi: 10.1186/1471-2164-14-745
- Oliveros, J. C. (2007-2015). Venny. An interactive tool for comparing lists with Venn's diagrams. (<https://bioinfo.gp.cnb.csic.es/tools/venny/index.html>)

- Osterman, A., & Overbeek, R. (2003). Missing genes in metabolic pathways: A comparative genomics approach. *Current Opinion in Chemical Biology*, 7(2), 238–251. doi: 10.1016/S1367-5931(03)00027-9
- Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F., & Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, 42, D206–D214. doi: 10.1093/nar/gkt1226
- Rodionov, D. A., Vitreschak, A. G., Mironov, A. A., & Gelfand, M. S. (2004). Comparative genomics of the methionine metabolism in Gram-positive bacteria: A variety of regulatory systems. *Nucleic Acids Research*, 32(11), 3340–3353. doi: 10.1093/nar/gkh659
- Sánchez-Romero, M. A., & Casadesús, J. (2019). The bacterial epigenome. *Nature Reviews Microbiology*, 18(January), 7–20. doi: 10.1038/s41579-019-0286-2
- Shelver, D., Rajagopal, L., Harris, T. O., & Rubens, C. E. (2003). MtaR, a Regulator of Methionine Transport, Is Critical for Survival of Group B Streptococcus In Vivo. *Journal of Bacteriology*, 185(22), 6592–6599. doi: 10.1128/JB.185.22.6592-6599.2003
- Struck, A.-W., Thompson, M. L., Wong, L. S., & Micklefield, J. (2012). S-Adenosyl-Methionine-Dependent Methyltransferases: Highly Versatile Enzymes in Biocatalysis, Biosynthesis and Other Biotechnological Applications. *ChemBioChem*, 13(18), 2642–2655. doi: 10.1002/cbic.201200556
- Udwary, D. W., Zeigler, L., Asolkar, R. N., Singan, V., Lapidus, A., Fenical, W., Jensen, P. R., & Moore, B. S. (2007). Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proceedings of the National Academy of Sciences of the United States of America*, 104(25), 10376–10381. doi: 10.1073/pnas.0700962104
- Vinci, C. R., & Clarke, S. G. (2007). Recognition of Age-damaged (R,S)-Adenosyl-L-methionine by Two Methyltransferases in the Yeast *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, 282(12), 8604–8612. doi: 10.1074/jbc.M610029200
- Vinci, C. R., & Clarke, S. G. (2010a). Homocysteine methyltransferases Mht1 and Sam4 prevent the accumulation of age-damaged (R,S)-AdoMet in the yeast *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, 285(27), 20526–20531. doi: 10.1074/jbc.M110.113076
- Vinci, C. R., & Clarke, S. G. (2010b). Yeast, Plants, Worms, and Flies Use a Methyltransferase to Metabolize Age-Damaged (R,S)-AdoMet, but What Do Mammals Do? *Rejuvenation Research*, 13(2–3), 362–364. doi: 10.1089/rej.2009.0956
- Wu, S. E., Huskey, W. P., Borchardt, R. T., & Schowen, R. L. (1983). Chiral Instability at Sulfur of S-Adenosylmethionine. *Biochemistry*, 22(12), 2828–2832. doi: 10.1021/bi00281a009
- Zallot, R., Harrison, K. J., Kolaczowski, B., & De Crécy-Lagard, V. (2016). Functional annotations of paralogs: A blessing and a curse. *Life*, 6(3). doi: 10.3390/life6030039

Zhu, X., Robinson, D. A., McEwan, A. R., O'Hagan, D., & Naismith, J. H. (2007). Mechanism of enzymatic fluorination in *Streptomyces cattleya*. *Journal of the American Chemical Society*, 129(47), 14597–14604. doi: 10.1021/ja0731569

Chapter 3
Mechanisms of polyketide biosynthesis and
diversification of polyketide structures via synthase engineering

Reproduced with permission from the Royal Society of Chemistry, Diversification of polyketide structures via synthase engineering. Kornfuehrer, T. and Eustáquio, A.S. (2019)

[*Med. Chem. Commun.*, 2019, 10, 1256-1272, <https://doi.org/10.1039/C9MD00141G>]

3.1 **Introduction**

Polyketides are a structurally diverse group of natural products with equally variable bioactivity. They exhibit intricate structures featuring complex stereochemistry and diverse functionality. Moreover, they provide privileged scaffolds for drug discovery as the structures have evolved to interact with biological systems. As such, many polyketides have been developed into pharmaceuticals such as the antibiotic erythromycin A, the antifungal agent amphotericin B, the immunosuppressant rapamycin, and the antiparasitic avermectins (Figure 3.1). They are biosynthesized by modular type I polyketide synthases and exhibit many characteristic features of the type I polyketide class such as i) carbon backbones exhibiting an alternating oxygenation pattern, ii) macrocyclic structures, iii) elegant stereochemistry, and iv) diverse functionality, including *O*-glycosidic bonds to unusual sugar moieties.

While polyketides have a firmly established role in treating human diseases, polyketide drug discovery and development efforts have waned in recent years, which is on trend with natural products drug discovery as a whole (Newman & Cragg, 2016; Aminov, 2017). The interplay of supply issues (Newman, 2016) and challenges faced by the synthetic chemist seeking to modify or fully synthesize a polyketide scaffold (Paterson & Lam, 2018) have contributed to the deprioritization of natural products research by pharmaceutical companies who opted to redirect their resources to efforts in combinatorial chemistry (Demain, 2014). An overall decline in the rate of drug discovery was observed concomitantly with the deprioritization of natural products research. The intrinsic advantage of natural product scaffolds to interact with biological targets makes it worthwhile to overcome their inherent challenges (Yñígez-Gutierrez & Bachmann, 2019).

Fortuitously, the biosynthetic logic and machinery that guides the formation of polyketides offers the opportunity to revitalize polyketide drug discovery via synthetic biology strategies. Polyketides are diverse members of a common biosynthetic class formed by multifunctional polyketide synthase (PKS) enzymes that combine simple carboxyacyl building blocks, or monomers, in a head-to-tail fashion to form

complex molecular scaffolds (Hertweck, 2009). Compounds such as those shown in Figure 3.1 are biosynthesized by a distinct subclass of PKS enzymes, called modular type I PKSs (Smith & Tsai, 2007).

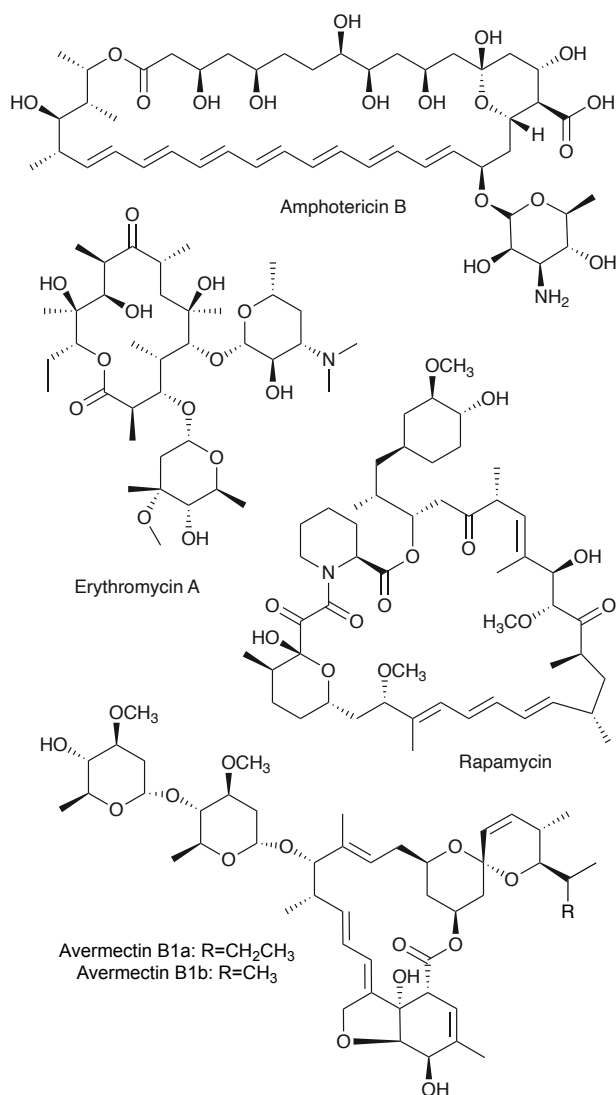


Figure 3.1: Chemical structures of clinically-relevant polyketides.

Modular type I PKS are distinguished from other subclasses of PKS – notably, iterative type I PKS and dissociable type II PKS – by their multimodular functional organization and assembly line-style architecture, which are both characteristics that impart utility for synthetic biology (Khosla *et al.*, 2009). Two evolutionarily distinct subtypes of modular type I PKSs are known and commonly termed *cis*- and *trans*-acyltransferase PKSs. Only polyketides biosynthesized by the prototypical modular, *cis*-acyltransferase PKSs are discussed herein, and the term PKS is used to mean this subtype unless otherwise stated.

This chapter describes how the logic of PKS biosynthesis facilitates the rational engineering of chemical diversity and provides recent examples of polyketide structure modifications achieved via synthetic biology. It presents examples that illustrate the types of structure modifications that can be accomplished. With medicinal chemists in mind, the goal was to focus more on the compounds themselves rather than the synthases. Therefore, details regarding the exact modifications done to the synthases have been deliberately left out. The reader is directed to the original citations for such details, if interested, and to several detailed reviews regarding strategies and challenges for PKS engineering (Bayly & Yadav, 2017; Kalkreuter & Williams, 2018; Klaus & Grninger, 2018; Barajas *et al.*, 2018).

3.2 Rationale for Polyketide Synthetic Biology

The opportunity to utilize PKSs for production of novel chemistry was recognized upon the initial discovery of the catalytic multifunctionality and modular organization of PKS genes involved in the biosynthesis of the macrolide erythromycin (Donadio *et al.*, 1991; Cortes *et al.*, 1990). The PKS enzymatic machinery facilitates access to complex, diverse chemical space from simple monomeric building blocks. Further, their modular organization and assembly line-style architecture allow rational prediction of chemical products from an analysis of genetic sequence – known as the principle of collinearity – although exceptions to the collinearity rule do exist (Moss *et al.*, 2004). Therefore, in principle, it should also be possible to engineer the biosynthetic assembly line to predictably effect a desired chemical change to the final polyketide product.

However, many early attempts to perform “combinatorial biosynthesis” with PKS machinery largely yielded nonfunctional chimeric assembly lines, and those chimeras that were functional suffered from low yields (Menzella *et al.*, 2005). Developments in PKS structural biology over the last decade – including the first reported structure of an intact PKS module (Dutta *et al.*, 2014) and its rearrangements during a catalytic cycle (Whicher *et al.*, 2014) – have dramatically increased our understanding of PKS structure and function. Thus, the field is now poised to achieve a number of targeted modifications to a polyketide scaffold in addition to being able to generate small fragments using highly combinatorial PKS chimeras (Hagen *et al.*, 2016).

The following sections will introduce polyketide synthases including the overall architecture and the enzymatic functionalities – or domains – that impart activity for monomer selection, β -carbon processing, α and β stereochemistry, and macrocyclization. For more detailed information, the reader is directed to excellent reviews of polyketide synthase logic/organization (Hertweck, 2009; Keatinge-Clay, 2017a), structure and mechanisms (Keatinge-Clay, 2012; Weissman, 2015; Robbins *et al.*, 2016), and stereocontrol (Keatinge-Clay, 2016; Weissman, 2017).

3.2.1 **Polyketide synthase architecture**

Representing many of the largest known proteins, PKSs are multifunctional megaenzymes with a modular organization (Weissman, 2015). Each *module* carries out a single round of polyketide chain extension via an enzyme-catalyzed decarboxylative Claisen-type condensation of the growing polyketide chain with a new monomeric extension unit. The modules are comprised of multiple catalytic *domains* that each play a distinct role during an extension cycle. The prototypical PKS, shown in Figure 3.2, is the 6-deoxyerythronolide B synthase (DEBS), which biosynthesizes the aglycone core of erythromycin.

The N-terminal of the PKS assembly line contains a loading module responsible for selecting the first monomer of the polyketide chain, thereby priming the assembly line for successive rounds of chain extension. The basic catalytic domains required of a module for polyketide chain extension are

acyltransferase (AT) for monomer selection, acyl carrier protein (ACP) for substrate shuttling and activation, and ketosynthase (KS) for decarboxylative condensation. In addition to these essential domains, each module may also contain one or more processing domains – ketoreductase (KR), dehydratase (DH), enoylreductase (ER) – which act sequentially to either partially or fully reduce the β -carbonyl functionality of the growing polyketide intermediate. The assembly line also contains a C-terminal thioesterase (TE) domain, which catalyzes release of the polyketide product from the assembly line (Figure 3.2)

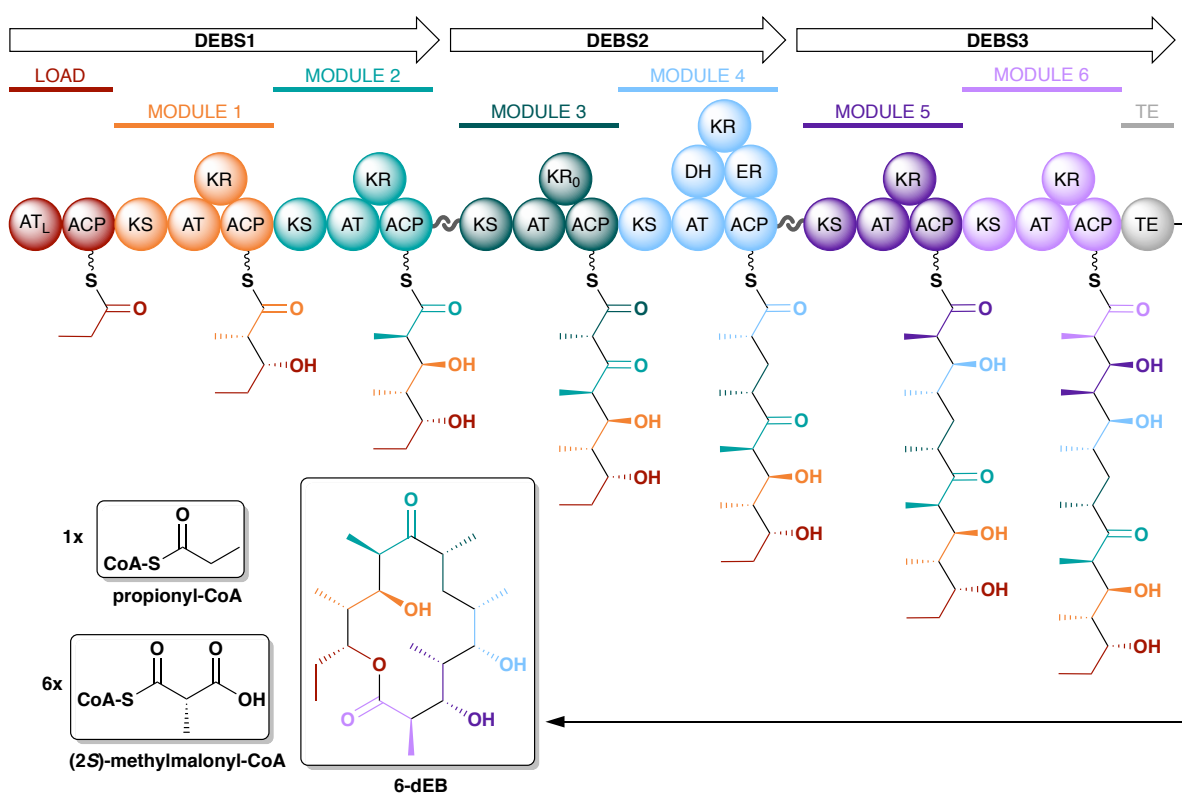


Figure 3.2: Organization of the 6-deoxyerythronolide B synthase.

In addition to the catalytic domains, there are critical structural domains associated with the mediation of protein-protein interactions. Because multiple genes are typically involved in encoding the entire PKS, the assembly line may be made of discrete proteins, termed here *subunits*. For example, DEBS is made of three polypeptide subunits – DEBS1, DEBS2, and DEBS3 (Figure 3.2). Regions on the extreme termini of PKS subunits, called *docking domains*, are responsible for facilitating recognition between the discrete subunits and subsequent self-assembly of the complete PKS assembly line (Broadhurst *et al.*, 2003; Richter *et al.*, 2008). Additionally, within a given subunit there are crucial *linker regions* that not only connect modules on the same protein but also serve to facilitate dynamic interactions between individual catalytic domains, thus maintaining the structural and catalytic integrity of the overall protein (Dutta *et al.*, 2014; Weissman, 2015).

3.2.2 Mechanisms of polyketide biosynthesis

3.2.2.1 Loading

Polyketide biosynthesis is initiated at the N-terminus of the assembly line with selection of a priming monomer, or *starter unit*, by the loading module. Diverse types of domain organizations exist among PKS loading modules, and their variation in enzymatic functionality allows each type to select for monomers of slightly different chemical classes. In the DEBS assembly line, loading is achieved by an AT_L-ACP didomain, though other notable loading module architectures include KS^Q-AT-ACP, A_L-ACP, and GNAT_L-ACP (Keatinge-Clay, 2017a). While their domain organizations and starter unit selectivity may differ, each type of loading module yields an assembly line that is primed with an ACP-bound thioester (Figure 3.3).

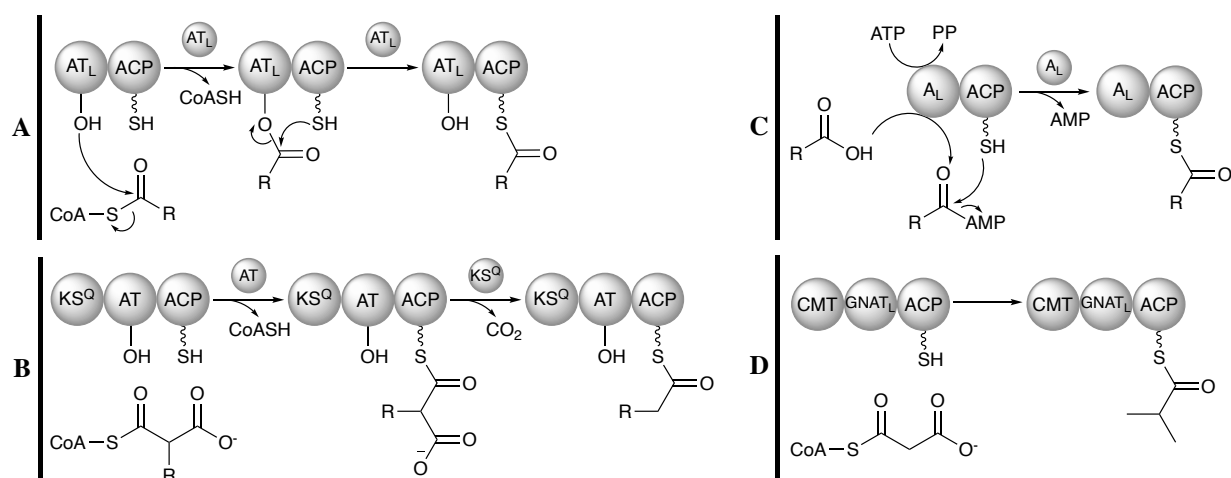


Figure 3.3: Domain architectures found in loading modules of polyketide synthases. (A) AT_L -ACP (B) KS^Q -AT-ACP (C) A_L -ACP (D) $GNAT_L$ -ACP

The didomain structure shown in Figure 3.3A features a loading acyltransferase (AT_L), which primes the assembly line directly by selecting an acylthioester-CoA and charging ACP with the starter unit. Examples include the erythromycin PKS loading module (Figure 3.2) that primes the assembly line with propionyl-CoA and the avermectin PKS loading module that loads isobutyryl-CoA (Lau *et al.*, 2000; Ikeda *et al.*, 1999). Relative to the other loading module architectures, AT_L domains can select for a broad range of acyl-CoA substrates, with some AT_L domains being promiscuous for multiple starter units.

The tridomain structure shown in Figure 3.3B features a condensation-incompetent ketosynthase (KS^Q). The “Q” superscript refers to a highly conserved glutamine residue that replaces the active-site cysteine residue found in KS domains, making KS^Q condensation-incompetent (Bisang *et al.*, 1999). Such KS^Q -AT-ACP modules load carboxyacyl-CoA starter units, as in extension modules. Following selection of either a malonyl-CoA or (2*S*)-methylmalonyl-CoA monomer by AT, the starter unit is transferred to ACP. Then, the KS^Q domain decarboxylates the ACP-bound starter unit to yield either an acetyl group (from malonyl-CoA) or a propionyl group (from (2*S*)-methylmalonyl-CoA) to start the polyketide chain. An example of a PKS containing such a loading module is that of amphotericin B (Caffrey *et al.*, 2001).

Some assembly lines utilize a loading adenylation (A_L) domain that is equivalent to the adenylation domains of non-ribosomal peptide synthase (NRPS) assembly lines (Keatinge-Clay, 2017a). These A_L -ACP modules (Figure 3.3C) can load a free carboxylic acid from the shikimate pathway, as seen in the rapamycin PKS (Lowden *et al.*, 2001). The A_L domain prepares a shikimate-derived carboxylic acid monomer for loading in an ATP-dependent manner and loads the activated monomer to ACP. The biosynthetic nature of the starter units A_L -ACP modules can load – in addition to their accepting free carboxylic acid monomers – make them noteworthy loading modules to the synthetic biologist.

The least common type of loading module utilizes a GCN5 *N*-Acetyltransferase-like ($GNAT_L$) domain. Like KS^Q -AT-ACP modules, $GNAT_L$ -ACP modules load carboxyacyl-CoA starter units to prime the assembly line (Figure 3.3D). However, the $GNAT_L$ domain itself both selects and decarboxylates the starter unit. These $GNAT_L$ domains are rare among PKS assembly lines and have not been extensively studied, so much detail regarding the mechanism remains to be elucidated. Further, these unusual loading modules sometimes contain a *C*-methyltransferase (CMT) which may install one or more methyl groups on the carboxyacyl starter unit. In the example shown in Figure 3.3D, the CMT domain methylates the acetyl-ACP starter unit (from malonyl-CoA) twice at the α -carbon to yield an assembly line primed with isobutyryl-CoA. Such a loading module exists in the gephyronic acid PKS (Young *et al.*, 2013). However, $GNAT_L$ modules may not contain a CMT at all, or they may contain CMTs that exhibit an alternative methylation pattern (Keatinge-Clay, 2017a).

3.2.2.2 Extension

A single round of polyketide chain elongation, or *extension*, in a prototypical module N is shown in Figure 3.4. It begins with translocation of the growing polyketide intermediate from the ACP of upstream module N-1 to the KS of downstream module N, a process called *transacylation*. Once there, the growing intermediate undergoes a KS-catalyzed decarboxylative condensation with a new monomer that is attached to the ACP.

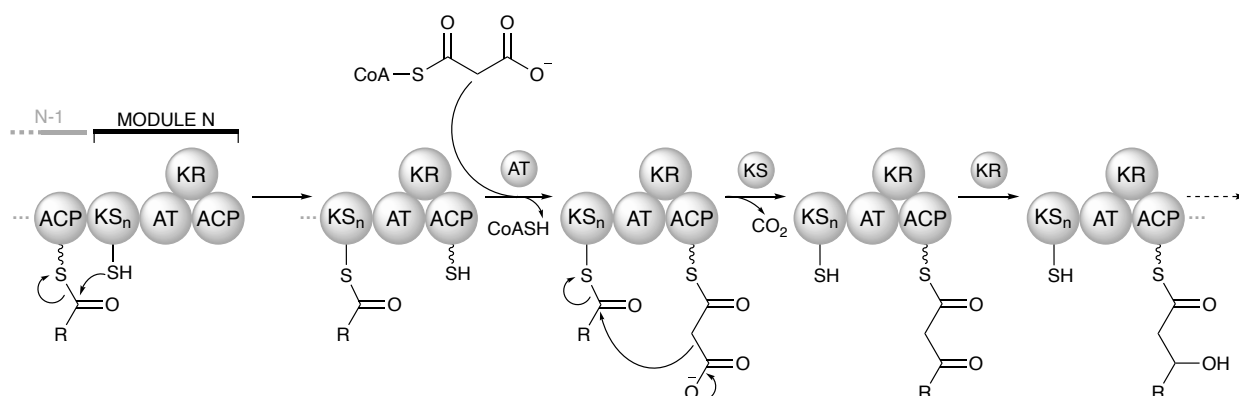


Figure 3.4: Mechanisms of polyketide chain extension in a prototypical module N.

The AT domain acts as a gatekeeper and is responsible for selection of the new monomer, often referred to as an *extender unit*. Mechanistically, acylation of the AT by an extender unit proceeds in a manner analogous to AT_L acylation (Figure 3.3A) by a starter unit. However, the monomer selected by the AT of an extension module is a carboxyacyl-CoA whereas an acyl-CoA starter unit is selected by AT_L. The most commonly used extender units are malonyl-CoA or (2*S*)-methylmalonyl-CoA, although functionalized extender units such as chloroethylmalonyl-CoA, (2*R*)-hydroxymalonyl-ACP, and (2*S*)-aminomalonyl-ACP also exist (Chan *et al.*, 2009; Ray & Moore, 2016).

After the AT selects the appropriate substrate, it loads it onto the downstream ACP, which contains a phosphopantethiene prosthetic group that serves as a flexible tethering point for both the carboxyacyl monomer and the growing polyketide intermediate. The KS domain is responsible for condensing the polyketide intermediate from the upstream module onto the downstream ACP-bound extender unit via a decarboxylative Claisen-type condensation.

Following condensation, the new ACP-bound intermediate may undergo full or partial reductive processing. This is mediated by the optional processing domains KR, DH, and ER – collectively termed the *reductive loop* (even though DH is not an oxidoreductase but, rather, a lyase). The domains of the reductive loop act in sequence to reduce the β-carbonyl functionality on the ACP-bound intermediate to a β-hydroxyl

(KR), to dehydrate the β -hydroxyl product to an α,β -unsaturated alkene (DH), or to reduce the alkene to a saturated alkane (ER). The reductive loop may contain one of four combinations of these domains – none, KR only, KR-DH, or KR-DH-ER. Once the processing by the reductive loop is complete, the ACP-bound intermediate is translocated to either a downstream KS for another round of extension in the next module or, if it is already in the final module, to the terminal TE domain where it is offloaded from the assembly line (Lowry *et al.*, 2016).

3.2.2.3 Offloading

The TE domain is responsible for releasing the polyketide chain from the assembly line, either as a macrocyclic lactone or a linear carboxylic acid. Both mechanisms involve nucleophilic attack of the TE-linked polyketide chain, though lactonization is initiated by hydroxyl functionality on the polyketide chain (Figure 3.5A) whereas a free water molecule is the nucleophile for linear products (Figure 3.5B). The nature of the offloading mechanism employed by a given PKS assembly line is a property of its TE domain (Horsman *et al.*, 2016), and TEs which catalyze lactonization are much more common than TEs that release a linear product via hydrolysis (Keatinge-Clay, 2017a). Atypical offloading mechanisms include two-electron and four-electron reductions to an aldehyde or alcohol, respectively (Du & Lou, 2010). Some assembly lines utilize reductive offloading, but this is more common in NRPS or PKS-NRPS hybrid systems (Du & Lou, 2010).

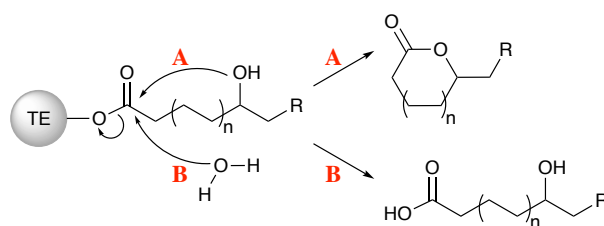


Figure 3.5: Mechanisms of polyketide offloading to release (A) cyclized or (B) linear products.

Those TEs that catalyze lactonization are usually highly selective in terms of ring size generated. For example, erythromycin producers generate only 14-membered macrolactones despite the availability of other intramolecular nucleophiles (Figure 3.2). An exception to ring size selectivity is known for the pikromycin (Pik) PKS in which 12- and 14-membered macrolactones are produced. However, the mechanism for this promiscuity was shown to be module skipping instead of the Pik-TE accepting alternate hydroxyl nucleophiles (Beck *et al.*, 2002). For DEBS-TE, the D-configuration of the C-13 hydroxyl group has been shown to be required for macrocyclization – smaller and larger ring sizes can be engineered only if the D-configuration of the hydroxyl nucleophile is maintained (Horsman *et al.*, 2016).

3.2.3 **Polyketide structure diversification via synthase engineering**

As evidenced from the above sections, the structural diversity of polyketides can be largely attributed to a few biosynthetic variables (Hertweck, 2009). The length of the polyketide chain is primarily determined by the number of modules present (Moss *et al.*, 2004). The selection of monomers to be incorporated into the final scaffold is determined by the gatekeeper AT domain (Chan *et al.*, 2009; Ray & Moore, 2016). Processing of the β -carbon, and the degree to which it is reduced, are governed by the domains of the reductive loop, which also hold significant influence over the complex stereochemistry of the final product (Keatinge-Clay, 2016; Weissman, 2017). Offloading mechanisms mediated by the C-terminal TE domain result in either a linear or cyclized product (Horsman *et al.*, 2016; Du & Lou, 2010). Finally, post-PKS tailoring enzymes can install additional chemical functionality to the polyketide product (Olano *et al.*, 2010).

Given this understanding of how PKS enzymes generate chemical diversity, individual catalytic domains become appealing candidates for modification as a means to alter the structure of the polyketide product formed by the assembly line. As such, domains have been previously described as “the combinatorial units” of PKS combinatorial biosynthesis (Weissman & Leadlay, 2005). The sections below highlight examples of modifications to polyketide structures that can be achieved by domain engineering.

3.3 **Monomer Diversification**

The AT domain exerts significant control over the final structure of the polyketide product by selecting which carboxyacyl monomer will be incorporated into the scaffold. By engineering the AT domain, one can achieve a novel polyketide scaffold architecture via the incorporation of non-native monomers with variable α -substituent functionality during either the loading or extension steps. The two most common monomers utilized in polyketide biosynthesis are malonyl-CoA and (2*S*)-methylmalonyl-CoA. However, there are a variety of more exotic monomers known to be incorporated by modular PKSs – including analogues exhibiting aromatic functionality, branched-chain alkyl substituents, and halogenation. The reader is referred to several comprehensive reviews for details on naturally occurring starter and extender units (Moore & Hertweck, 2002; Chan *et al.*, 2009; Ray & Moore, 2016). Additionally, alternative monomers may be synthetically produced or engineered, allowing site-specific incorporation of designed functionality.

3.3.1 **Acyltransferase engineering**

To date, several strategies have been employed for AT engineering including domain exchange, site-directed mutagenesis, and AT inactivation with cross-complementation (Dunn & Khosla, 2013; Musiol-Kroll & Wohlleben, 2018). Structures of polyketide analogues obtained through AT engineering strategies and the methods used to generate them, which will be covered in the following sections, are shown in Figure 3.6.

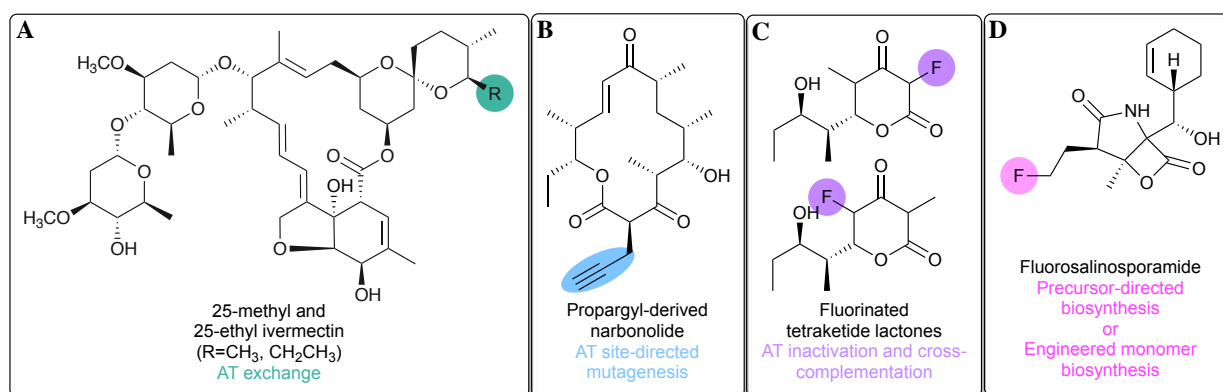


Figure 3.6: Examples of polyketide analogues obtained via monomer diversification and the methods employed.

3.3.1.1 Domain exchange

One of the early examples of AT domain exchange was the replacement of the (2*S*)-methylmalonyl-CoA specific AT domain in module 4 of DEBS with a malonyl-CoA specific AT domain from the rapamycin PKS to afford 6-desmethyl erythromycin D, albeit at low yields of ~0.3 mg/L (Petkovic *et al.*, 2003). A more recent example involved the production of 25-methyl and 25-ethyl ivermectin analogues (Figure 3.6A) at yields of 2 g/L and 0.95 g/L, respectively, by exchanging the loading AT_L-ACP didomain of the avermectin PKS with the loading AT_L-ACP from the milbemycin PKS (Zhang *et al.*, 2015).

These examples show that domain exchange may be used to alter monomer incorporation at either loading or extension modules of PKS assembly lines. Yet, domain exchange has proven to be a particularly complicated feat as many attempts have resulted in structurally and/or catalytically compromised assembly lines, leading to low product yields or no product at all. As we will cover in section 3.6.2, maintaining the integrity of ACP interactions with its cognate enzymes, including AT domains, is crucial in order to generate functional PKS chimeras. Further, systematic investigations by Yuzawa *et al.* have helped identify best regions to “cut and paste” in order to enable AT swaps in the context of small chimeras (Yuzawa *et al.*, 2017).

3.3.1.2 Site-directed mutagenesis

Site-directed mutagenesis provides a route for AT engineering that may be less likely to significantly perturb protein structure and catalytic activity than domain exchange. This technique has been employed to relax substrate specificity of an AT domain, causing promiscuous incorporation of both native and non-native starter/extender units into the polyketide backbone. Thus, the engineered assembly line may produce multiple analogues in a single fermentation broth. One such early example involved mutation of DEBS-AT4, leading to *in vivo* production of a mixture of the natural polyketide, 6-deoxyerythronolide B (6-dEB), and an unnatural analogue, 6-desmethyl-6-dEB, which is formed when DEBS-AT4 incorporates malonyl-CoA rather than methylmalonyl-CoA, its native substrate (Reeves *et al.*, 2001).

Site-directed mutagenesis experiments may be rationally guided by PKS structural information. For instance, one study used molecular modeling to inform targeted mutagenesis of DEBS-AT6 and produce unnatural derivatives including 2-propargylerythromycin A (Bravo-Rodriguez *et al.*, 2015). Moreover, recently obtained structural data of a promiscuous AT, that of SpnD-AT from the splenocin PKS, further informed efforts to broaden the substrate scope of DEBS-AT6 to include larger functional groups such as benzyl substrates (Li *et al.*, 2018). These studies suggest that broadening substrate specificity may prove a useful approach for generating suites of analogues, allowing subsequent investigation into structure-activity relationships. However, it may also be desirable to invert the substrate specificity rather than simply relax it.

Inversion of substrate specificity would allow generation of a single non-natural polyketide analogue with site-specific incorporation of a particular monomer. Recently, inversion of substrate specificity was achieved in the last extension module of DEBS when modification of AT6 via site-directed mutagenesis caused it to preferentially accept propargylmalonyl-CoA over the native substrate methylmalonyl-CoA (Koryakina *et al.*, 2017).

In a similar *in vitro* system, Kalkreuter *et al.* recently employed site-directed mutagenesis to invert substrate specificity in the last two extension modules (5 and 6) of the pikromycin PKS. Both Pik-

AT5 and Pik-AT6 harbor natural specificity for methylmalonyl-CoA with low native promiscuity. The implemented mutations altered the active site such that the AT was able to accommodate malonyl-CoA derivatives with bulkier α -substituents of propargyl (Figure 3.6B), ethyl, allyl or butyl. Importantly, they were able to see single incorporation of non-native extender units by each module as well as double incorporation, where the non-native extender unit was incorporated by both mutated modules in sequence. This is the first time a full-length polyketide has been generated with the incorporation of two non-natural extender units (Kalkreuter *et al.*, 2019).

They also observed that incorporation of bulky side chains by Pik-AT5 results in low efficiency of downstream reductive processing by Pik-KR5. The final round of extension does not occur for intermediates from module 5 that were not reduced by Pik-KR5, generating a truncated polyketide product with additional keto functionality. The authors propose this may be the result of the downstream KS harboring specificity for an intermediate exhibiting β -hydroxyl rather than β -carbonyl functionality. If true, this would suggest that alternative monomer incorporation by a non-terminal module may additionally require that downstream domains be engineered to accept the new substrate (Kalkreuter *et al.*, 2019).

3.3.1.3 Inactivation and cross-complementation

As briefly mentioned in the Introduction, modular type I PKSs are categorized into *cis*-AT and *trans*-AT systems. In *cis*-AT systems, such as DEBS and all other assembly lines discussed in this review thus far, each module contains an AT domain that is part of the megaenzyme structure. In contrast, modules of *trans*-AT systems lack their own AT domain and instead utilize a free-standing, separately encoded AT for substrate selection and loading (Helfrich & Piel, 2016). The existence of *trans*-AT domains allows for generating chemical diversity via inactivation of native *cis*-AT domains and cross-complementation with a heterologous *trans*-AT domain.

This strategy was employed by Walker *et al.* to achieve site-selective incorporation of fluorinated monomers by a truncated PKS consisting of DEBS modules 2-3 and the TE domain (Walker *et al.*, 2013;

Ad *et al.*, 2017). They tested the ability of each of the two DEBS modules to incorporate the fluoromalonyl-CoA monomer *in vitro*. While the fluorinated product was produced (Figure 3.6C), the catalytic efficiency of the system dropped drastically, partially due to substantial hydrolysis of the fluorinated building block by the methylmalonyl-CoA specific AT of DEBS. The authors hypothesized that a malonyl-CoA specific AT may interact more favorably with the fluorinated building block, so they devised a system in which the AT domain of the same DEBS modules were inactivated via site-directed mutagenesis, and the activity of the inactivated domain was complemented with a malonyl-CoA specific *trans*-AT. Doing so significantly increased fluoromalonyl-CoA incorporation by the truncated PKS system (Thuronyi & Chang, 2015). This work shows that it is possible to utilize PKS machinery to incorporate fluorine into a PKS scaffold in a site-specific manner, an application that is particularly attractive to medicinal chemistry (Gillis *et al.*, 2015).

Another example of cross-complementation used the promiscuous KirCII *trans*-AT to introduce allyl- and propargyl functional groups into kirromycin (Musiol-Kroll *et al.*, 2017). Although kirromycin is encoded by a *trans*-AT PKS, one can envision that KirCII has the potential to be applied with *cis*-AT systems in a manner analogous to what was done for site-specific incorporation of fluorine. Introduction of propargyl groups is significant because alkynes can be further derivatized via click chemistry, allowing the integration of synthetic biology and synthetic chemistry for natural product structure diversification.

3.3.2 **Precursor-directed biosynthesis and monomer biosynthesis**

In addition to AT engineering, the introduction of modified monomers into polyketide scaffolds can also be achieved by a process called mutasynthesis or precursor-directed biosynthesis. This involves harnessing the promiscuity of certain AT domains coupled with interruption of the biosynthesis of the natural monomer and supplementation of mutant cultures with an alternate monomer.

For instance, fluorosalinosporamide – a hybrid polyketide-nonribosomal peptide (Figure 3.6D) – was engineered by supplementing a fluorinated precursor to cultures of a mutant blocked in the biosynthesis of chloroethylmalonyl-CoA, the natural polyketide monomer (Eustáquio & Moore, 2008).

Alternatively, fluorosalinosporamide can also be obtained via total biosynthesis by replacing the chlorinase gene involved in native monomer biosynthesis with a fluorinase gene (Eustáquio *et al.*, 2010). The latter was possible because the fluoroacetate pathway from which the fluorinase gene was derived (Chan & O'Hagan, 2012) shares similarities with the chloroethylmalonyl-CoA pathway (Eustáquio *et al.*, 2008; Eustáquio *et al.*, 2009).

3.4 Control of Oxidation State and Stereochemistry

If present in a given module, the optional reductive loop spans from the C-terminus of the AT domain to the N-terminus of the ACP domain and comprises some combination of the processing domains KR, DH, and ER. In addition to determining the oxidation state of the polyketide scaffold, the reductive loop is also the primary determinant of scaffold stereochemistry (Figure 3.7). The KR domain, in particular, is responsible for setting the majority of stereocenters present in a polyketide scaffold, and KR domains are classified based on the stereochemistry of the polyketide intermediate following the KR's action upon it (Keatinge-Clay, 2016).

In the absence of an α -substituent (Figure 3.7A), the KR domain stereoselectively reduces β -ketoacyl intermediates to yield either L- β -hydroxyls (A-type KR) or D- β -hydroxyls (B-type KR). The DH domain introduces alkene functionality into the polyketide scaffold through *syn*-coplanar elimination of the β -hydroxyl group and α -proton (Gay *et al.*, 2013). This occurs stereospecifically such that a substrate with an L- β -hydroxyl group will yield a *cis*-double bond while a substrate with a D- β -hydroxyl group will yield a *trans*-double bond (Labonte & Townsend, 2013; Keatinge-Clay, 2016). Therefore, the stereochemical outcome of a DH-mediated dehydration is tied to the nature of the KR domain that immediately precedes it. The final activity of the reductive loop is carried out by the ER domain, which reduces the α,β -alkene functionally to a fully saturated alkane.

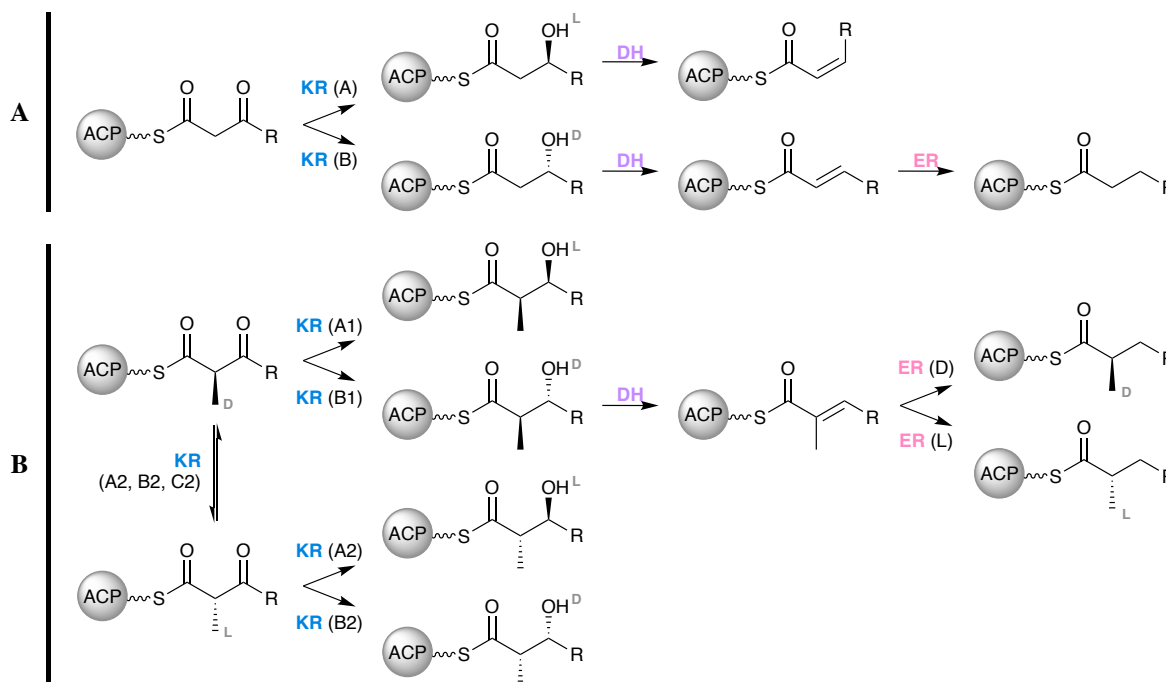


Figure 3.7: Control of oxidation state and stereochemistry effected by domains of the reductive loop when (A) an α -unsubstituted monomer, such as malonate, is added and (B) when an α -substituted monomer, such as (2S)-methylmalonate, is added.

In the absence of an α -substituent (Figure 3.7A), the KR domain stereoselectively reduces β -ketoacyl intermediates to yield either L- β -hydroxyls (A-type KR) or D- β -hydroxyls (B-type KR). The DH domain introduces alkene functionality into the polyketide scaffold through *syn*-coplanar elimination of the β -hydroxyl group and α -proton (Gay *et al.*, 2013). This occurs stereospecifically such that a substrate with an L- β -hydroxyl group will yield a *cis*-double bond while a substrate with a D- β -hydroxyl group will yield a *trans*-double bond (Labonte & Townsend, 2013; Keatinge-Clay, 2016). Therefore, the stereochemical outcome of a DH-mediated dehydration is tied to the nature of the KR domain that immediately precedes it. The final activity of the reductive loop is carried out by the ER domain, which reduces the α,β -alkene functionally to a fully saturated alkane.

The stereocontrol exerted by the reductive loop becomes more nuanced when there is an α -substituent present, which necessarily creates a chiral center at the α -carbon of the polyketide

intermediate (Figure 3.7B). In addition to their stereoselective reduction of β -carbonyl groups, KR domains also exhibit stereospecificity by reducing substrates with either a D- α -substituent (A1- and B1-type KRs) or those with an L- α -substituent (A2- and B2-type KRs). The A2- and B2-type KRs exhibit additional epimerase activity that allows inversion of α -substituent orientation from D to L prior to reduction of the carbonyl. There also exist redox-incompetent KR domains – C2-type KRs – that exhibit this same epimerase activity when presented with a D- α -substituted intermediate, thereby yielding an L- α -substituted- β -ketoacyl intermediate. Such a KR can be seen in the third extension module of DEBS, indicated by notation KR₀. The final reductive domain, ER, also has the ability to exert stereocontrol over α -substituted intermediates. When presented with a *trans*- α,β -unsaturated intermediate possessing an α -substituent, an L-type ER will yield an L- α -substituent, and a D-type ER will yield a D- α -substituent (Keatinge-Clay, 2016).

3.4.1 **Ketoreductase engineering**

3.4.1.1 **Site-directed mutagenesis**

As with other domains of the reductive loop, the KR can be inactivated via site-directed mutagenesis to block reduction, thereby altering the oxidation state of the polyketide backbone. One such example achieved inactivation of KR domains in the amphotericin PKS (Power *et al.*, 2008). The authors identified the KR domains of modules 12 and 16 as appropriate candidates for modification 1) by eliminating KR domains whose action is known to be essential for downstream biosynthetic steps or pharmacological activity, and 2) because the importance of C-7 (KR16) and C-15 (KR12) to the structure-activity relationship of amphotericin (Figure 3.8A) had not yet been investigated.

Each of the two KR domains was successfully inactivated, affording production of several analogues that retained antifungal activity but exhibited additional carbonyl functionality on the polyketide backbone. Notably, inactivation of KR16 generated the analogue 7-oxo-amphotericin (Figure 3.8B), which showed reduced hemolytic activity and has, therefore, the potential to improve the therapeutic window of

this clinically relevant drug. Further, the authors argue that addition of carbonyl functionality serves to prime the structures for further modifications via chemical synthesis, such as modifications to increase polarity and solubility.

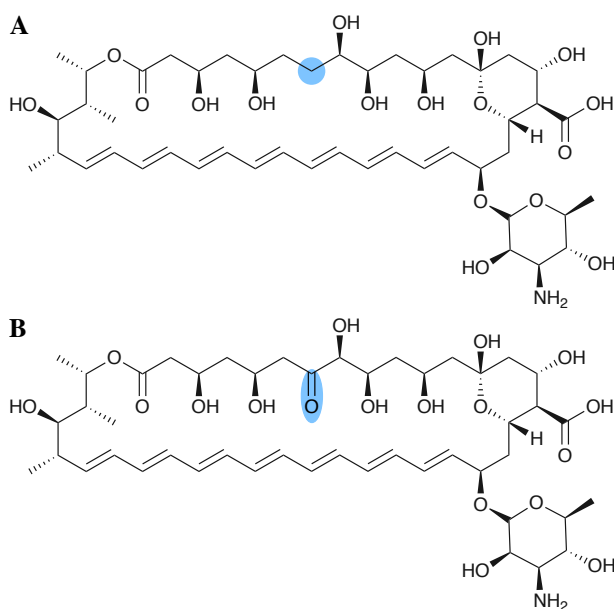


Figure 3.8: Alteration of oxidation state in the amphotericin B backbone via ketoreductase engineering. Structures of (A) amphotericin B and (B) 7-oxo-amphotericin, a novel analogue

3.4.1.2 Domain exchange

Most efforts in KR engineering have focused on understanding how to harness the exceptional stereocontrol exhibited by this domain. One early study on the mechanism of stereocontrol by KR domains identified several active site residues that, when modified, altered stereochemistry of the product *in vitro* (Baerga-Ortiz *et al.*, 2006). However, incorporation of these same mutations did not result in the desired stereochemistry when the KR was part of an intact PKS *in vivo*, indicating gaps in our understanding of the

molecular mechanism for KR stereocontrol (Kwan *et al.*, 2011). Consequently, domain exchange is the most commonly employed method to change stereochemical outcome. However, domain exchange is still not a perfectly refined approach, and much of the ongoing research utilizes truncated PKS model systems to aid in devising methods and principles for predictably altering stereochemical outcomes via KR exchange.

One such method involves use of polylinkers to achieve swaps of an entire reductive loop (Kellenberger *et al.*, 2008). Recognizing the common evolutionary origins of the domains in the reductive loop and, consequently, the need to preserve the intimate protein-protein interactions between them, Kellenberger *et al.* devised a polylinker-mediated approach for systematic engineering and evaluation of reductive loop swaps in the truncated PKS model system DEBS1-TE (the first subunit of DEBS fused to the C-terminal TE from DEBS3). By employing this approach *in vitro*, it was possible to alter both oxidation state and stereochemistry of the intermediate β -hydroxyl group via exchange of reductive loop domains in the second module of DEBS.

This same method was used some years later by another group, who used the same model system to probe KR stereocontrol over α -substituents (Annaval *et al.*, 2015). The authors sought to find a donor KR that would afford epimerized α -substituents in the second module of DEBS, which natively harbors and A1-type KR. They tested a donor pool comprised of both A2- and B2-type KRs, which are all known to exhibit epimerase activity in their native context. Not all chimeras were active, but all of the active chimeras generated hydroxyl configuration consistent with the donor KR type in a new context. This validated that direction of ketoreduction is a fundamental property of the KR active site. However, almost half of the donor KRs did not switch α -methyl configuration. This result indicates that the PKS context of the KR affects activity and that we still lack a detailed understanding of the molecular basis for epimerization.

In a recent advance towards fully realizing the ability to alter stereochemistry via KR domain swap, Eng *et al.* were able to affect a stereochemical change from *anti*- to *syn*- via a reductive loop swap. This study utilized the truncated PKS model system LipPKS1-TE (the first module of the lipomycin PKS

fused to the TE from DEBS), and the authors replaced the native A2-type KR with a panel of donor A1-type KRs. In all cases, the stereochemistry was altered from *anti*- to *syn*- via epimerization of the α -substituent. Further, the native module contains a dimerization element (DE), a structural motif essential for catalytic activity. They found that activity is retained when the native DE is retained but attenuated when the native DE is replaced with a donor DE (Eng *et al.*, 2016). This is a significant finding for the advancement of KR domain swapping as this DE is present in approximately half of PKS modules where the KR is the sole domain in the reductive loop (Zheng *et al.*, 2013).

3.4.2 **Dehydratase engineering**

3.4.2.1 **Site-directed mutagenesis**

The DH has not been extensively explored for engineering relative to other PKS domains, but examples of targeted DH engineering are present in the literature. Mostly, site-directed mutagenesis is employed to inactivate DH functionality. One such study by Kim *et al.* inactivated the DH domain in the first module of the PKS responsible for biosynthesis of geldanamycin (Figure 4.9A), which shows anticancer activity. Further, one of the post-PKS oxidative tailoring enzymes was disrupted in order to prevent installation of quinone functionality, which is thought to be associated with dose-limiting hepatotoxicity of the natural product. The combinatorial modifications made to the biosynthetic pathway led to *in vivo* production of a suite of analogues, all featuring the desired hydroxylation at C15 but varying slightly in their extent of post-PKS tailoring (Kim *et al.*, 2009).

One novel analogue, DHQ3, was detected at titers comparable to the production of geldanamycin in the wild type strain. In addition to retaining the expected hydroxylation at C15 and lacking quinone functionality, DHQ3 also lacks *O*-methylation of the aromatic ring moiety and an alkene group on the polyketide backbone, both of which are installed during post-PKS tailoring steps (Figure 3.9B). Significantly, DHQ3 exhibited 4.6-fold higher biological activity than geldanamycin and is the only C15 hydroxylated analogue to show any significant activity at all (Kim *et al.*, 2009). By comparing the structural

differences between the analogues, the authors were able to determine the structural features likely responsible for the enhanced activity, showing the ability of PKS engineering to produce novel chemistry that may help tease out structure-activity relationships.

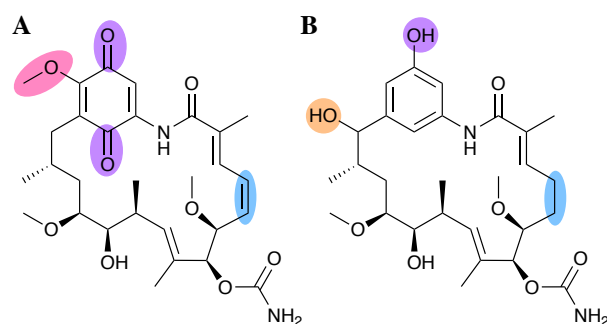


Figure 3.9: Alteration of oxidation state in the geldanamycin backbone via dehydratase engineering and disruption of post-assembly line tailoring. Structures of (A) Geldanamycin and (B) DHQ3, a novel analogue

3.4.2.2 Domain exchange

Exchange of a DH domain played a key role in a recent study by Hagen *et al.*, which sought to achieve *in vitro* production of the commodity chemical adipic acid via engineered PKS machinery. The authors systematically performed reductive loop swaps in the first extension module of the borrelidin PKS, which contains only a KR domain for β -carbon processing. They utilized a variety of donor loops containing the full set of processing domains – KR, DH, and ER. However, they found that processing by the reductive loop stalled at the DH domain in each case (Hagen *et al.*, 2016). Presumably, none of the DH domains in any of the donor loops were capable of processing the unnatural substrate – succinyl-ACP – which features a terminal carboxylic acid group that is atypical of polyketide intermediates.

As DH domains have been previously reported to display substrate specificity (Gay *et al.*, 2013), Hagen *et al.* sought to identify a DH domain known to process a substrate with a carboxylic acid group. One such domain was found in the second extension module of the borrelidin PKS. In an exciting combinatorial feat, the authors replaced the DH domain of the already chimeric modules with the DH domain from the second module of the borrelidin PKS. Finally, the TE from the DEBS assembly line was attached to the end, allowing release of adipic acid from the chimeric PKS (Hagen *et al.*, 2016).

3.4.3 Enoylreductase engineering

3.4.3.1 Site-directed mutagenesis

Early efforts to engineer the ER domain were focused on altering the stereochemical outcome of the reduction when a methyl branch is present as an α -substituent. One study revealed a point-mutation (Y52V) in the active site of an erythromycin ER that is able to effect inversion of methyl branch orientation *in vivo* (Kwan *et al.*, 2008). However, the reverse mutation (V52Y) in an ER shown to have the opposite stereospecificity (rapamycin ER of module 13) did not alter the orientation of the methyl branch, indicating an incomplete understanding of the molecular mechanisms of ER-mediated reduction. Recent efforts to engineer ER domains have focused on reductive loop swaps or using site-directed mutagenesis to inactivate the ER domain.

3.4.3.2 Domain exchange

Two related examples center on affording *in vivo* production of ivermectin, which is a clinically relevant derivative of avermectin, an antiparasitic agent. Avermectin features a double bond at C22-C23, but this is reduced to an alkane in ivermectin (Figure 3.10). Ivermectin is currently produced semi-synthetically via reduction of the natural product with Wilkinson's catalyst (Chabala *et al.*, 1980; Ōmura & Crump, 2004).

The first successfully engineered *in vivo* production of ivermectin was reported in 2006. The relevant reductive loop from the second extension module of the avermectin PKS (aveDH2-KR2) was replaced with a fully-reducing loop from the fourth module of the pikromycin PKS (pikDH4-KR4-ER4). However, the engineered ivermectin producer suffered from low yields (1-3% of parent) that precluded it from commercial viability (Zhang *et al.*, 2006). More recently, a separate group attempted to perform a similar reductive loop swap to afford engineered ivermectin production. The authors replaced the aveDH2-KR2 motif with the milbemycin reductive loop milDH2-KR2-ER2. Excitingly, this led to *in vivo* production of ivermectin B1 in high yield nearly equivalent to the yield of avermectin in the parent strain (Zhang *et al.*, 2015).

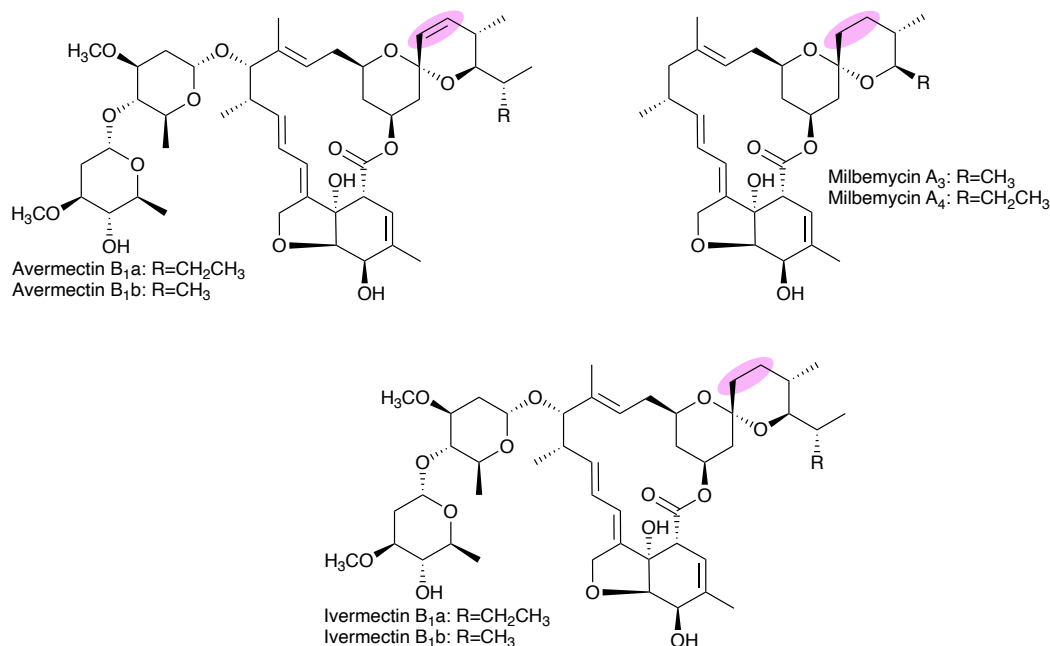


Figure 3.10: Structures of avermectins B1, ivermectins B1, and milbemycins A.

Milbemycin, also an antiparasitic agent, has a polyketide scaffold that is strikingly similar to avermectin (Figure 3.10). Furthermore, the donor milbemycin reductive loop shows higher sequence similarity (51.50%) to the avermectin reductive loop than the pikromycin donor loop (46.48%) used previously (Zhang *et al.*, 2015). Both sequence similarity of donor and acceptor modules and structural similarity of their native intermediate compounds are major factors to consider when exchanging PKS domains. While the original group did take sequence similarity of the donor loop into consideration, they were limited at that time by the available PKS sequence information. The sequence of the characterized milbemycin PKS was not reported until 2010 (Wang *et al.*, 2010). This example highlights the strides we have made in PKS engineering and how increased PKS sequence knowledge can contribute to engineering efforts.

3.5 Engineering Cyclized or Linear Products

Located on the C-terminus of the PKS, the TE catalyzes offloading of the nascent polyketide product from the assembly line. Therefore, the TE domain is the final biosynthetic determinant originating in the PKS machinery that is capable of altering the primary core structure of the polyketide product (Hertweck, 2009). There are two principal mechanisms for TE-mediated offloading – hydrolysis to generate a linear carboxylic acid product and cyclization to generate a macrocyclic lactone or lactam (Figure 3.5).

The linear polyketide tautomycin (Figure 3.11A) has gained attention for its potent immunosuppressant activity and novel mechanism of action (Choi *et al.*, 2017). Investigation of its biosynthetic pathway resulted in the first high-resolution X-ray crystal structure of a TE domain responsible for offloading of a linear polyketide product (Scaglione *et al.*, 2010). This structure revealed that the tautomycin TE contains a narrow substrate tunnel, which significantly constricts the substrate and precludes macrocyclization of the untailored polyketide product (Figure 3.11B).

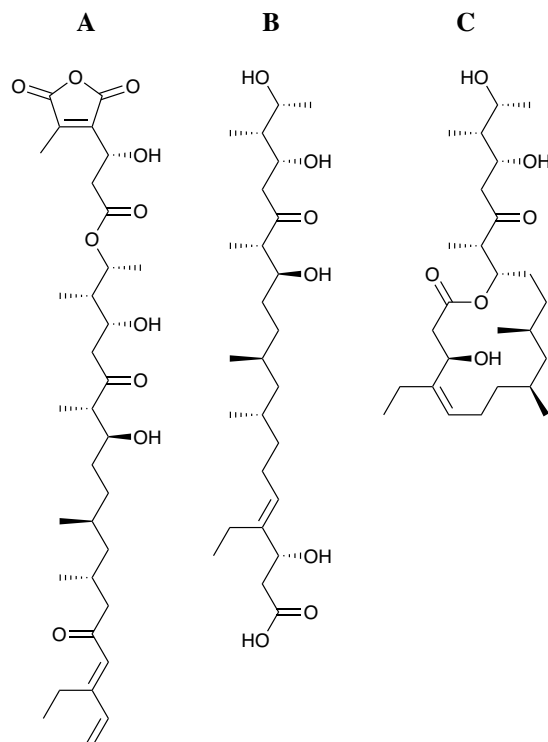


Figure 3.11: Effecting macrocyclization of the tautomycetin via thioesterase domain exchange. Structures of (A) tautomycetin in its mature, bioactive form, (B) the untailored, linear product of the tautomycetin PKS, and (C) the novel macrocyclic tautomycetin analogue.

The native tautomycetin TE was exchanged for the pikromycin TE, which catalyzes macrocyclization, in an effort to generate a macrocyclic analogue of tautomycetin *in vivo* (Tripathi *et al.*, 2016). In contrast to the tautomycetin TE, the pikromycin TE has a large substrate binding tunnel capable of accommodating both 12- and 14-membered rings (Scaglione *et al.*, 2010). The chimeric PKS generated by Tripathi *et al.* was able to produce a 14-membered macrocyclic analogue of tautomycetin (Figure 3.11C), proving the utility of TE domain engineering for the production of novel chemistry. However, only traces of the cyclic analogue were observed, and yields were not quantified (Tripathi *et al.*, 2016). This indicates that TE engineering still requires refinement as a synthetic biology strategy in order to become practical.

In fact, recent reports based on the pikromycin PKS suggest that TE domains may have limited substrate flexibility. One such study showed that the wild-type pikromycin TE has drastically reduced

activity with unnatural substrates containing epimerized chiral centers, failing to produce macrocycles (Hansen *et al.*, 2017). Yet, another study demonstrated that TE site-directed mutagenesis can be used to increase catalytic efficiency and broaden substrate scope, enabling effective cyclization (Koch *et al.*, 2017). The latter study found that mutation of the catalytic serine to a cysteine improves reaction kinetics. It does so by changing the reaction mechanism from a step-wise addition-elimination to a concerted acyl substitution, as predicted from quantum mechanical calculations (Koch *et al.*, 2017). These studies highlight that the nature of the TE should also be considered when attempting AT or reductive loop engineering to modify the polyketide scaffold.

3.6 The Present and Future of Polyketide Structure Diversification

The seminal discovery that polyketide macrolides such as erythromycin are biosynthesized by uncharacteristically large, multifunctional enzymes (Cortes *et al.*, 1990; Donadio *et al.*, 1991) has inspired chemists and biologists alike to harness this biosynthetic modularity to diversify polyketide structures. Although initial attempts to perform large-scale, combinatorial biosynthesis by module shuffling failed due to complete loss or drastic reduction in product yields, a greater understanding of PKS structure and function and of PKS evolution have vitally informed recent efforts that led to the production of polyketide derivatives and small polyketide fragments in acceptable yields.

Further developments are still necessary to make synthetic biology of PKS assembly lines a reliable tool for generating structural diversity in polyketide chemical space. This includes a) continued efforts by natural products researchers to characterize PKS biosynthetic gene clusters; b) additional understanding of the structure and function of PKS enzymes; and c) development of empirically-based engineering principles to better guide PKS engineering attempts.

3.6.1 The continued role of natural products research

Microbial natural products have long been known to exhibit remarkable bioactivity, and decades of sustained efforts by natural products researchers have led to the discovery of diverse natural product

scaffolds. Further, the advances of the genomics era have given us privileged access to the complex synthetic strategies employed by Nature to generate these compounds. The development of bioinformatic tools and databases such as antiSMASH (Medema *et al.*, 2011), MIBiG (Medema *et al.*, 2015), and others (Medema, 2018), have created a space for natural products researchers from around the globe to contribute and analyze their sequence data.

Recently, the number of microbial genomes being sequenced has been increasing at an exponential rate, facilitating discovery of previously unknown natural products via genome mining strategies (Ziemert *et al.*, 2016). It is important for natural products researchers to continue to probe cryptic PKS clusters uncovered by genome sequencing (*i.e.* those PKS clusters not yet connected to a characterized polyketide product). In order for PKS machinery to be useful to the synthetic biologist, it is essential for the PKS genes to be characterized, well-annotated, and associated with their cognate natural product.

Recent estimates indicate that only 10% of the known polyketide structures have been associated with their biosynthetic gene cluster (Dejong *et al.*, 2016). Therefore, connecting cryptic PKS gene clusters to the compounds they encode and continued efforts to characterize PKSs will contribute to expanding the PKS toolkit. The ivermectin example presented here (Figure 3.10) highlights the importance of continued efforts in PKS and polyketide discovery, since availability of sequence information for the milbemycin PKS was crucial for enabling engineered ivermectin production in commercially viable yields (Zhang *et al.*, 2015).

3.6.2 The emergent landscape of polyketide synthase engineering

3.6.2.1 A developing synthetic biology toolkit

Given the supposed modularity of polyketide synthases, one can envision, for instance, starting from a compound of interest and assembling a polyketide synthase from known “parts/devices” (domains/modules) that can catalyze biosynthesis of the target compound (Yuzawa *et al.*, 2018). Such polyketide synthetic biology experiments currently rely on “design-build-test-learn” cycles, as does any

other synthetic biology effort. Computational tools help with design of engineered constructs, such as the recently developed ClusterCAD (Eng *et al.*, 2018), which facilitates the identification of PKS domains and modules that can be used to generate a compound of interest *de novo* (Hagen *et al.*, 2016), or that can be used to modify a given polyketide scaffold. Engineered polyketide synthases are then built and tested for activity. The more variants are tested, the more we can learn regarding not only which parts to use but also how to connect the parts to produce catalytically active chimeric PKS assembly lines. These kinds of studies are critical for development of reliable principles for PKS engineering.

For example, recent efforts by Yuzawa *et al.* have helped to better identify well-conserved regions of the AT domain and surrounding linker region that serve as “cut-paste” sites for a donor AT domain (Yuzawa *et al.*, 2017). The authors created several variants of a single chimeric module – differing in the boundaries used for the module exchange – then systematically analyzed the variants of this system. This led to identification of a putatively optimized domain boundary that maintained both protein stability and catalytic activity. Importantly, this domain boundary is highly conserved across most PKS modules, meaning that their findings should facilitate AT domain replacements in systems beyond their model. Indeed, tests of this method on a second model system indicate that the identified domain boundaries are likely generalizable (Yuzawa *et al.*, 2017). These results can inform future efforts to perform AT domain exchanges as a method of generating novel polyketide chemistry.

Such insights into AT domain exchanges may provide guidance on how to exchange other catalytic domains, such as those of the reductive loop. While controlling stereochemistry is arguably one of the most exciting engineering outcomes, it has proven to be one of the most challenging tasks. For instance, the KR domain has been described as the “stereochemical workhorse” (Keatinge-Clay, 2016) of PKS assembly lines due to the exquisite stereocontrol it can exert over polyketide scaffolds (Figure 3.7). However, KR domain swaps to achieve stereochemical changes have been demonstrated only in truncated model systems thus far, generally with mixed results (Kellenberger *et al.*, 2008; AnnaVal *et al.*, 2015). The

ability to control the stereochemical outcome of complex polyketides in a predictable way – although a welcoming addition to the PKS toolkit – remains outstanding and requires significant continued research.

3.6.2.2 Synthase structure and function

Most early efforts in PKS synthetic biology failed due to an incomplete understanding of the structure and function of PKS megaenzymes. Much work has been done in the last decade to address this gap in our knowledge, and development of techniques such as electron cryo-microscopy have led to significant increases in available structural information, including the first structure of an intact PKS module (Dutta *et al.*, 2014).

This expanded understanding of PKS structure has also elucidated the importance of protein-protein interactions for not only the structural integrity of the assembly line but also the catalytic integrity (Dodge *et al.*, 2018). Additional structural information, including structures of PKS domains/modules at various points in their catalytic cycle, will help to further elucidate which regions of PKS assembly lines are more or less susceptible to perturbation during engineering efforts and how to avoid any issues that may compromise the integrity of the assembly line.

Moreover, the multimodular architecture of PKSs raises an interesting question regarding how synthetic progress along the assembly line is achieved. Mechanisms that prevent a module from being used more than once during biosynthesis of one polyketide molecule must be in place because, although there are exceptions (Sugimoto *et al.*, 2015), PKS modules are not used iteratively. In other words, a mechanism that prevents the just-generated product from being acted upon again by the same module, and instead to be translocated to the correct downstream module must exist. Understanding how one-directional, or so-called vectorial biosynthesis is controlled is important for PKS engineering purposes as it allows directing biosynthesis towards translocation. While the molecular underpinnings of vectorial control remain to be fully elucidated, models have been proposed that include kinetic and structural elements (Klaus & Grninger, 2018).

Protein-protein interactions are expected to play a crucial role in vectorial control. For instance, the ACP domain has received much attention due to its central role in shuttling the polyketide intermediate not only to downstream modules but also to processing domains within a module (Whicher *et al.*, 2014). Thus, the ACP domain is known to contain a variety of elements necessary for recognition by its cognate partners. While the ACP domain has not been considered as an engineering target in this review due to its lack of influence on chemical structure, it should be noted that maintaining interactions between an ACP and its cognate partners is of critical importance in generating functional, engineered assembly lines (Chandran *et al.*, 2006; Kapur *et al.*, 2010).

3.6.2.3 **Evolution of polyketide synthases**

In addition to structure/function studies on PKS machinery and systematic construction of chimeric PKSs, it is essential to study how these assembly lines evolved. Insights on PKS evolution led to recent discoveries that have expanded our understanding of the modular boundaries and various protein-protein interactions present in a PKS (Figure 3.12). By studying giant, >25-module PKSs encoding aminopolyols such as mediomycin, Zhang *et al.* were able to infer that KS domains move together with the upstream ACP and reductive loop during recombination and evolution of modular PKSs. The authors were also able to show that KSs form substrate-specific, phylogenetic clades and to pinpoint previously unknown sequence motifs presumably associated with KS substrate specificity (Zhang *et al.*, 2017). These findings are extremely significant in that they change our understanding of the KS domain by suggesting it to have more substrate specificity than previously thought and by showing that the KS is evolutionarily linked to the domains of the preceding module.

This evolutionary insight led to a proposed change in the modular boundaries of PKS assembly lines (Keatinge-Clay, 2017b). It suggests that, in order to enable synthetic biology efforts, we ought to update the canonical module boundaries (Figure 3.12B) used for *cis*-AT PKS assembly lines such that the KS is grouped with its cognate ACP and reductive loop (Figure 3.12C). In fact, it was recently demonstrated that recognition of ACP_n and KS_{n+1} plays a major role in turnover (Klaus *et al.*, 2016), corroborating the

proposed module redefinition based on phylogenetic analyses (Zhang *et al.*, 2017). A similar phylogenetic analysis of ACP and KS domains from *trans*-AT systems suggests that this boundary update applies to these assembly lines as well (Vander Wood & Keatinge-Clay, 2018).

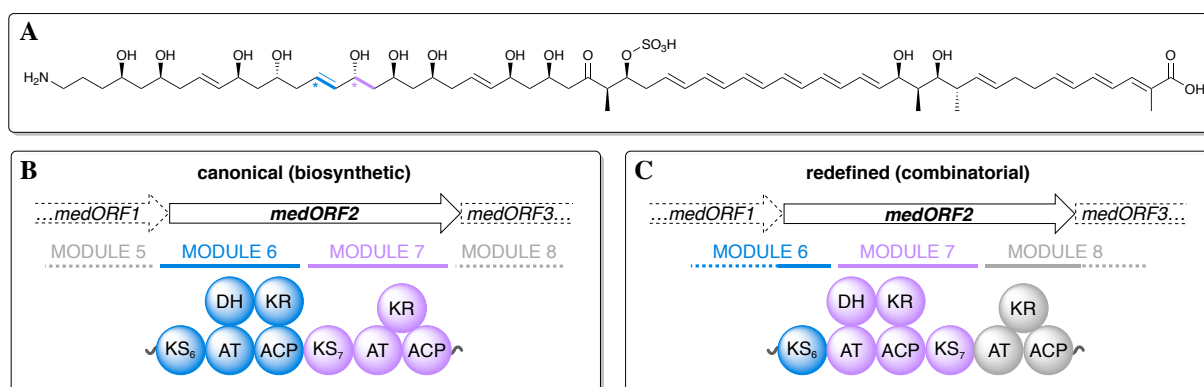


Figure 3.12: Evolutionary studies of large aminopolyol assembly lines leads to proposed redefinition of modular boundaries in polyketide synthases. (A) Structure of the aminopolyol mediomycin 2A and modular schematic of *medORF2* shown with (B) canonical modular boundaries and (C) newly proposed redefinition of modular boundaries.

The DEBS assembly line shown in Figure 3.2 conforms to the originally proposed boundaries, for it seems appropriate to wait for further consensus on the matter before changing the representation of the canonical PKS assembly line. Further, it is worth noting that this combinatorial definition of a PKS module conflicts with the canonical biosynthetic definition of a PKS module. Until now, a PKS module has been primarily defined as a functional unit capable of catalyzing a single round of polyketide chain extension. However, the combinatorial module shown in Figure 3.12C no longer fulfils this criterion, as the KS is no longer grouped with the downstream AT and ACP (compare to Figure 3.4). Still, we encourage the reader to consider the module redefinition when attempting PKS synthetic biology.

Evolution-guided approaches reveal the likelihood of encountering downstream bottlenecks in engineering of larger PKS assembly lines. This prediction seems supported by several examples discussed in this review, such as a) the failure of downstream KR domains to process unnaturally bulky intermediates generated via engineered control of monomer incorporation (Kalkreuter *et al.*, 2019) described in Section 3.3.1.2, and b) the drastically reduced activity of the wild-type pikromycin TE with epimerized, unnatural substrates (Hansen *et al.*, 2017) described in Section 3.5. As the combinatorial abilities of PKS synthetic biology continue to develop, additional downstream bottlenecks are sure to be revealed.

This means that we are still not at the point of designing large PKS chimeras with ease, as we lack a holistic understanding of the interdependency between domains of disparate modules. However, certain targeted modifications can be achieved as highlighted in this review. Moreover, PKS synthetic biology has proven to be a useful method for generating small, chiral, functionalized fragments (Hagen *et al.*, 2016), which will likely prove to be a useful tool to the medicinal chemist.

3.6.3 Bridging natural products and medicinal chemistry research

Medicinal chemistry has long picked up where natural products research leaves off; it works to take bioactive compounds and develop them into optimized pharmaceutical therapies. In the wake of technological developments surrounding combinatorial chemistry, pharmaceutical companies have opted to deprioritize natural products research. However, a concomitant decline in the overall rate of drug discovery was observed.

Synthetic biology has the potential to help reunite these two fields and revitalize drug discovery by using the machinery of natural products biosynthesis to generate novel chemistry. At present, PKS synthetic biology provides the ability to either generate small fragments, or “oligoketides”, *de novo* from a few disparate PKS parts or to make small, targeted structural modifications to large natural product scaffolds, both of which are valuable tools for the field of medicinal chemistry. The ability of synthetic biology to generate chiral, functionalized synthons for incorporation into synthetic schemes would allow

integration of privileged natural product scaffolds with the technological advancements of combinatorial chemistry (Liu *et al.*, 2017).

Synthetic biology may be used to make novel analogues of polyketide natural products that help to tease out structure-activity relationships, as highlighted in Figure 3.9 with geldanamycin (Kim *et al.*, 2009). Alternatively, it may be used to create sustainable sources of clinically relevant natural product analogues by setting up *in vivo* production of compounds that are currently generated through total or semi-synthesis, as shown in Figure 3.10 with the ivermectin example (Zhang *et al.*, 2015). Medicinal chemists may also meaningfully contribute to the field of synthetic biology by generating functionalized precursors for use in precursor-directed biosynthesis studies as well as by providing guidance for synthetic biologists regarding how and where to modify polyketide scaffolds to confer an enhanced therapeutic profile to the natural product.

3.7 **Conclusions**

Owing to developments in PKS structural biology and enzymology, and the persistence of synthetic biologists to develop methods and principles for PKS engineering, polyketide synthetic biology has emerged as a viable tool for generating chemical diversity. Although further refinement of principles and tools is still necessary, the time is ripe to bridge natural products and medicinal chemical research by incorporating polyketide synthetic biology into medicinal chemistry efforts. Medicinal chemical knowledge can substantively inform polyketide engineering efforts by directing engineering towards improved biological activity. Conversely, polyketide synthetic biology can be used to generate difficult-to-access, chiral fragments that can then be incorporated into synthetic schemes. Alternatively, targeted modifications to polyketide scaffolds can be made via synthetic biology to help tease out structure-activity relationships or address supply issues of natural product analogues. With so much to gain from establishing such a partnership, all that is left is to collaborate.

3.8 References Cited

- Ad, O., Thuronyi, B. W., & Chang, M. C. Y. (2017). Elucidating the mechanism of fluorinated extender unit loading for improved production of fluorine-containing polyketides. *Proceedings of the National Academy of Sciences*, 114, E660–E668. doi: 10.1073/pnas.1614196114
- Aminov, R. (2017). History of antimicrobial drug discovery: Major classes and health impact. *Biochemical Pharmacology*, 133, 4–19. doi: 10.1016/j.bcp.2016.10.001
- Annaval, T., Paris, C., Leadlay, P. F., Jacob, C., & Weissman, K. J. (2015). Evaluating Ketoreductase Exchanges as a Means of Rationally Altering Polyketide Stereochemistry. *ChemBioChem*, 16, 1357–1364. doi: 10.1002/cbic.201500113
- Baerga-Ortiz, A., Popovic, B., Siskos, A. P., O'Hare, H. M., Spiteller, D., Williams, M. G., Campillo, N., Spencer, J. B., & Leadlay, P. F. (2006). Directed Mutagenesis Alters the Stereochemistry of Catalysis by Isolated Ketoreductase Domains from the Erythromycin Polyketide Synthase. *Chemistry & Biology*, 13(March), 277–285. doi: 10.1016/j.chembiol.2006.01.004
- Barajas, J. F., Blake-Hedges, J. M., Bailey, C. B., Curran, S., & Keasling, J. D. (2017). Engineered polyketides: Synergy between protein and host level engineering. *Synthetic and Systems Biotechnology*, 2(3), 147–166. doi: 10.1016/j.synbio.2017.08.005
- Bayly, C. L., & Yadav, V. G. (2017). Towards Precision Engineering of Canonical Polyketide Synthase Domains: Recent Advances and Future Prospects. *Molecules*, 22, 235. doi: 10.3390/molecules22020235
- Beck, B. J., Yoon, Y. J., Reynolds, K. A., & Sherman, D. H. (2002). The Hidden Steps of Domain Skipping : Macrolactone Ring Size Determination in the Pikromycin Modular Polyketide Synthase. *Chemistry & Biology*, 9, 575–583.
- Bisang, C., Long, P. F., Cortés, J., Westcott, J., Crosby, J., Matharu, A.-L., Cox, R. J., Simpson, T. J., Staunton, J., & Leadlay, P. F. (1999). A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature*, 401, 502–505.
- Bravo-Rodriguez, K., Klopries, S., Koopmans, K. R. M., Sundermann, U., Yahiaoui, S., Arens, J., Kushnir, S., Schulz, F., & Sanchez-Garcia, E. (2015). Substrate Flexibility of a Mutated Acyltransferase Domain and Implications for Polyketide Biosynthesis. *Chemistry & Biology*, 22(11), 1425–1430. doi: 10.1016/j.chembiol.2015.02.008
- Broadhurst, R. W., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F., & Weissman, K. J. (2003). The Structure of Docking Domains in Modular Polyketide Synthases. *Chemistry & Biology*, 10, 723–731. doi: 10.1016/S
- Caffrey, P., Lynch, S., Flood, E., Finnan, S., & Oliynyk, M. (2001). Amphotericin biosynthesis in *Streptomyces nodosus*: deductions from analysis of polyketide synthase and late genes. *Chemistry & Biology*, 8, 713–723.
- Chabala, J. C., Mrozik, H., Tolman, R. L., Eskola, P., Lusi, A., Peterson, L. H., Woods, M. F., Fisher, M. H., Campbell, W. C., Egerton, J. R., & Ostlind, D. A. (1980). Ivermectin, a New Broad-Spectrum Antiparasitic Agent. *Journal of Medicinal Chemistry*, 23, 1134–1136. doi: 10.1021/jm00184a014
- Chan, K. K. J., & O'Hagan, D. (2012). The Rare Fluorinated Natural Products and Biotechnological Prospects for Fluorine Enzymology. In *Methods in Enzymology* (1st ed., Vol. 516). Elsevier Inc. doi: 10.1016/B978-0-12-394291-3.00003-4

- Chan, Y. A., Podevels, A. M., Kevany, B. M., & Thomas, M. G. (2009). Biosynthesis of polyketide synthase extender units. *Natural Product Reports*, 26, 90–114. doi: 10.1039/b801658p
- Chandran, S. S., Menzella, H. G., Carney, J. R., & Santi, D. V. (2006). Activating Hybrid Modular Interfaces in Synthetic Polyketide Synthases by Cassette Replacement of Ketosynthase Domains. *Chemistry & Biology*, 13(May), 469–474. doi: 10.1016/j.chembiol.2006.02.011
- Choi, S.-S., Nah, H.-J., Pyeon, H., & Kim, E.-S. (2017). Biosynthesis, regulation, and engineering of a linear polyketide tautomycin: a novel immunosuppressant in *Streptomyces* sp. CK4412. *Journal of Industrial Microbiology & Biotechnology*, 44(4), 555–561. doi: 10.1007/s10295-016-1847-2
- Cortes, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J., & Leadlay, P. F. (1990). An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature*, 348, 176–178.
- Dejong, C. A., Chen, G. M., Li, H., Johnston, C. W., Edwards, M. R., Rees, P. N., Skinnider, M. A., Webster, A. L. H., & Magarvey, N. A. (2016). Polyketide and nonribosomal peptide retrobiosynthesis and global gene cluster matching. *Nature Chemical Biology*, 12, 1007–1014. doi: 10.1038/nchembio.2188
- Demain, A. L. (2014). Importance of microbial natural products and the need to revitalize their discovery. *Journal of Industrial Microbiology and Biotechnology*, 41(2), 185–201. doi: 10.1007/s10295-013-1325-z
- Dodge, G. J., Maloney, F. P., & Smith, J. L. (2018). Protein–protein interactions in “cis-AT” polyketide synthases. *Natural Product Reports*, 35, 1082–1096. doi: 10.1039/c8np00058a
- Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J., Katz, L., Science, S., Series, N., May, N., Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J., & Katz, L. (1991). Modular Organization of Genes Required for Complex Polyketide Biosynthesis. *Science*, 252(5006), 675–679.
- Du, L., & Lou, L. (2010). PKS and NRPS release mechanisms. *Natural Product Reports*, 27, 255–278. doi: 10.1039/b912037h
- Dunn, B. J., & Khosla, C. (2013). Engineering the acyltransferase substrate specificity of assembly line polyketide synthases. *Journal of The Royal Society Interface*, 10, 20130297.
- Dutta, S., Whicher, J. R., Hansen, D. A., Hale, W. A., Chemler, J. A., Congdon, G. R., Narayan, A. R. H., Håkansson, K., Sherman, D. H., Smith, J. L., & Skinotis, G. (2014). Structure of a modular polyketide synthase. *Nature*, 510, 512–517. doi: 10.1038/nature13423
- Eng, C. H., Yuzawa, S., Wang, G., Baidoo, E. E. K., Katz, L., & Keasling, J. D. (2016). Alteration of Polyketide Stereochemistry from anti to syn by a Ketoreductase Domain Exchange in a Type I Modular Polyketide Synthase Subunit. *Biochemistry*, 55, 1677–1680. doi: 10.1021/acs.biochem.6b00129
- Eng, C. H., Backman, T. W. H., Bailey, C. B., Magnan, C., García Martín, H., Katz, L., Baldi, P., & Keasling, J. D. (2018). ClusterCAD: A computational platform for type i modular polyketide synthase design. *Nucleic Acids Research*, 46(D1), D509–D515. doi: 10.1093/nar/gkx893

- Eustáquio, A. S., Hagan, D. O., & Moore, B. S. (2010). Engineering Fluorometabolite Production: Fluorinase Expression in *Salinispora tropica* Yields Fluorosalinoporamide. *Journal of Natural Products*, 73(3), 378–382. doi: 10.1021/np900719u.Engineering
- Eustáquio, A. S., McGlinchey, R. P., Liu, Y., Hazzard, C., Beer, L. L., Florova, G., Alhamadsheh, M. M., Lechner, A., Kale, A. J., Kobayashi, Y., Reynolds, K. A., & Moore, B. S. (2009). Biosynthesis of the salinoporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from S-adenosyl-L-methionine. *Proceedings of the National Academy of Sciences*, (12), 12295–12300.
- Eustáquio, A. S., & Moore, B. S. (2008). Mutasynthesis of Fluorosalinoporamide, a Potent and Reversible Inhibitor of the Proteasome. *Angewandte Chemie - International Edition*, 47, 3936–3938. doi: 10.1002/anie.200800177
- Eustáquio, A. S., Pojer, F., Noel, J. P., & Moore, B. S. (2008). Discovery and characterization of a marine bacterial SAM-dependent chlorinase. *Nature Chemical Biology*, 4(1), 69–74.
- Gay, D., You, Y.-O., Keatinge-Clay, A. T., & Cane, D. E. (2013). Structure and Stereospecificity of the Dehydratase Domain from the Terminal Module of the Rifamycin Polyketide Synthase. *Biochemistry*, 52, 8916–8928. doi: 10.1021/bi400988t
- Gillis, E. P., Eastman, K. J., Hill, M. D., Donnelly, D. J., & Meanwell, N. A. (2015). Applications of Fluorine in Medicinal Chemistry. *Journal of Medicinal Chemistry*, 58, 8315–8359. doi: 10.1021/acs.jmedchem.5b00258
- Hagen, A., Poust, S., Rond, T. De, Fortman, J. L., Katz, L., Petzold, C. J., & Keasling, J. D. (2016). Engineering a Polyketide Synthase for In Vitro Production of Adipic Acid. *ACS Synthetic Biology*, 5, 21–27. doi: 10.1021/acssynbio.5b00153
- Hansen, D. A., Koch, A. A., & Sherman, D. H. (2017). Identification of a Thioesterase Bottleneck in the Pikromycin Pathway through Full-Module Processing of Unnatural Pentaketides. *Journal of the American Chemical Society*, 139, 13450–13455. doi: 10.1021/jacs.7b06432
- Helfrich, E. J. N., & Piel, J. (2016). Biosynthesis of polyketides by trans-AT polyketide synthases. *Natural Product Reports*, 33, 231–316. doi: 10.1039/c5np00125k
- Hertweck, C. (2009). The biosynthetic logic of polyketide diversity. *Angewandte Chemie - International Edition*, 48(26), 4688–4716. doi: 10.1002/anie.200806121
- Horsman, M. E., Hari, T. P. A., & Boddy, C. N. (2016). Polyketide synthase and non-ribosomal peptide synthetase thioesterase selectivity: logic gate or a victim of fate? *Natural Product Reports*, 33, 183–202. doi: 10.1039/c4np00148f
- Ikeda, H., Nonomiya, T., Usami, M., Ohta, T., & Omura, S. (1999). Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*. *Proceedings of the National Academy of Sciences*, 96(August), 9509–9514.
- Kalkreuter, E., Crowetipton, J. M., Lowell, A. N., Sherman, D. H., & Williams, G. J. (2019). Engineering the Substrate Specificity of a Modular Polyketide Synthase for Installation of Consecutive Non-Natural Extender Units. *Journal of the American Chemical Society*. doi: 10.1021/jacs.8b10521
- Kalkreuter, E., & Williams, G. J. (2018). Engineering enzymatic assembly lines for the production of new antimicrobials. *Current Opinion in Microbiology*, 45, 140–148. doi: 10.1016/j.mib.2018.04.005

- Kapur, S., Chen, A. Y., Cane, D. E., & Khosla, C. (2010). Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase. *Proceedings of the National Academy of Sciences*, 107, 22066–22071. doi: 10.1073/pnas.1014081107/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1014081107
- Keatinge-Clay, A. T. (2017a). The Uncommon Enzymology of Cis-Acyltransferase Assembly Lines. *Chemical Reviews*, 117, 5334–5366. doi: 10.1021/acs.chemrev.6b00683
- Keatinge-Clay, A. T. (2017b). Polyketide Synthase Modules Redefined. *Angewandte Chemie - International Edition*, 56(17), 4658–4660. doi: 10.1002/anie.201701281
- Keatinge-Clay, A. T. (2016). Stereocontrol within polyketide assembly lines. *Natural Product Reports*, 33(2), 141–149. doi: 10.1039/c5np00092k
- Keatinge-Clay, A. T. (2012). The structures of type I polyketide synthases. *Natural Product Reports*, 29, 1050–1073. doi: 10.1039/c2np20019h
- Kellenberger, L., Galloway, I. S., Sauter, G., Böhm, G., Hanefeld, U., Cortés, J., Staunton, J., & Leadlay, P. F. (2008). A Polylinker Approach to Reductive Loop Swaps in Modular Polyketide Synthases. *ChemBioChem*, 9, 2740–2749. doi: 10.1002/cbic.200800332
- Khosla, C., Kapur, S., & Cane, D. E. (2009). Revisiting the modularity of modular polyketide synthases. *Current Opinion in Chemical Biology*, 13, 135–143. doi: 10.1016/j.cbpa.2008.12.018
- Kim, W., Lee, D., Hong, S. S., Na, Z., Shin, J. C., Roh, S. H., Wu, C. Z., Choi, O., Lee, K., Shen, Y. M., Paik, S. G., Lee, J. J., & Hong, Y. S. (2009). Rational biosynthetic engineering for optimization of geldanamycin analogues. *ChemBioChem*, 10(7), 1243–1251. doi: 10.1002/cbic.200800763
- Klaus, M., & Grninger, M. (2018). Engineering strategies for rational polyketide synthase design. *Natural Product Reports*, 35(10), 1070–1081. doi: 10.1039/c8np00030a
- Klaus, M., Ostrowski, M. P., Austerjost, J., Robbins, T., Lowry, B., Cane, D. E., & Khosla, C. (2016). Protein-Protein Interactions , Not Substrate Recognition , Dominate the Turnover of Chimeric Assembly Line Polyketide Synthases. *The Journal of Biological Chemistry*, 291(31), 16404–16415. doi: 10.1074/jbc.M116.730531
- Koch, A. A., Hansen, D. A., Shende, V. V., Furan, L. R., Houk, K. N., Jiménez-Osés, G., & Sherman, D. H. (2017). A Single Active Site Mutation in the Pikromycin Thioesterase Generates a More Effective Macrocyclization Catalyst. *Journal of the American Chemical Society*, 139, 13456–13465. doi: 10.1021/jacs.7b06436
- Koryakina, I., Kasey, C., McArthur, J. B., Lowell, A. N., Chemler, J. A., Li, S., Hansen, D. A., Sherman, D. H., & Williams, G. J. (2017). Inversion of Extender Unit Selectivity in the Erythromycin Polyketide Synthase by Acyltransferase Domain Engineering. *ACS Chemical Biology*, 12, 114–123. doi: 10.1021/acscchembio.6b00732
- Kwan, D. H., Sun, Y., Schulz, F., Hong, H., Popovic, B., Sim-stark, J. C. C., Haydock, S. F., & Leadlay, P. F. (2008). Prediction and Manipulation of the Stereochemistry of Enoylreduction in Modular Polyketide Synthases. *Chemistry & Biology*, 15(11), 1231–1240. doi: 10.1016/j.chembiol.2008.09.012
- Kwan, D. H., Tosin, M., Schläger, N., Schulz, F., & Leadlay, P. F. (2011). Insights into the stereospecificity of ketoreduction in a modular polyketide synthase. *Organic and Biomolecular Chemistry*, 9, 2053–2056. doi: 10.1039/c1ob00022e

- Labonte, J. W., & Townsend, C. A. (2013). Active Site Comparisons and Catalytic Mechanisms of the Hot Dog Superfamily. *Chemical Reviews*, 113, 2182–2204. doi: 10.1021/cr300169a
- Lau, J., Cane, D. E., & Khosla, C. (2000). Substrate Specificity of the Loading Didomain of the Erythromycin Polyketide. *Biochemistry*, 39, 10514–10520. doi: 10.1021/bi000602v
- Li, Y., Zhang, W., Zhang, H., Tian, W., Wu, L., Wang, S., Zheng, M., Zhang, J., Sun, C., Deng, Z., Sun, Y., Qu, X., & Zhou, J. (2018). Structural Basis of a Broadly Selective Acyltransferase from the Polyketide Synthase of Splenocin. *Angewandte Chemie - International Edition*, 57, 5823–5827. doi: 10.1002/anie.201802805
- Liu, R., Li, X., & Lam, K. S. (2017). Combinatorial chemistry in drug discovery. *Current Opinion in Chemical Biology*, 38, 117–126. doi: 10.1016/j.cbpa.2017.03.017
- Lowden, P. A. S., Wilkinson, B., Böhm, G. A., Handa, S., Floss, H. G., Leadlay, P. F., & Staunton, J. (2001). Origin and True Nature of the Starter Unit for the Rapamycin Polyketide Synthase. *Angewandte Chemie - International Edition*, 40, 777–779.
- Lowry, B., Li, X., Robbins, T., Cane, D. E., & Khosla, C. (2016). A Turnstile Mechanism for the Controlled Growth of Biosynthetic Intermediates on Assembly Line Polyketide Synthases. *ACS Central Science*, 2, 14–20. doi: 10.1021/acscentsci.5b00321
- Medema, M. H., Blin, K., Cimermancic, P., Jager, V. De, Zakrzewski, P., Fischbach, M. A., Weber, T., Takano, E., & Breitling, R. (2011). antiSMASH : rapid identification , annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research*, 39(June), 339–346. doi: 10.1093/nar/gkr466
- Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., de Bruijn, I., Chooi, Y. H., Claesen, J., Coates, R. C., Cruz-Morales, P., Duddela, S., Dusterhus, S., Edwards, D. J., Fewer, D. P., Garg, N., Geiger, C., Gomez-Escribano, J. P., Greule, A., Hadjithomas, M., Haines, A. S., Helfrich, E. J. N., Hillwig, M. L., Ishida, K., Jones, A. C., Jones, C. S., Jungmann, K., Kegler, C., Kim, H. U., Kötter, P., Krug, D., Masschelein, J., Melnik, A. V., Mantovani, S. M., Monroe, E. A., Moore, M., Moss, N., Nützmänn, H.-W., Pan, G., Pati, A., Petras, D., Reen, F. J., Rosconi, F., Rui, Z., Tian, Z., Tobias, N. J., Tsunematsu, Y., Wiemann, P., Wyckoff, E., Yan, X., Yim, G., Yu, F., Xie, Y., Aigle, B., Apel, A. K., Balibar, C. J., Balskus, E. P., Barona-Gómez, F., Bechthold, A., Bode, H. B., Borriss, R., Brady, S. F., Brakhage, A. A., Caffrey, P., Cheng, Y.-Q., Clardy, J., Cox, R. J., De Mot, R., Donadio, S., Donia, M. S., van der Donk, W., Dorrestein, P. C., Doyle, S., Driessen, A. J. M., Ehling-Schulz, M., Entian, K.-D., Fischbach, M. A., Gerwick, L., Gerwick, W. H., Gross, H., Gust, B., Hertweck, C., Höfte, M., Jensen, S. E., Ju, J., Katz, L., Kaysser, L., Klassen, J. L., Keller, N. P., Kormanec, J., Kuipers, O. P., Kuzuyama, T., Kyrpides, N. C., Kwon, H.-J., Lautru, S., Lavigne, R., Lee, C. Y., Lincuan, B., Liu, X., Liu, W., Luzhetskyy, A., Mahmud, T., Mast, Y., Méndez, C., Metsä-Ketelä, M., Micklefield, J., Moore, B. S., Moreira, L. M., Mitchell, D. A., Müller, R., Neilan, B. A., Nett, M., Nielsen, J., O’Gara, F., Oikawa, H., Osbourn, A., Osburne, M. S., Ostash, B., Payne, S. M., Pernodet, J.-L., Petricek, M., Piel, J., Ploux, O., Raaijmakers, J. M., Salas, J. A., Schmitt, E. K., Scott, B., Seipke, R. F., Shen, B., Sherman, D. H., Sivonen, K., Smanski, M. J., Sosio, M., Stegmann, E., Süßmuth, R. D., Tahlan, K., Thomas, C. M., Tang, Y., Truman, A. W., Viaud, M., Walton, J. D., Walsh, C. T., Weber, T., van Wezel, G. P., Wilkinson, B., Willey, J. M., Wohlleben, W., Wright, G. D., Ziemert, N., Zhang, C., Zotchev, S. B., Breitling, R., Takano, E., & Glöckner, F. O. (2015). Minimum Information about a Biosynthetic Gene cluster. *Nature Chemical Biology*, 11, 625–631. doi: 10.1038/nchembio.1890

- Medema, M. H. (2018). Computational Genomics of Specialized Metabolism : from Natural Product Discovery to Microbiome Ecology. *MSystems*, 3.
- Menzella, H. G., Reid, R., Carney, J. R., Chandran, S. S., Reisinger, S. J., Patel, K. G., Hopwood, D. A., & Santi, D. V. (2005). Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nature Biotechnology*, 23(9), 1171–1176. doi: 10.1038/nbt1128
- Moore, B. S., & Hertweck, C. (2002). Biosynthesis and attachment of novel bacterial polyketide synthase starter units. *Natural Product Reports*, 19, 70–99. doi: 10.1039/b003939j
- Moss, S. J., Martin, C. J., & Wilkinson, B. (2004). Loss of co-linearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity. *Natural Product Reports*, 21, 575–593.
- Musirol-Kroll, E. M., Zubeil, F., Schafhauser, T., Härtner, T., Kulik, A., McArthur, J., Koryakina, I., Wohlleben, W., Grond, S., Williams, G. J., Lee, S. Y., & Weber, T. (2017). Polyketide Bioderivatization Using the Promiscuous Acyltransferase KirCII. *ACS Synthetic Biology*, 6, 421–427. doi: 10.1021/acssynbio.6b00341
- Musirol-Kroll, E. M., & Wohlleben, W. (2018). Acyltransferases as Tools for Polyketide Synthase Engineering. *Antibiotics*, 7, 62. doi: 10.3390/antibiotics7030062
- Newman, D. J. (2016). Developing natural product drugs: Supply problems and how they have been overcome. *Pharmacology and Therapeutics*, 162, 1–9. doi: 10.1016/j.pharmthera.2015.12.002
- Newman, D. J., & Cragg, G. M. (2016). Natural Products as Sources of New Drugs from 1981 to 2014. *Journal of Natural Products*, 79(3), 629–661. doi: 10.1021/acs.jnatprod.5b01055
- Olano, C., Méndez, C., & Salas, J. A. (2010). Post-PKS tailoring steps in natural product-producing actinomycetes from the perspective of combinatorial biosynthesis. *Natural Product Reports*, 27, 571–616. doi: 10.1039/b911956f
- Ōmura, S., & Crump, A. (2004). The life and times of ivermectin — a success story. *Nature Reviews Microbiology*, 2, 984–989.
- Paterson, I., & Lam, N. Y. S. (2018). Challenges and discoveries in the total synthesis of complex polyketide natural products. *Journal of Antibiotics*, 71(2), 215–233. doi: 10.1038/ja.2017.111
- Petkovic, H., Lill, R. E., Sheridan, R. M., Wilkinson, B., McCormick, E. L., McArthur, H. A. I., Staunton, J., Leadlay, P. F., & Kendrew, S. G. (2003). A Novel Erythromycin, 6-Desmethyl Erythromycin D, Made by Substituting an Acyltransferase Domain of the Erythromycin Polyketide Synthase. *The Journal of Antibiotics*, 56, 543–551.
- Power, P., Dunne, T., Murphy, B., Lochlainn, L. N., Rai, D., Borissow, C., Rawlings, B., & Caffrey, P. (2008). Engineered Synthesis of 7-Oxo- and 15-Deoxy-15-Oxo- Amphotericins: Insights into Structure-Activity Relationships in Polyene Antibiotics. *Chemistry & Biology*, (January), 78–86. doi: 10.1016/j.chembiol.2007.11.008
- Ray, L., & Moore, B. S. (2016). Recent advances in the biosynthesis of unusual polyketide synthase substrates. *Natural Product Reports*, 33, 150–161. doi: 10.1039/c5np00112a
- Reeves, C. D., Murli, S., Ashley, G. W., Piagentini, M., Hutchinson, C. R., & Mcdaniel, R. (2001). Alteration of the Substrate Specificity of a Modular Polyketide Synthase Acyltransferase Domain through Site-Specific Mutations. *Biochemistry*, 40, 15464–15470. doi: 10.1021/bi015864r

- Richter, C. D., Nietlispach, D., Broadhurst, R. W., & Weissman, K. J. (2008). Multienzyme docking in hybrid megasynthetases. *Nature Chemical Biology*, 4(1), 75–81. doi: 10.1038/nchembio.2007.61
- Robbins, T., Liu, Y. C., Cane, D. E., & Khosla, C. (2016). Structure and mechanism of assembly line polyketide synthases. *Current Opinion in Structural Biology*, 41, 10–18. doi: 10.1016/j.sbi.2016.05.009
- Scaglione, J. B., Akey, D. L., Sullivan, R., Kittendorf, J. D., Rath, C. M., Kim, E.-S., Smith, J. L., & Sherman, D. H. (2010). Biochemical and Structural Characterization of the Tautomycetin Thioesterase: Analysis of a Stereoselective Polyketide Hydrolase. *Angewandte Chemie - International Edition*, 49, 5726–5730. doi: 10.1002/anie.201000032
- Smith, S., & Tsai, S. (2007). The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Natural Product Reports*, 24, 1041–1072. doi: 10.1039/b603600g
- Sugimoto, Y., Ishida, K., Traitcheva, N., Busch, B., Dahse, H.-M., & Hertweck, C. (2015). Freedom and Constraint in Engineered Noncolinear Article Freedom and Constraint in Engineered Noncolinear Polyketide Assembly Lines. *Chemistry & Biology*, 22, 229–240. doi: 10.1016/j.chembiol.2014.12.014
- Thuronyi, B. W., & Chang, M. C. Y. (2015). Synthetic Biology Approaches to Fluorinated Polyketides. *Accounts of Chemical Research*, 48, 584–592. doi: 10.1021/ar500415c
- Tripathi, A., Choi, S.-S., Sherman, D. H., & Kim, E.-S. (2016). Thioesterase domain swapping of a linear polyketide tautomycetin with a macrocyclic polyketide pikromycin in *Streptomyces* sp. CK4412. *Journal of Industrial Microbiology & Biotechnology*, 43(8), 1189–1193. doi: 10.1007/s10295-016-1790-2
- Vander Wood, D. A., & Keatinge-Clay, A. T. (2018). The modules of trans-acyltransferase assembly lines redefined with a central acyl carrier protein. *Proteins*, 86(September 2017), 664–675. doi: 10.1002/prot.25493
- Walker, M. C., Thuronyi, B. W., Charkoudian, L. K., Lowry, B., Khosla, C., & Chang, M. C. Y. (2013). Expanding the Fluorine Chemistry of Living Systems Using Engineered Polyketide Synthase Pathways. *Science*, 341(September), 1089–1095.
- Wang, X., Yan, Y., Zhang, B., An, J., Wang, J., Tian, J., Jiang, L., Chen, Y., Huang, S.-X., Yin, M., Zhang, J., Gao, A., Liu, C.-X., Zhu, Z.-X., & Xiang, W.-S. (2010). Genome Sequence of the Milbemycin-Producing Bacterium. *Journal of Bacteriology*, 192(17), 4526–4527. doi: 10.1128/JB.00596-10
- Weissman, K. J. (2017). Polyketide stereocontrol: a study in chemical biology. *Beilstein Journal of Organic Chemistry*, 13, 348–371. doi: 10.3762/bjoc.13.39
- Weissman, K. J. (2015). The structural biology of biosynthetic megaenzymes. *Nature Chemical Biology*, 11, 660–670. doi: 10.1038/nchembio.1883
- Weissman, K. J., & Leadlay, P. F. (2005). Combinatorial biosynthesis of reduced polyketides. *Nature Reviews Microbiology*, 3(12), 925–936. doi: 10.1038/nrmicro1287
- Whicher, J. R., Dutta, S., Hansen, D. A., Hale, W. A., Chemler, J. A., Dosey, A. M., Narayan, A. R. H., Håkansson, K., Sherman, D. H., Smith, J. L., & Skinotis, G. (2014). Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature*, 510, 560–564. doi: 10.1038/nature13409

- Yñíguez-Gutierrez, A. E., & Bachmann, B. O. (2019). Fixing the Unfixable: The Art of Optimizing Natural Products for Human Medicine. *Journal of Medicinal Chemistry*. doi: 10.1021/acs.jmedchem.9b00246
- Young, J., Stevens, D. C., Carmichael, R., Tan, J., Rachid, S., Boddy, C. N., Müller, R., & Taylor, R. E. (2013). Elucidation of Gephyronic Acid Biosynthetic Pathway Revealed Unexpected SAM-Dependent Methylations. *Journal of Natural Products*, 76, 2269–2276. doi: 10.1021/np400629v
- Yuzawa, S., Backman, T. W. H., Keasling, J. D., & Katz, L. (2018). Synthetic biology of polyketide synthases. *Journal of Industrial Microbiology & Biotechnology*, 45(7), 621–633. doi: 10.1007/s10295-018-2021-9
- Yuzawa, S., Deng, K., Wang, G., Baidoo, E. E. K., Northen, T. R., Adams, P. D., Katz, L., & Keasling, J. D. (2017). Comprehensive in Vitro Analysis of Acyltransferase Domain Exchanges in Modular Polyketide Synthases and Its Application for Short-Chain Ketone Production. *ACS Synthetic Biology*, 6, 139–147. doi: 10.1021/acssynbio.6b00176
- Zhang, J., Yan, Y.-J., An, J., Huang, S.-X., Wang, X.-J., & Xiang, W.-S. (2015). Designed biosynthesis of 25-methyl and 25-ethyl ivermectin with enhanced insecticidal activity by domain swap of avermectin polyketide synthase. *Microbial Cell Factories*, 14, 1–13. doi: 10.1186/s12934-015-0337-y
- Zhang, L., Hashimoto, T., Qin, B., Hashimoto, J., Kozono, I., Kawahara, T., Okada, M., Awakawa, T., Ito, T., Asakawa, Y., Ueki, M., Takahashi, S., Osada, H., Wakimoto, T., Ikeda, H., Shin-ya, K., & Abe, I. (2017). Characterization of Giant Modular PKSs Provides Insight into Genetic Mechanism for Structural Diversification of Aminopolyol Polyketides. *Angewandte Chemie - International Edition*, 56(7), 1740–1745. doi: 10.1002/anie.201611371
- Zhang, X., Chen, Z., Li, M., & Wen, Y. (2006). Construction of ivermectin producer by domain swaps of avermectin polyketide synthase in *Streptomyces avermitilis*. *Applied Microbiology and Biotechnology*, 72, 986–994. doi: 10.1007/s00253-006-0361-2
- Zheng, J., Fage, C. D., Demeler, B., Ho, D. W., & Keatinge-clay, A. T. (2013). The Missing Linker: A Dimerization Motif Located within Polyketide Synthase Modules. *ACS Chemical Biology*, 8, 1263–1270. doi: 10.1021/cb400047s
- Ziemert, N., Alanjary, M., & Weber, T. (2016). The evolution of genome mining in microbes-a review. *Natural Product Reports*, 33(8), 988–1005. doi: 10.1039/c6np00025h

Chapter 4

**Synthetic biology efforts to generate
stereochemical analogues of polyketide natural products**

4.1 Introduction

4.1.1 Targeting the spliceosome for treatment of cancer

In eukaryotic organisms, nascent RNA transcripts – known as pre-mRNA – must undergo an additional round of processing, called *splicing*, prior to translation by the ribosome. Splicing involves removal of non-coding intronic regions from the pre-mRNA to yield the mature mRNA template used for translation. This process is carried out by the spliceosome, an elaborate and dynamic protein-RNA complex that assembles on the pre-mRNA at each instance of splicing (Abelson, 2017). As the scientific understanding of splicing has grown, so has the number of human diseases attributed to aberrant splicing of pre-mRNA transcripts (Novoyatleva *et al.*, 2006). Thus, the spliceosome has emerged as a potential therapeutic target for human diseases, particularly cancer (Lee & Abdel-Wahab, 2016).

There are three known classes of polyketide natural products that act on the spliceosome – pladienolides, herboxidienes, and spliceostatins (Figure 4.1). They possess a common conjugated diene pharmacophore which interacts with the splicing factor 3b (SF3B) component of the spliceosome (Lagisetti *et al.*, 2008; Lagisetti *et al.*, 2014).

The diene moiety imparts a planar structure, and this portion of the molecule is able to insert itself into a narrow tunnel formed by the proteins SF3B1 and PHF5A during the early stages of spliceosome assembly (Cretu *et al.*, 2018). Importantly, binding of the natural product at this site precludes assembly of a functional spliceosome by competing with the pre-mRNA substrate (Finci *et al.*, 2018). While these splice modulators share a common cellular target, the molecular effect of binding to this target varies between cell lines and compounds (Wu *et al.*, 2018; León *et al.*, 2017).

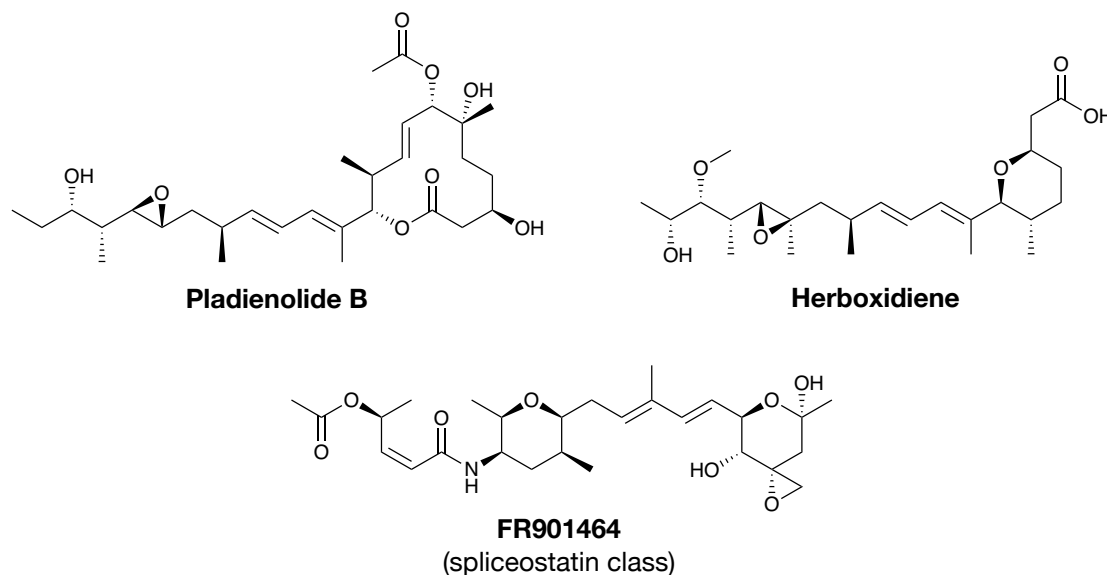


Figure 4.1: The three structural classes of polyketide natural product spliceosome inhibitors.

4.1.2 Using synthetic biology to produce non-natural polyketide analogues with promising therapeutic properties

The pladienolide family of splice modulating polyketides are of particular interest in our lab due to their potential as a therapy for acute myeloid leukemia (Crews *et al.*, 2016) and chronic lymphocytic leukemia (Kashyap *et al.*, 2015). In addition to the diene pharmacophore, these splice modulators are typified by a 12-membered macrolide ring. Shown in Figure 4.1, the most well-known natural product of this class is Pladienolide B (PldB), which was isolated from the fermentation broth of *Streptomyces platensis* Mer-11107 (Sakai *et al.*, 2004). Notably, two synthetic analogues of PldB progressed to clinical trials – E7107 and H3B-8800 (Figure 4.2A). E7107 trials in patients with advanced solid tumors were halted in phase I due to adverse effects (Hong *et al.*, 2014; Eskens *et al.*, 2013). H3B-8800 trials in patients with myelodysplastic syndromes, acute myeloid leukemia, and chronic myelomonocytic leukemia are ongoing (H3 Biomedicine Inc., 2016).

The first natural product of the pladienolide class to be reported was FD-895 from the fermentation broth of *Streptomyces hygroscopicus* A-9561 (Seki-Asano *et al.*, 1994), though its absolute stereochemistry – determined *via* total synthesis (Villa *et al.*, 2012) – was not reported for nearly twenty years. In addition to verifying the structure and stereochemistry of FD-895, Villa *et al.* assayed the full suite of synthesized C16/C17 stereochemical analogues against HCT-116 tumor cells. One diastereomeric analogue, 17*S*-FD-895 (Figure 4.2B), exhibited superior therapeutic properties compared to the natural product. However, the lengthy (30 steps) and low-yield synthesis of 17*S*-FD-895 may prevent this promising compound from making it to the clinic. Thus, there is a need for an alternative method of producing 17*S*-FD-895.

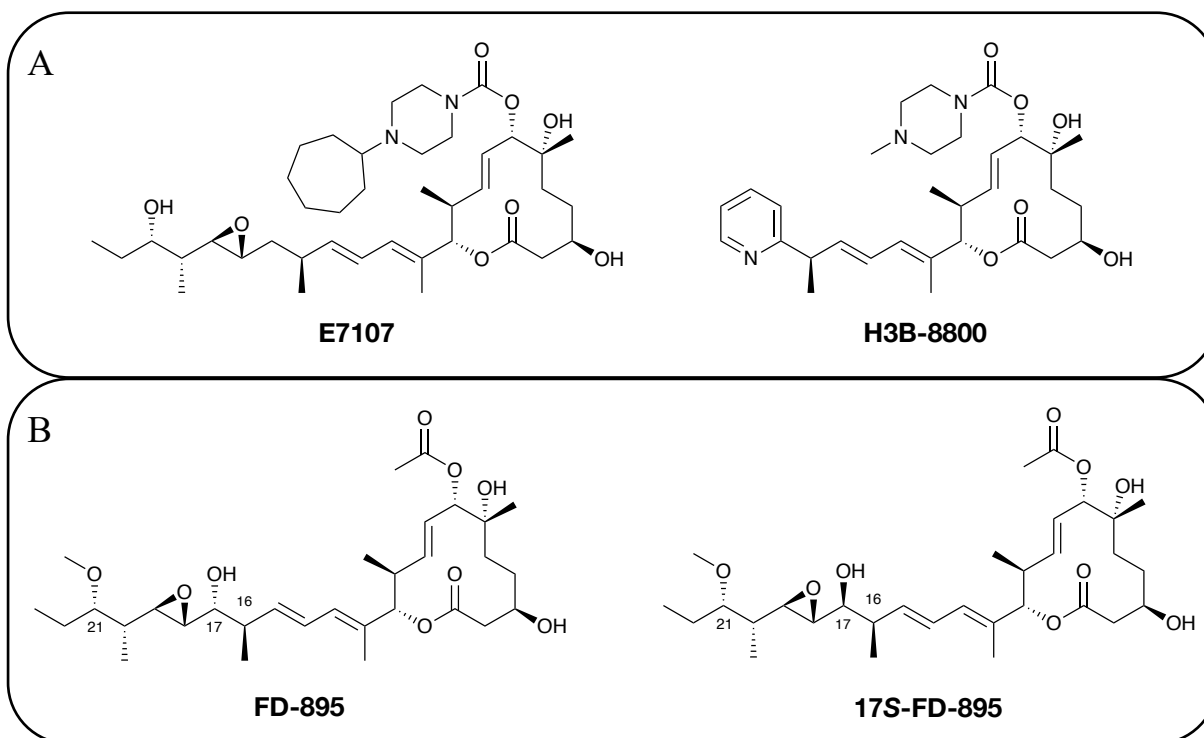


Figure 4.2: Notable synthetic and natural analogues from the pladienolide structural class. (A) The synthetic PldB analogues E7107 and H3B-8800, which progressed to stage I clinical trials. **(B)** The natural product FD-895 and its synthetically-generated diastereomer, 17*S*-FD-895.

As discussed in Chapter 3, the modular architecture of PKS assembly lines makes them promising candidates for synthetic biology, and engineering domains of the reductive loop can effect structural changes in stereochemistry and oxidation state. Unfortunately, the FD-895 producer is not available nor has the sequence of the FD-895 biosynthetic gene cluster (BGC) been reported, rendering it impossible to engineer the cluster. As PldB and FD-895 differ only at two positions – FD-895 features *O*-methylation at C21 and a hydroxyl group at C17 – a potential solution may be to engineer the PKS responsible for pladienolide biosynthesis in *S. platensis* Mer-11107 (Machida *et al.*, 2008).

The third extension module of the Pladienolide PKS – referred to herein as PldMod3 – controls the oxidation state and stereochemistry at positions C16/C17 via domains of its reductive loop. The polyketide intermediate is fully reduced following extension in PldMod3, as the reductive loop in this module contains the full suite of reductive domains – ketoreductase (KR), dehydratase (DH), and enoylreductase (ER) – shown in blue in Figure 4.3A. Inactivation of the DH and ER domains in PldMod3 would yield a product with a hydroxyl group at C17. However, the native Pld-KR3 is a B1-type KR, which does not afford the desired stereochemistry (Figure 4.3B). We propose a reductive loop swap with a donor B2-type KR, which would afford both the desired oxidation state and stereochemistry at C17 and stereochemistry at C16/C17. The resulting polyketide product would be nearly identical to 17*S*-FD-895, though lacking *O*-methylation at C21 (Figure 4.3C).

This chapter is primarily concerned with cultivation of the pladienolide producing organism (Sakai *et al.*, 2004) and preliminary efforts to employ synthetic biology for fermentative production of 17*S*-FD-895. Experiments and results relating to the herboxidiene producer, *Streptomyces chromofuscus*, can be found in Appendix A.

Figure 4.3: Proposed reductive loop swap of the pladienolide polyketide synthase. (A) Native pladienolide PKS, as reported in Machida *et al.* (B) Stereochemical outcomes afforded by different KR subtypes, a topic covered in detail in chapter 3 (section 3.4). (C) Modular schematic of PldMod3 engineered via reductive loop swap and the hypothetical polyketide product that such a modification to the PKS would yield.

4.2 Results and Discussion

4.2.1 Propagation of *Streptomyces* sp. FERM BP-7812 and analysis of phenotypically distinct isolates

Streptomyces platensis Mer-11107, the native producer of PldB, was obtained from the International Patent Organism Depository (IPOD) at the National Institute of Bioscience and Human Technology, Agency of Industrial Science and Technology, Japan, where it was deposited with accession number FERM BP-7812. Cultivation of *Streptomyces* sp. FERM BP-7812 involved an initial liquid culture, which was both cryopreserved and propagated on multiple types of solid media. The solid media cultures revealed that the culture was received from IPOD as a nonhomogeneous mixture (Figure 4.4A). It was unclear whether this was the result of a contaminating organism or if the strain itself was phenotypically variable.

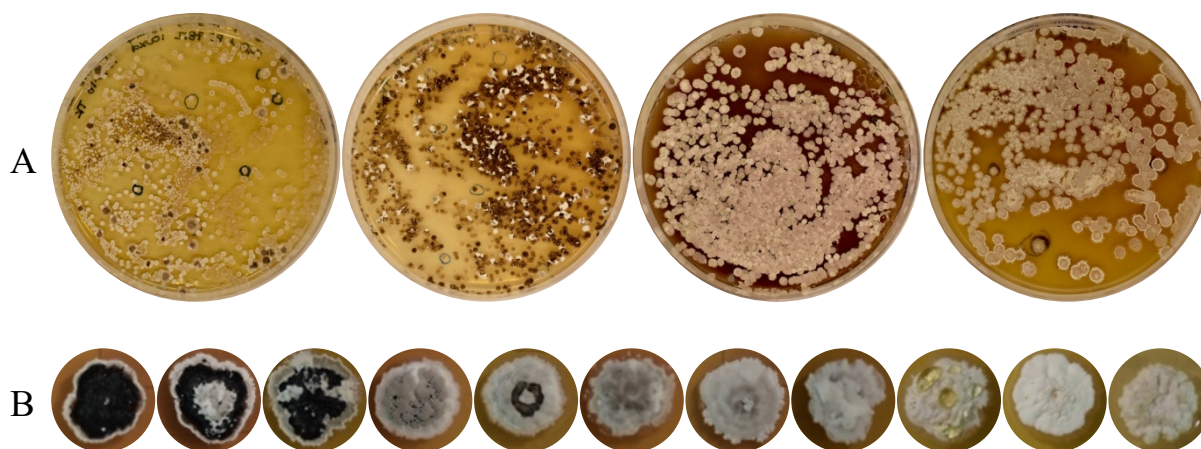


Figure 4.4: Phenotypic heterogeneity of *Streptomyces* sp. FERM BP-7812 on solid media. (A) Diversity plates. Media type from left to right: A1, ISP2, MS, NZSG. **(B)** Isolates cultured on ISP2 agar, showing a spectrum-like array of phenotypes. Isolate ID (from left to right): 1I-i11, 0N10-n3, 0I-i6, 0N-n2, 0I-i2, 1I-i6, 0I10-i1, 1I-i4, 0I-i1, 1I-i12, and 0I-i8.

Thus, in an effort to determine which member(s) of the mixture were the desired pladienolide-producing organism, it was necessary to isolate single colonies from the plates, to cultivate and preserve each one individually, and to examine their differential characteristics. Propagation of single colonies facilitated more detailed observation of the phenotypes expressed by the FERM BP-7812 isolates (Figure 4.4B). There was marked variation in phenotype among the isolates, particularly with respect to color, sporulation behavior, and excretion of pigmented metabolites. The phenotypic variation of growing colonies is most clearly observed on ISP2 agar, thus ISP2 was chosen as the solid media used for routine culturing. None of the isolates appeared to be particularly unique in their phenotype; rather, the isolated colonies seemed to all lie along the same spectrum of phenotype. It remained unclear whether all isolates were variations of the same strain or whether the culture contained a contaminating organism.

To further investigate the possibility of a contaminated culture, two experimental methods were employed – IDBac and partial 16S sequencing. IDBac facilitated indirect assessment of the phylogenetic relatedness between isolates *via* bioinformatic analysis of colonial protein content, which was assessed via mass spectrometry (Clark *et al.*, 2018). This analysis revealed that the isolates form three distinct groupings, correlative to phenotype, at the species level or below (Figure 4.5).

Half of the isolates were subjected to 16S analysis. This involved isolation of gDNA from the isolates (Figure 4.6A), PCR amplification of the ~1400 bp 16S rRNA gene (Figure 4.6B), and Sanger sequencing of the amplicons. This revealed that the fifteen isolates – seven black, five grey, and three white (Figure 4.6C) – possess identical partial 16S rRNA sequences. This result, along with IDBac clustering, suggests the presence of a contaminating organism is unlikely. Thus, the pladienolide production ability of the isolates must be compared.

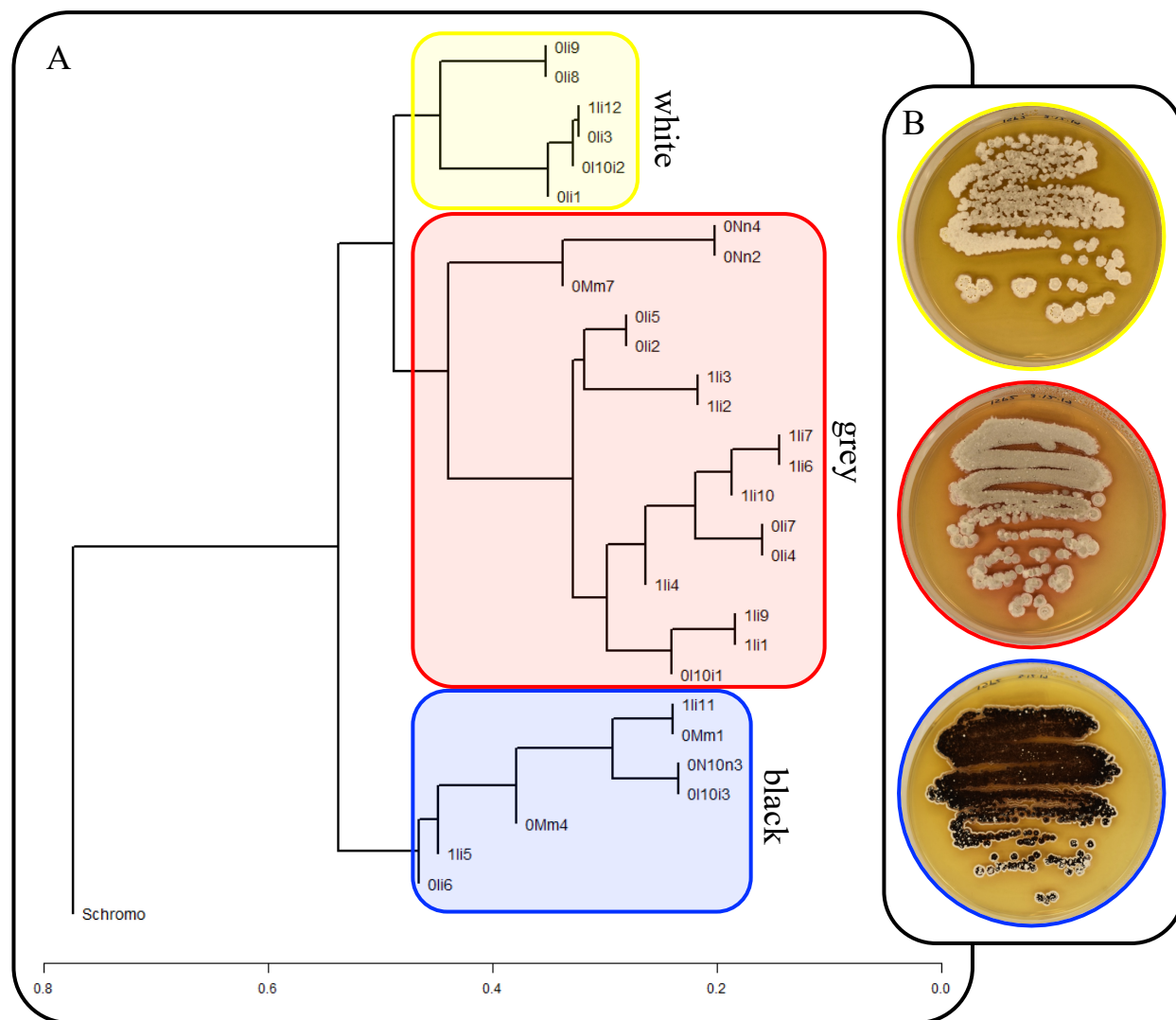


Figure 4.5: IDBac analysis of isolates. (A) Dendrogram generated in IDBac shows three distinct groupings of isolates, which correlate to phenotype. White phenotypes are boxed in yellow. Grey phenotypes are boxed in red. Black isolates are boxed in blue. **(B)** Representative isolates of each grouping – white, 0I10-i2; grey, 0I10-i1; and black, 0I10-i3.

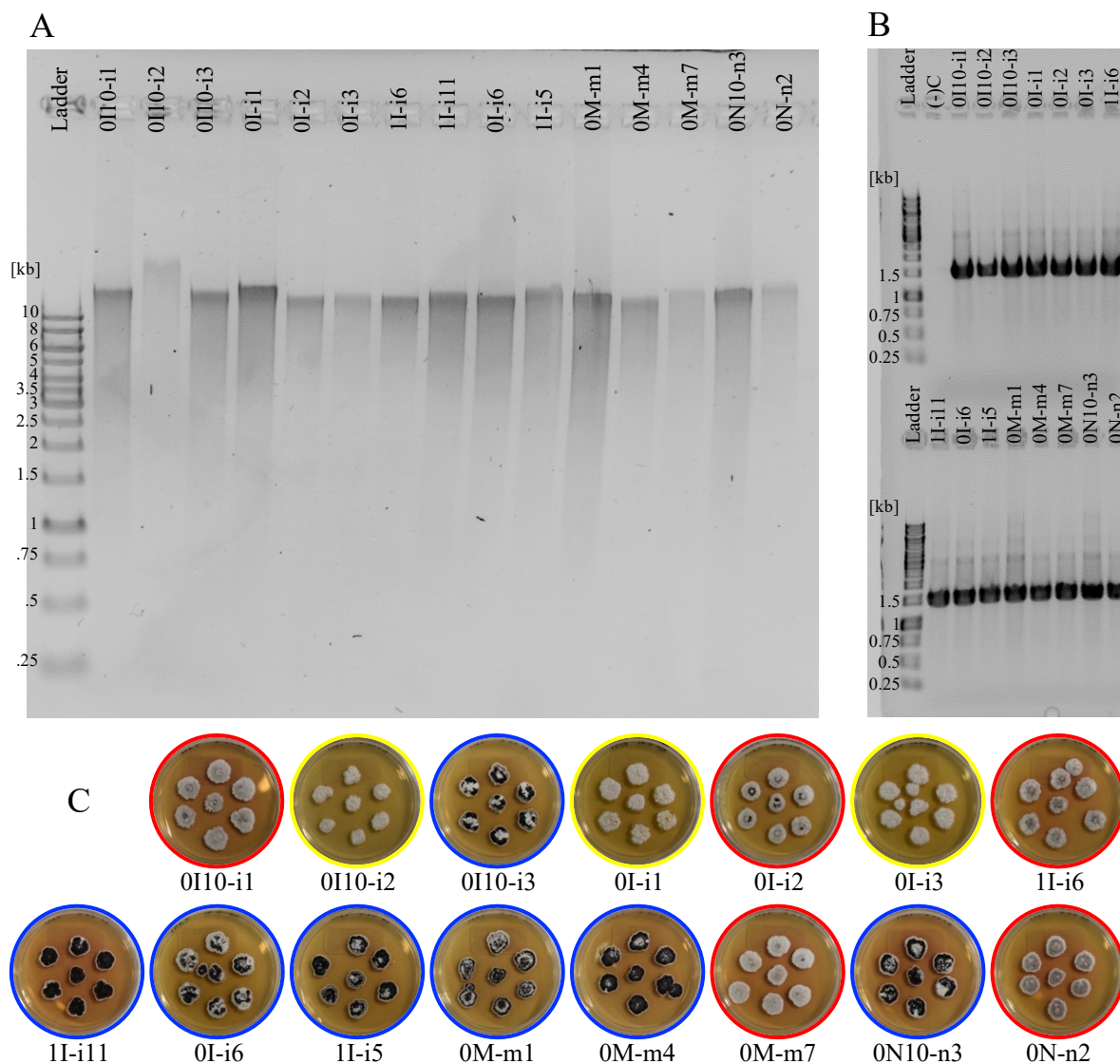


Figure 4.6: Genomic DNA isolation and 16S analysis for half of the cryopreserved *Streptomyces* sp. FERM BP-7812 isolates. Gels showing (A) isolate gDNA and (B) 16S PCR amplicon. (C) Isolates selected for 16S analysis, shown cultured on ISP2 agar (blue outline = black phenotype group; red outline = grey phenotype group; yellow outline = white phenotype group)

4.2.1.1 Analysis of pladienolide B production in isolates of *Streptomyces* sp. FERM BP-7812

As a preliminary assessment, targeted metabolomics was employed to determine whether the heterogeneous mixed culture produced PldB via fermentation. An extract of the production culture supernatant was compared with an authentic PldB standard using high performance liquid chromatography (HPLC). This analysis confirmed that the heterogeneous culture received from IPOD possesses the ability to biosynthesize PldB (Figure 4.7). The PldB standard elutes at 40.384 min, and the extract has a correlating peak at 40.287 min. The solvent scheme used here for analysis of PldB in extracts required optimization to achieve better resolution of the various compounds present in the extract.

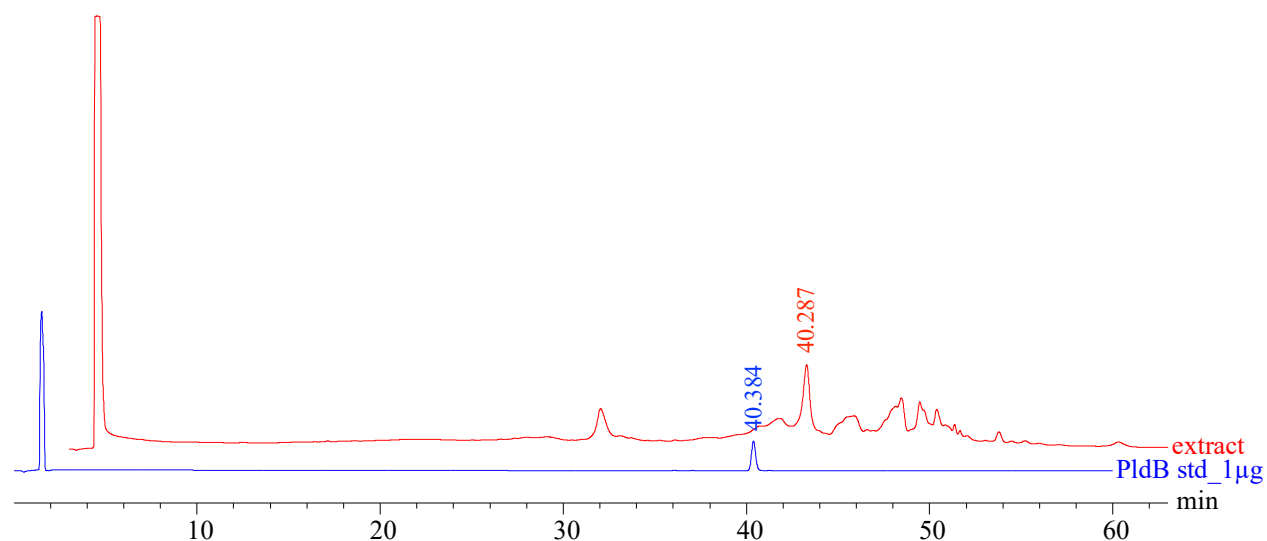


Figure 4.7: Preliminary assessment of pladienolide B production in mixed culture of *Streptomyces* sp. FERM BP-7812.

Notably, the solvent schemes differ slightly between all PldB production analyses described herein, as method optimization was concurrent with each new analysis. All relevant solvent schemes and the analysis they are associated with are described in the methods section.

Pladienolide titers were next tested between phenotype groups. Three isolates – 0I10-i1, 0I10-i2, and 0I10-i3 (Figure 4.5B) – were selected as the experimental set as they originated from the same diversity plate. Pladienolide production cultures were made for each isolate, followed by extraction of supernatant with EtOAc. A 100 μ g aliquot of each extract (dissolved in MeOH) was subjected to HPLC analysis, which revealed that all three phenotype groups can biosynthesize detectable amounts of PldB in liquid culture (Figure 4.8).

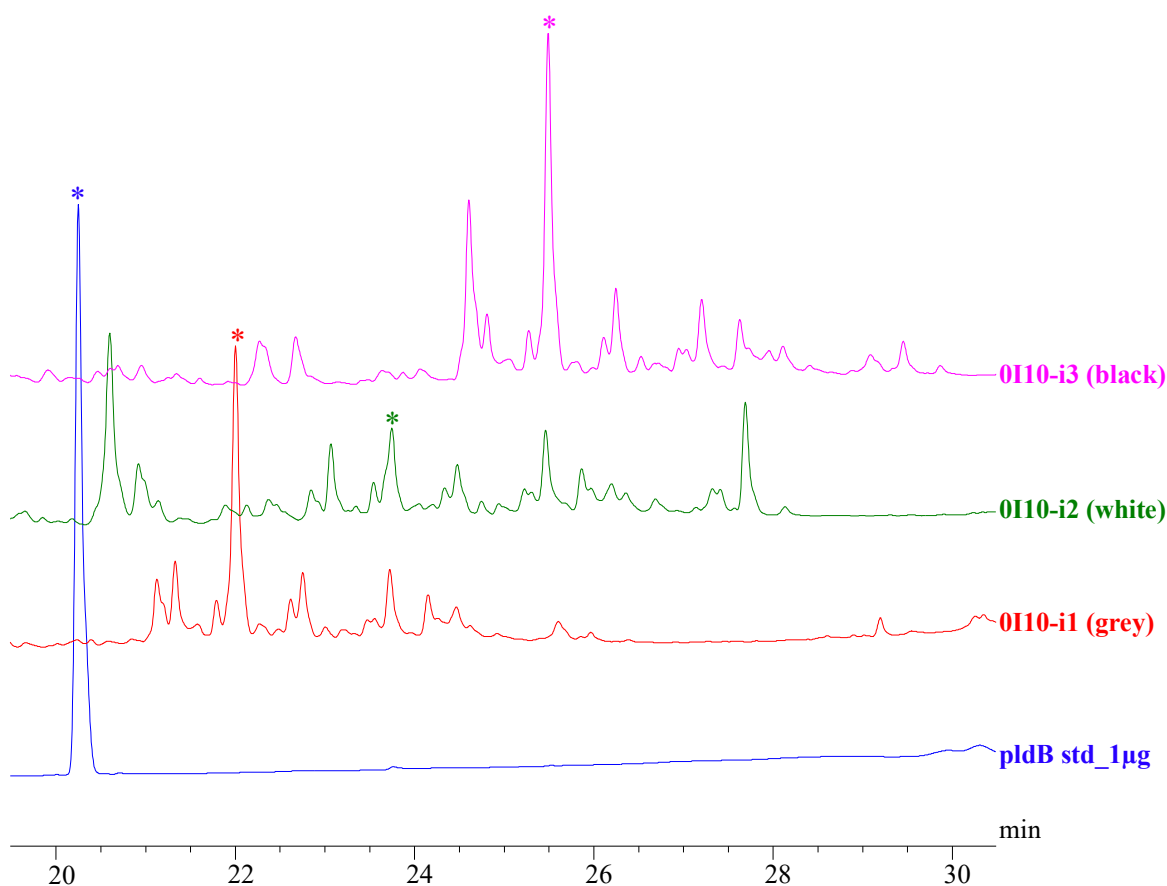


Figure 4.8: Comparison of pladienolide B(*) production between isolate phenotype groups.

While all three isolates produced PldB, the white isolate showed significantly lower titers than the black and grey isolates. Further, the black isolate slightly outproduced the grey isolate (TABLE VIII). It was assumed that this trend would generalize across isolates as several qualitative observations suggest the white phenotype is the result of laboratory domestication. This is discussed in more detail in section 4.2.1.2.

TABLE VIII: COMPARISON OF PLADIENOLIDE B TITERS BETWEEN PHENOTYPE GROUPS

Isolate	Phenotype Group	PldB titers (mg/L)
0I10-i1	Grey	0.716
0I10-i2	White	0.238
0I10-i3	Black	0.772

Pladienolide titers were next obtained for the seven black isolates – 0I10-i3, 0I-i6, 1I-i5, 1I-i11, 0M-m1, 0M-m4, and 0N10-n3. Two production cultures were made for each isolate, and the culture supernatant of each replicate was extracted separately with. The stacked chromatograms shown in Figure 4.9 compare the highest yielding replicate of each isolate and 0.25 μ g of PldB standard. Calculated titers for each replicate and the average of both replicates for each isolate are reported in TABLE IX.

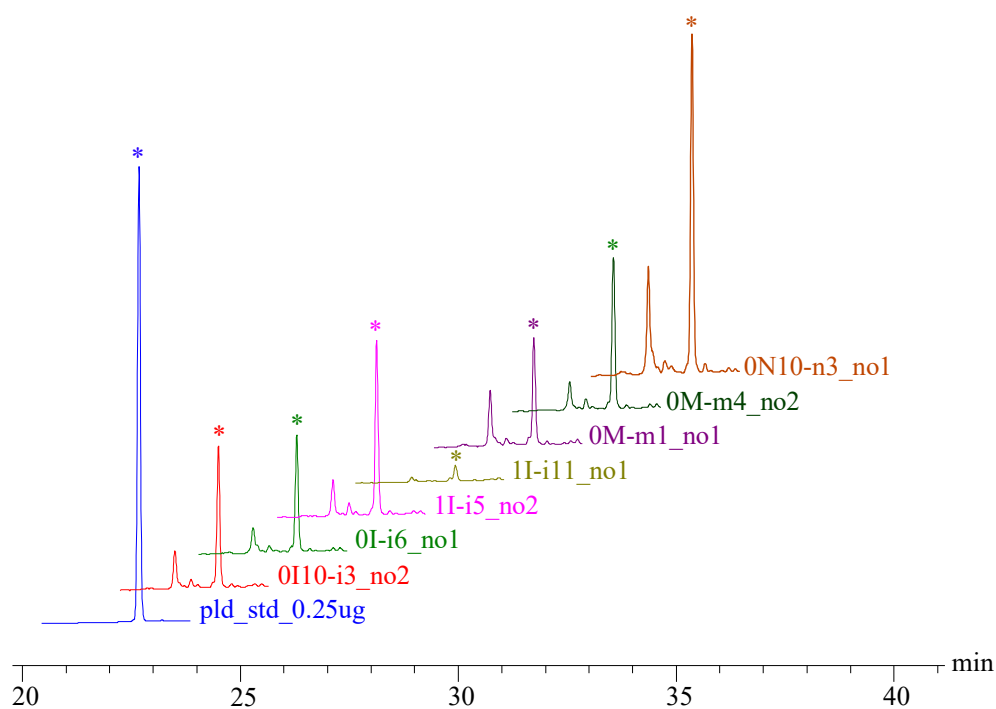


Figure 4.9: Comparison of pladienolide B(*) production between black isolates.

TABLE IX: COMPARISON OF PLADIENOLIDE B TITERS BETWEEN BLACK ISOLATES

Isolate	PldB titers (mg/L)		Average PldB titers (mg/L)
	Replicate #1	Replicate #2	
0I10-i3	4.2	6.4	5.3
0I-i6	5.4	0.3	2.9
1I-i5	0.4	8.0	4.2
1I-i11	0.8	0.4	0.6
0M-m1	5.1	1.7	3.4
0M-m4	5.0	7.0	6.0
0N10-n3	15.4	4.1	9.8

The titers reported in TABLE IX show much variation between replicates, reflecting general issues with reproducibility in cultivation of filamentous bacteria. One such issue is variation in coil density of the stainless-steel spring used during cultivation. The spring serves not only to aerate the culture, but also to break up the filamentous chain into a finer cell suspension. Nonuniform cell suspensions between isolate cultures may result from variation in the density of the coiled spring, thus impacting the accuracy of compared titers. Additionally, the seed cultures of each isolate grow at slightly different rates, likely due to differing cell densities of the cryostocks. This makes it difficult to normalize the number of cells used to inoculate each seed culture across all isolates, and the standard technique of measuring the optical density of the culture is not reliable with filamentous bacteria.

Furthermore, the titers reported in TABLE IX are likely inaccurate (too low) for isolate 0I-i6 (replicate #2) and 1I-i5 (replicate #1) as these two production cultures had significant amounts of cell mass growing up the side of the flask by the end of the cultivation period, which results in a thin culture growing in the base of the flask. This source of error is, again, not uncommon in the cultivation of filamentous bacteria. Nonetheless, 0N10-n3 stood out amongst the black isolates and has been selected as the “best” pladienolide producer.

Notably, reported titers for isolate 0I10-i3 from TABLE VIII and TABLE IX vary by an order of magnitude. This can likely be attributed to differences in incubation time – the cultures for comparison of titers between black isolates (TABLE IX) were grown for six days, while the cultures for comparison of titers between phenotype groups (TABLE VIII) were only grown for five days. The extra day of incubation for the black isolates was aberrant, as all other PldB production cultures were cultivated for five days in accordance with Machida *et al.* However, future studies should utilize six days given the order of magnitude increase in PldB titers the additional day affords.

4.2.1.2 Laboratory domestication of *Streptomyces* sp. FERM BP-7812

The phenotypic variability of *Streptomyces* sp. FERM BP-7812 is striking. For instance, the white isolates do not exhibit sporulation, nor do they produce a red pigment that is secreted into solid media by both the black and grey isolates. Further, the IDBac clustering (Figure 4.4A) and spectrum of isolate phenotypes (Figure 4.3B) suggest that the grey phenotype is an intermediate between the black and white phenotypes. We hypothesize that the black phenotype is the wild-type organism and the white phenotype arises due to strain domestication in the laboratory, a phenomenon which has been reported in *Streptomyces* and leveraged to study their development and evolutionary history (Chater & Chandra, 2006).

A passaging study of four black isolates – 0N10-n3, 0M-m4, 0I-i6, and 0I10-i3 – was performed to test this hypothesis. The study involved serial passaging of isolates via liquid culture and visualization of the phenotypic composition in each passage via plating on solid media. Photographs of solid media cultures were taken after nine days of incubation. (Figure 4.10) It is important to note that Passage 2 is not represented in Figure 4.10 as minimal (1-2 colonies) or no growth was observed on the solid media plates, though the liquid culture for each isolate appeared healthy and was successfully utilized for inoculation of Passage 3. It is unknown why this occurred.

Though not phenotypically homogenous, the black phenotype predominates Passage 1 for all four isolates. With increasing passage number, the black phenotype makes up a smaller proportion of the colonies observed on solid media. Concomitantly, colonies exhibiting grey, white or tan phenotypes appear in increasing proportion. Interestingly, the observed rate of phenotypic drift varied between the isolates.

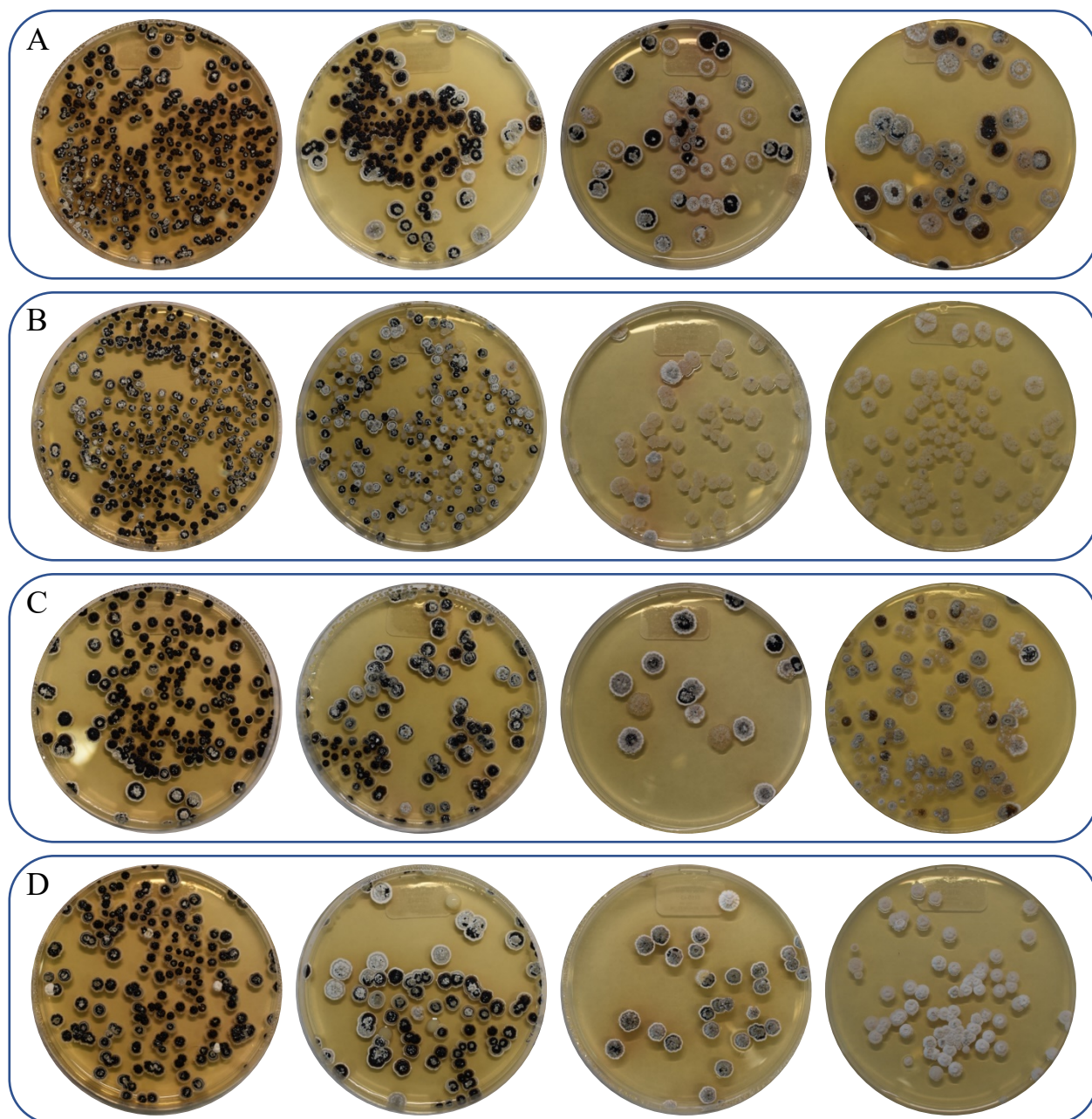


Figure 4.10: Passing study of black isolates. (A) 0N10-n3 (B) 0M-m4 (C) 0I-i6 (D) 0I10-i3. From left to right – passage 1, passage 3, passage 4, passage 5.

The black phenotype is well retained into Passage 5 of isolate 0N10-n3 (Figure 4.10A), which was identified in the previous section as the best producer of PldB. In contrast, isolate 0M-m4 (Figure 4.10B) loses the black phenotype by Passage 4, which contains mostly tan colonies and four colonies that appear to be chimeras of the white and grey phenotype. By Passage 5, 0M-m4 contains entirely tan colonies. Isolate 0I-i6 (Figure 4.10C) also retains the black phenotype into Passage 5, but there is a higher proportion of tan colonies than observed in Passage 5 of 0N10-n3. Finally, isolate 0I10-i3 (Figure 4.10D) retains the black phenotype in Passage 3, but in Passage 4 the black phenotype is only observed in chimeric black-grey colonies. While the black and grey phenotypes are absent in Passage 5 of 0I10-i3 (unlike 0N10-n3 and 0I-i6), it is not yet fully composed of tan colonies as the white phenotype is still well retained (unlike 0M-m4). Collectively, these observations do suggest that the black phenotype is that of the wild-type organism, which is lost in the laboratory setting. However, given the previously discussed reproducibility issues regarding strain cultivation, the passage experiment would need to be repeated to confirm that the observed variation in drift is really isolate-specific.

Streptomyces are remarkable not only for their biosynthetic potential, but also for their complex developmental cycle (Manteca & Yagüe, 2018). The emergence of the white phenotype and tan (or “bald”) phenotype in *Streptomyces* has been associated with regulatory changes in key developmental genes involved in sporulation and formation of aerial mycelium, respectively (Chater & Chandra, 2006). While the rapid phenotypic drift observed in the passaging study of *Streptomyces* sp. FERM BP-7812 supports our hypothesis that this strain is susceptible to domestication in the laboratory setting, additional evidence – such as genome sequencing and comparative metabolomics of low and high passage numbers – is needed to conclusively prove this.

4.2.2 **Efforts to use synthetic biology for production of 17S-FD-895**

4.2.2.1 **in silico analysis of donor reductive loops**

ClusterCAD, a computational webtool designed to facilitate synthetic biology studies of PKS assembly lines, was used to find potential donor B2-type KR domains for the PldMod3 reductive loop swap. ClusterCAD helps users find donor PKS “parts” (*i.e.* catalytic domains) that can affect desired modifications to the polyketide scaffold (Eng *et al.*, 2018). The ClusterCAD entry for erythromycin (<https://clustercad.jbei.org/pks/BGC0000055.1/>) was used as a starting point to find potential B2-type KR donors because the first extension module of DEBS, the erythromycin PKS, contains a well-characterized B2-type KR (Østergaard *et al.*, 2002; Keatinge-Clay & Stroud, 2006; Zheng & Keatinge-Clay, 2013).

ClusterCAD allows users to perform blastp (protein-protein BLAST) searches directly in the webtool. Thus, the amino acid sequence of the B2-type KR1 from DEBS was submitted to find all KRs in the database with a similar amino acid sequence. The KR subtype of the resulting hits were readily verifiable due to the detailed domain annotations offered in ClusterCAD. Fifteen B2-type KRs were identified as potential donors for the proposed (Figure 4.3) reductive loop swap of PldMod3. They are described in TABLE X, which lists the PKS of origin, the polypeptidic subunit, and the KR number. Additionally, the table lists the organism that harbors the PKS.

TABLE X: B2-TYPE KETOREDUCTASE DONOR CANDIDATES

PKS	Subunit	KRn ^a	Native Producer
Aculeximycin (Acu)	AcuAII	KR5	<i>Kutzneria albida</i> DSM 43870
Brasilinolide (Nbr)	NbrG	KR13	<i>Nocardia terpenica</i> IFM 0406
Brasilinolide	NbrH	KR16	<i>Nocardia terpenica</i> IFM 0406
Brasilinolide	NbrH	KR17	<i>Nocardia terpenica</i> IFM 0406
Erythromycin (Ery)	EryAI	KR1	<i>Saccharopolyspora erythraea</i> NRRL2338
Halstoctacosanolide (Hls)	HlsC	KR7	<i>Streptomyces halstedii</i> HC34
Lankamycin (Lkm)	LkmAI	KR1	<i>Streptomyces rochei</i> 7434AN4 (pSLA2-L)
Lasalocid (Las)	LasAIV	KR7	<i>Streptomyces lasaliensis</i> NRRL3382R
Megalomicin (Meg)	MegAI	KR1	<i>Micromonospora megalomicea</i> NRRL3275
Meridamycin (Mer)	MerC	KR10	<i>Streptomyces violaceusniger</i> DSM 4137
Pikromycin (Pik)	PikAI	KR1	<i>Streptomyces venezuelae</i> ATCC 15439
Monensin (Mon)	MonAVIII	KR12	<i>Streptomyces cinnamomensis</i> ATCC 15413
Mycolactone (Mls)	MlsAI	KR2	<i>Mycobacterium ulcerans</i> AGY99 (pMUM001)
PM100117 (Gon)	GonP5	KR14	<i>Streptomyces caniferus</i> GUA-06-06-006A
Salinomycin (Sln)	SlnA7	KR11	<i>Streptomyces albus</i> XM211

^a KRn (or KR_n) is the KR number, and the value of n corresponds to extension module number (*i.e.* loading module is considered n=0 and first extension module is considered n=1).

As covered in Chapter 3, it is important to consider phylogeny and evolutionary relationships between domains of the assembly line when engineering a PKS via domain exchange. Therefore, we conducted a phylogenetic analysis of B2-type donor KR_n and the evolutionarily relevant ACP_n and KS_{n+1} domains associated with them.

The phylogenetic tree of Pld-KR3 and the B2-type donors is shown in Figure 4.11A. Given that the B1-type KR of PldMod3 is inherently distinct from the B2-type donor KRs, the relatively high distance of Pld-KR3 from the potential donors is not a surprise. However, that KR12 from the monensin (Mon) PKS and KR2 from the mycolactone (Mls) PKS are more distinct from the rest of the donor pool than Pld-KR3 is surprising. Inspection of the aligned amino acid sequences reveals that neither Mon-KR12 nor Mls-KR2 contain the Leu-Asp-Asp (LDD) motif (Figure 4.11B, residues 94-96). In Mon-KR12 it appears as VND and in Mls-KR2 it appears as LAD.

The LDD motif is considered the strongest sequence predictor of a B-type (both B1- and B2-type) KR domain (Weissman, 2016), and it is present in Pld-KR3 and all other candidate donors. Further, both Mon-KR12 and Mls-KR2 feature an alanine two residues C-terminal of the catalytic tyrosine (Figure 4.11B, residue 153). All other donor candidates feature a proline at this position, which is a known sequence motif that can be used to typify B2-type KR domains (Keatinge-Clay, 2007). As expected of a B1-type KR, Pld-KR3 does not possess a proline at this site.

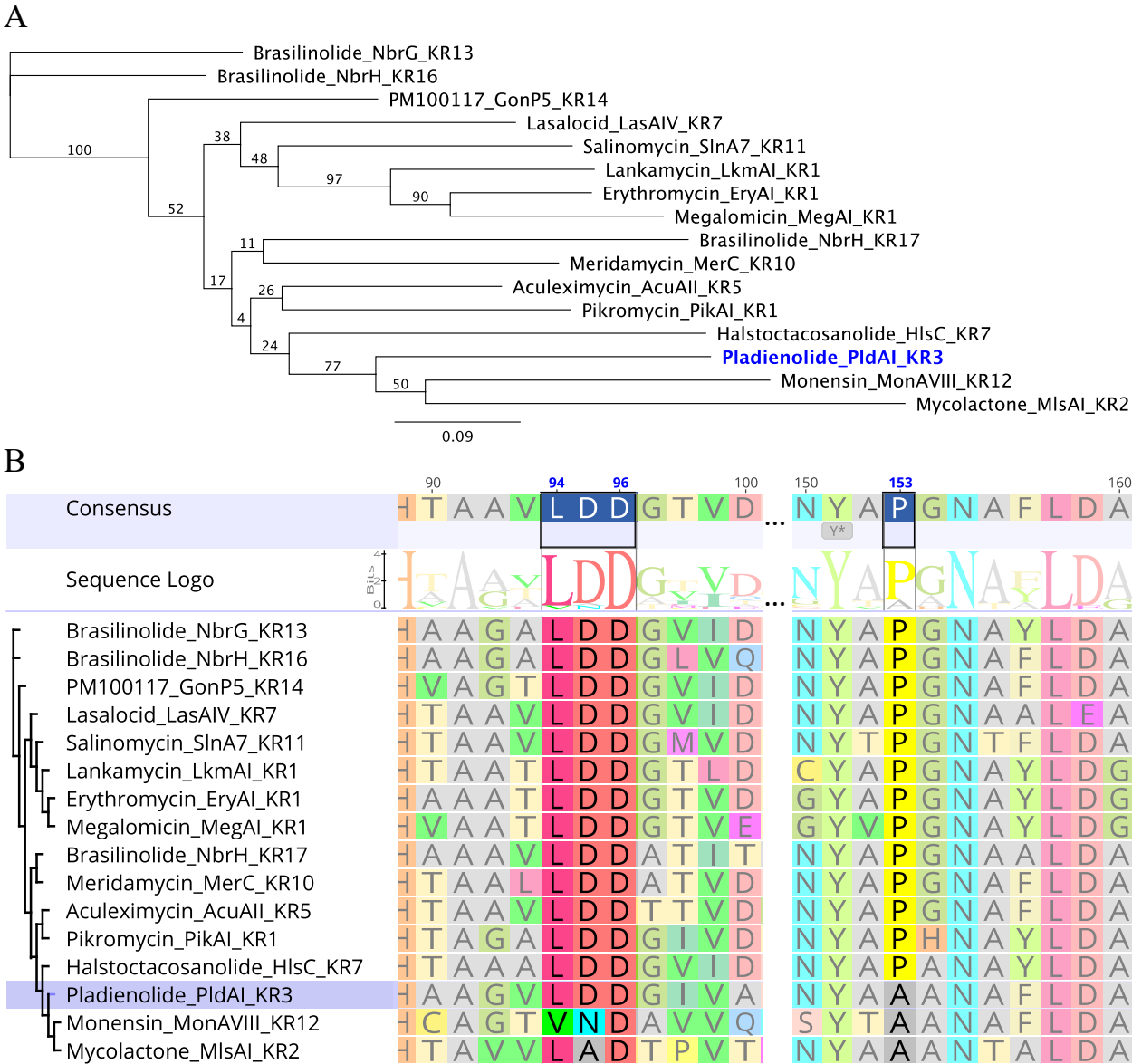


Figure 4.11: Phylogenetic tree and sequence alignment of donor ketoreductase domains
(A) Phylogenetic tree (B) Protein alignment showing residues characteristic of B-type KR domains. Residues 101-149 are omitted.

The phylogenetic tree of Pld-ACP3 and ACP_n domains from the candidate donors is shown in Figure 4.12. Due to the short sequence length and high degree of sequence conservation, it is difficult to construct reliable phylogenetic trees of ACP domains (Jenke-Kodama *et al.*, 2006). Still, there are two distinct groupings at the top and the bottom of the tree worth mentioning.

At the top of the tree in Figure 4.12, ACP_n domains from erythromycin and megalomycin PKSs form a distinct pair, and the domain from the lankamycin PKS is a relatively short distance away. The N-1 modules of these assembly lines are loading modules (TABLE XI). They possess an AT_L-ACP architecture (see section 3.2.2.1) and utilize an acylthioester-CoA starter unit. Thus, their ACP_n domains likely exhibit unusual substrate specificity with respect to the other ACP_n in this analysis. At the bottom of the tree in Figure 4.12, the ACP_n from mycolactone, brasilinolide, and PM100117 PKSs are distinct from the rest, including Pld-ACP3, which casts doubt on the suitability of these donors for engineering of the Pld PKS.

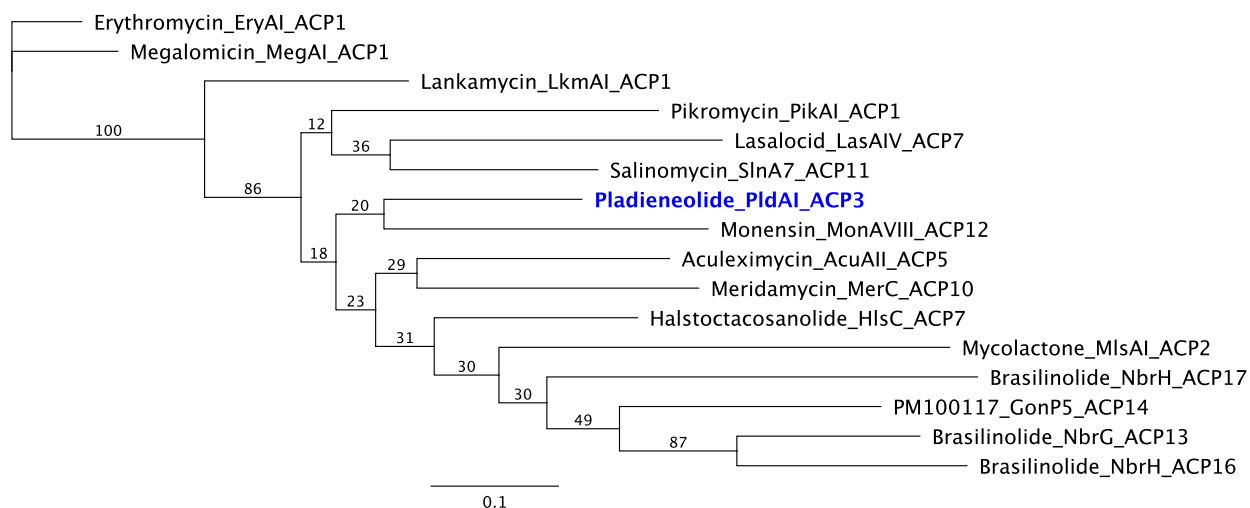


Figure 4.12: Phylogenetic tree of evolutionarily relevant acyl carrier protein domains from donor candidates.

TABLE XI: ASSEMBLY LINE CONTEXT OF DONOR B2-TYPE KETOREDUCTASE DOMAINS

PKS-KRn	Proceeding Module (N-1)		Donor Module (N)		Succeeding Module (N+1)	
	reductive loop	extender unit ^c	reductive loop	extender unit ^c	reductive loop	extender unit ^c
Pld-KR3	KR _{B1} , DH	mal	KR _{B1} , DH, ER	mmal	KR _{B1} , DH	mal
Acu-KR5	KR, DH, ER	mal	KR _{B2}	mmal	KR _{B1} , DH	mal
Nbr-KR13	DH	mal	KR _{B2}	mmal	KR, DH, ER	mal
Nbr-KR16	KR _{A1}	mal	KR _{B2}	mmal	KR _{B2}	mal
Nbr-KR17	KR _{B2}	mmal	KR _{B2}	mal	KR, DH	mal
Ery-KR1	N/A ^a	propionyl-CoA	KR _{B2} , DH ₀	mmal	KR _{A1}	mmal
Hls-KR7	KR _{A1}	mal	KR _{B2}	mmal	KR _{B1} , DH	mal
Lkm-KR1	N/A ^a	2metbut-CoA	KR _{B2} , DH ₀	mmal	KR _{A1}	mmal
Las-KR7	none	mmal	KR _{B2}	mmal	KR _{B1} , DH, ER	mal
Meg-KR1	N/A ^a	propionyl-CoA	KR _{B2}	mmal	KR _{A1}	mmal
Mer-KR10	KR _{B1}	mal	KR _{B2}	mmal	KR _{A1}	mal
Pik-KR1	N/A ^a	mmal	KR _{B2} , DH ₀	mmal	KR _{B1} , DH	mal
Mon-KR12	KR _{A1}	mmal	KR _{B2} , DH ₀	mmal	N/A ^b	N/A ^b
Mls-KR2	KR _{B1} , DH ₀	mal	KR _{B2} , DH ₀ , ER ₀	mmal	KR _{B1} , DH	mmal
Gon-KR14	DH ₀	mal	KR _{B2}	mmal	KR _{B1} , DH, ER	mal
Sln-KR11	DH ₀	mmal	KR _{B2}	mmal	KR _{A2}	mmal

^a Loading module – no reductive loop domains are present.

^b Monensin donor module is terminal to PKS.

^c Extender unit abbreviations: mal = malonyl-CoA ; mmal = methylmalonyl-CoA

The phylogenetic tree of Pld-KS4 and KS_{n+1} domains from the candidate donors is shown in Figure 4.13. The tree has two distinct groupings – the top half of the tree contains eight KS_{n+1} domains which cover a wide range of sequence distance, while the bottom half contains seven KS_{n+1} domains, including Pld-KS4, that cover a narrow range of sequence distance.

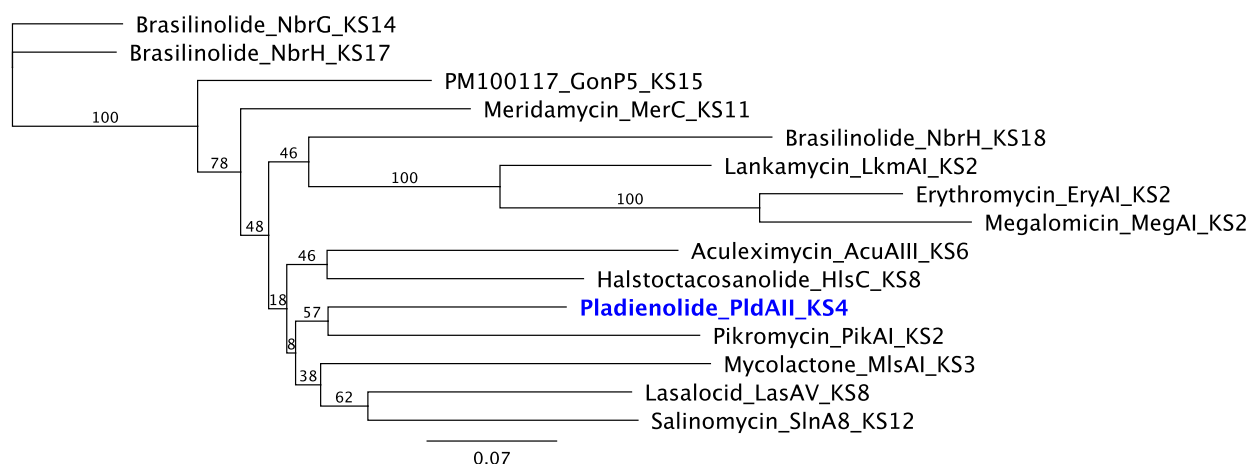


Figure 4.13: Phylogenetic tree of evolutionarily relevant ketosynthase domains from donor candidates

Looking at the bottom half first, the three KS_{n+1} domains Pld-KS4 most closely clades with are from the pikromycin, halstoctacosanolide, and aculeximycin PKSs. Their corresponding N+1 extension modules utilize the same extender unit and have the same reductive loop architecture as PldMod4 (TABLE XI). Together, this indicates that these donors may be compatible for engineering of the Pld PKS. The lasalocid and salinomycin make up the bottom-most clade. These are the only two assembly lines that do not possess reductive loop elements in the N-1 extension module, thus these KS_{n+1} domains may recognize substrates with an unreduced carbonyl at the corresponding position.

The top half, which covers a wide range of sequence distance, contains all KS_{n+1} domains from brasilinolide as well as those from PM100117, meridamycin, lankamycin, erythromycin, and megalomycin. In this group, the clade formed by KS_{n+1} from lankamycin, erythromycin, and megalomycin may be explained by substrate specificity they are the only KS domains from the top half that catalyze condensation of the growing chain with methylmalonyl-CoA. As mentioned in the discussion of the ACP_n tree, these are also the assembly lines where the N-1 is a loading module with AT_L -ACP didomain architecture. Also observed in the top half of the tree is that Nbr-KS18 from the brasilinolide PKS clades distinctly from the two other KS domains from the same assembly line (Nbr-KS14 and Nbr-KS17). This may be explained by disparate substrate specificity of the polyketide intermediate. Nbr-KS14 and Nbr-KS17 catalyze extension on an α -substituted intermediate while Nbr-KS18 accepts a polyketide intermediate without an α -substituent (TABLE XI).

Many different strategies could be employed for engineering PldMod3 with sequences from the above described donors. The simplest approach would be to swap out only the native reductive loop and replace it with the reductive loop from a donor containing a B2-type KR. However, additional modifications – such as exchanging cognate ACP_n and KS_{n+1} – may also be necessary to achieve an engineered construct with sufficient catalytic activity.

The phylogenetic disparity between Pld-ACP3 and the ACP_n domains from mycolatone, brasilinolide, and PM100117 strongly indicate that these donors would not be suitable for engineering a reductive loop swap as Pld-ACP3 may be unable to make necessary contacts with the new reductive domains. Additionally, replacement of Pld-ACP3 with any of these domains may preclude the new ACP_n from making necessary contacts with other domains in the assembly line that have been retained from the native Pld PKS.

The most promising B2-type KR donor candidates are from the aculeximycin, halstoctacosanolide, and pikromycin PKSs. This is due to the close phylogenetic relationship between their cognate KS_{n+1} domains and Pld-KS4 as well as the similar assembly line architecture in the regions their

KR_n are sourced from. These donors would be suitable not only for exchange of reductive loops, but also for the exchange of additional catalytic domains, which may be necessary to achieve functional chimeric constructs.

4.2.2.2 Cloning and purification of recombinant PldMod3

Chimeric constructs generated by swapping out the reductive loop of PldMod3 would have their activity assessed *in vitro* with synthetic extender units. This would allow determination of the relative activity of each construct as multiple will likely be made in parallel. The chimeric constructs would additionally have their activity compared to a wild-type standard. Therefore, we cloned the third extension module of the pladienolide PKS into pHis8 (Jez *et al.*, 2000) via traditional, restriction-ligation cloning methods (Figure 4.14).

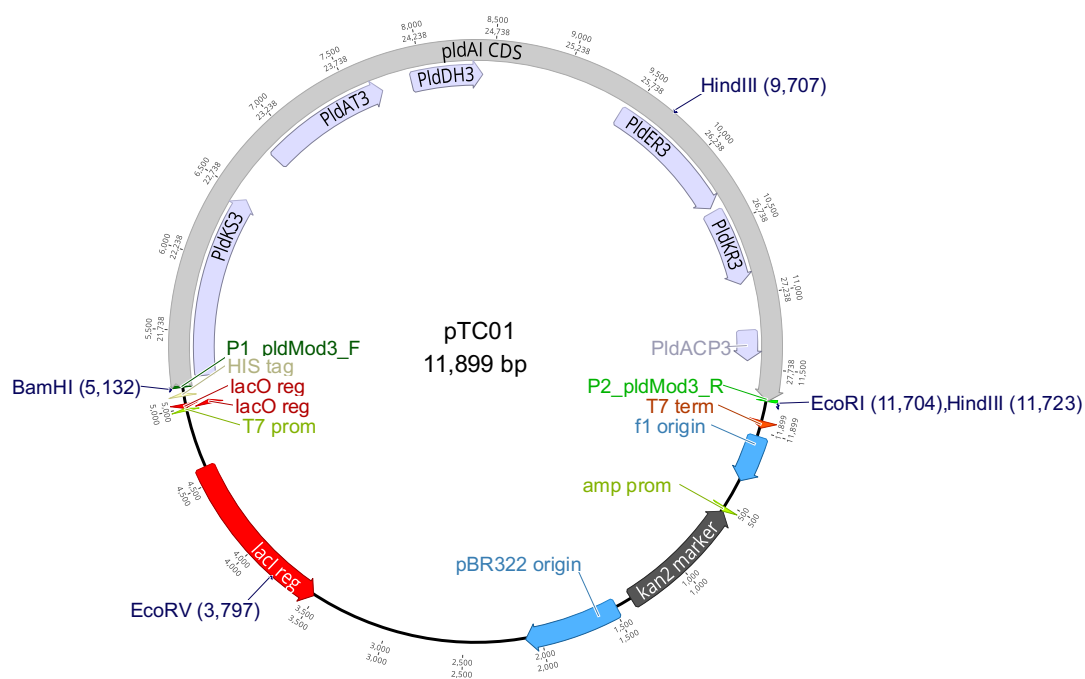


Figure 4.14 Vector map of pTC01. *pldMod3* cloned into expression vector pHis8. EcoRI and BamHI were used as cloning sites. EcoRV and HindIII sites used to screen clones.

Because the third extension module is C-terminal on the PldAI subunit, the 8x-His tag was positioned at the N-terminus of the cloned construct to preserve the subunit's C-terminal docking domain. Primers with added restriction sites (BamHI and EcoRI) were designed such that the translated protein product would include the upstream linker region which connects PldMod3 to the second module and all subsequent residues of the subunit including the docking domain.

Amplification of *pldMod3* by PCR resulted in significant non-specific amplification (Figure 4.15A), and the desired 6590 bp amplicon had to be carefully gel-excised. As PKSs are known for their highly repetitive sequences, this complication was not surprising. Following digest of the purified amplicon and pHis8 with BamHI and EcoRI, the two were ligated and cloned in *E. coli* 10-beta, yielding pTC01 (Figure 4.15B). Cloning was first attempted in *E. coli* DH5 α but was not successful, potentially due to the size of the cloned construct. Once the cloned construct was confirmed, the plasmid was then transformed into *E. coli* BAP1 for heterologous expression (Figure 4.15C).

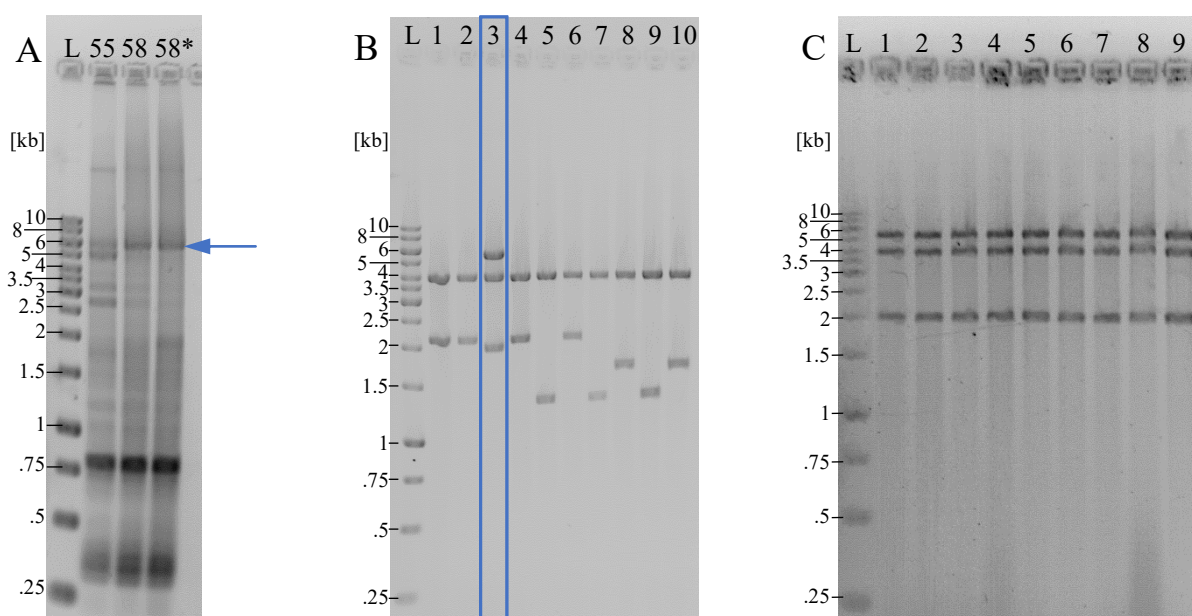


Figure 4.15: Construction of pTC01. (A) PCR of PldMod3 (B) Digest screen of *E. coli* 10-beta clones (C) Digest screen of *E. coli* BAP1 clones.

The ~232 kDa recombinant protein was successfully expressed in BAP1 (Figure 4.16). Excitingly, the protein was present in the soluble fraction of the cell lysate (lane 4) after a 4-hour induction, as well as following elution (lane 8) and desalting (lane 10), indicating that it binds to the Ni-NTA resin. However, much of the protein is still present in the lysate upon separation from the resin (lane 5), and two lower molecular weight *E. coli* proteins (~75 and ~25 kDa) appear to bind to the resin more efficiently. Thus, much of the desired recombinant protein is lost during purification, and that which has been captured by the resin requires further purification (*e.g.* using size-exclusion chromatography).

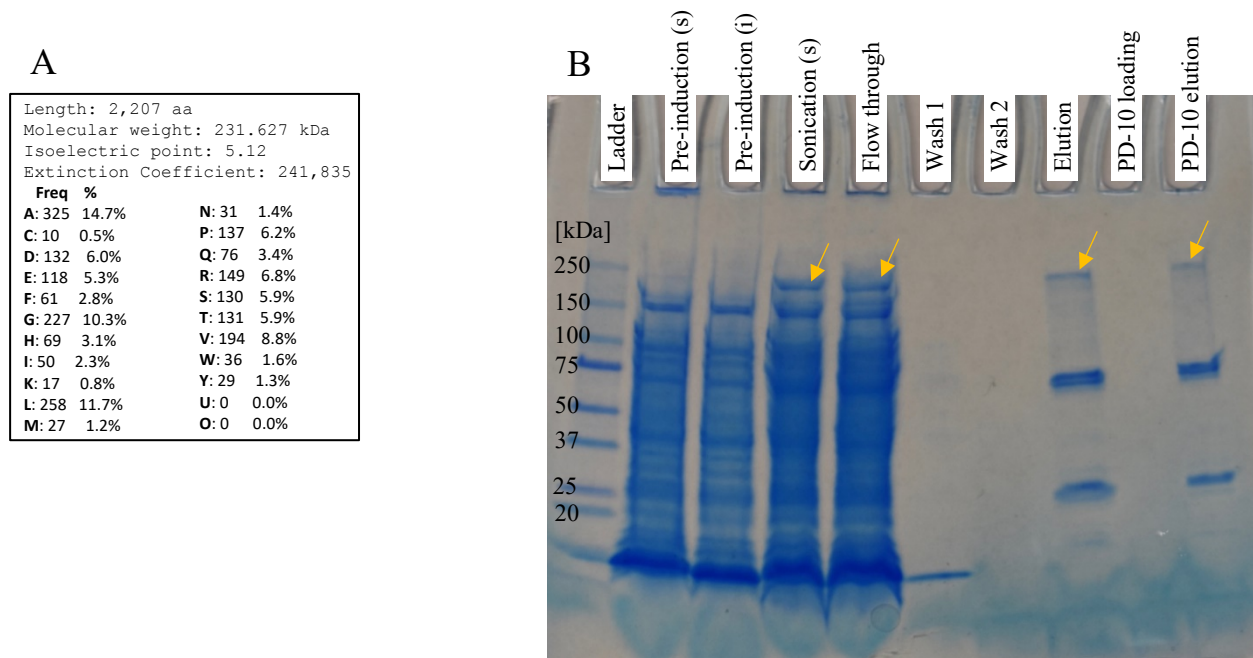


Figure 4.16: Purification of recombinant PldMod3. (A) Translated protein analysis from geneious. **(B)** SDS-PAGE gel of PldMod3 purification from 4-h induction culture. PldMod3 is indicated by a yellow arrow.

A second attempt to isolate recombinant PldMod3 in pure form was made (Figure 4.17) with several modifications, both to production and the purification process. First, the induction time was increased, in an effort to increase the overall amount of recombinant protein in the cell lysate. Two production cultures were induced in parallel for 12 h (Figure 4.17A) and 17 h (Figure 4.17B). Modifications to the purification process include increased resin bed capacity as well as gradient wash and gradient elution steps (increasing imidazole concentration gradient). Unfortunately, the modified process still did not afford isolation of PldMod3 in pure form.

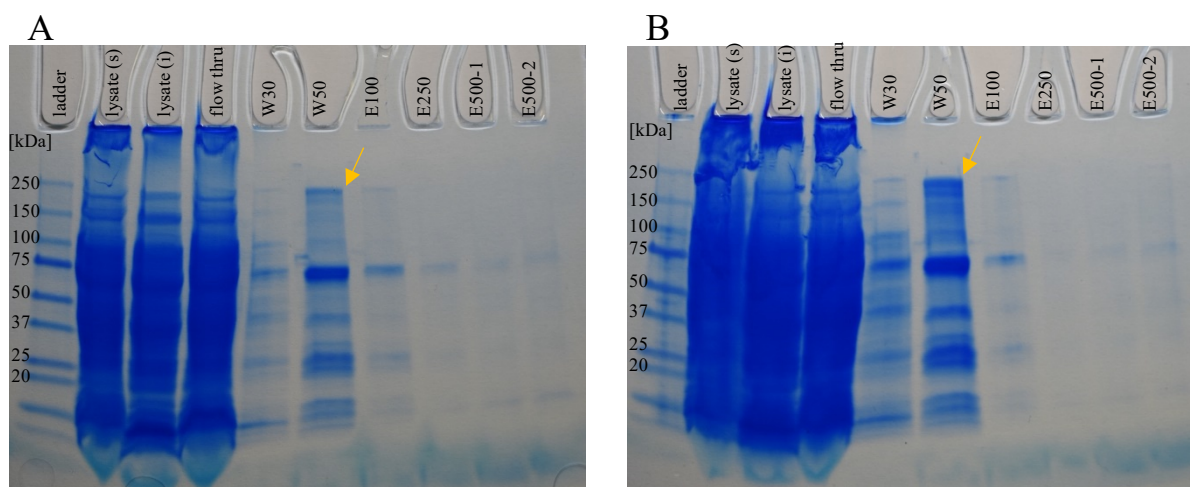


Figure 4.17: Modified purification of recombinant PldMod3. (A) 12-hour induction and **(B)** 17-hour induction. Band of recombinant PldMod3 indicated with yellow arrow.

While the increased induction time and resin bed volume improved the yield of captured recombinant protein, many more contaminating proteins were captured as well. The gradient wash and elution failed to selectively remove contaminating proteins. Future efforts to purify recombinant PldMod3 would be best served by using a 12 h induction time as the 17 h induction seems to result in far more

contaminating proteins bound to the resin with minimal gains in terms of bound recombinant PldMod3. Additionally, the resin should be washed with very low concentrations of imidazole, and a single high-concentration imidazole elution step would suffice to dislodge bound protein from the resin. As the eluent will contain many contaminating *E. coli* proteins, fast protein liquid chromatography is recommended to obtain recombinant PldMod3 in pure form.

4.3 Conclusions

Polyketide natural products provide privileged scaffolds for drug discovery, but their structures often require modification in order to translate these compounds into the clinic. This is true of the pladienolide family of polyketides, which includes several synthetic analogues that show promise as pharmaceutical therapies for cancers such as acute myeloid leukemia and chronic lymphocytic leukemia. One such synthetic analogue, 17S-FD-895, shows nearly 25-fold higher potency and more than 100-fold improved metabolic stability over the parent compound. However, the laborious chemical synthesis of this modified analogue poses a significant barrier to the realization of its therapeutic potential. In order to address this barrier, we seek to engineer the pladienolide PKS via reductive loop swap to afford fermentative production of 17S-FD-895 (Figure 4.3). As discussed in Chapter 3, harnessing the stereocontrol of PKS assembly lines remains one of the most outstanding challenges of PKS engineering, though much progress has been made.

Herein, we have identified a suite of PKS domain donors whose catalytic machinery can affect the desired modification to the pladienolide scaffold. Our donor pool is limited by the overall rarity of the desired catalytic domain, a B2-type KR, among PKS assembly lines. Phylogenetic analyses may provide insight into which donors are most suitable, but it is impossible to say which donors – if any – will provide us with an engineered construct that is structurally stable and catalytically active. These uncertainties can only be resolved through empirical analysis of chimeric constructs. That we were able to clone and heterologously express a truncated PKS subunit – PldMod3 – is a step in the right direction.

In addition to obstacles posed by PKS engineering, the work herein described additionally exemplifies fundamental challenges of working with natural products, such as instability of native producers in the laboratory setting. Once the pladienolide producing strain *Streptomyces* sp. FERM BP-7812 was obtained, it took careful analysis of many phenotypically distinct isolates to identify an isolate with suitable pladienolide production abilities. Even still, strain instability will have to be addressed in order to attain sufficient and scalable fermentative production of 17S-FD-895.

4.4 **Future Directions**

Cloning and heterologous expression of pladienolide BGC. As the native producer of PldB is phenotypically unstable under laboratory conditions, engineering of the native producer is not ideal. It was very recently reported that overexpression of the native pathway-specific activator, PldR, restored pladienolide production in the domesticated strain *Streptomyces platensis* AS6200 (Booth *et al.*, 2020). However, this strategy may not work for the strain described in this work nor does it address the underlying instability observed in our strain. As the ultimate goal of engineering the pladienolide PKS is stable production of an unnatural analogue, it is critical to have a stable host. Heterologous expression in a well-characterized *Streptomyces* host provides a promising mechanism for systematically evaluating engineered constructs and scaling fermentative production to industrial levels. Moreover, fermentation and media optimization will be important to identify best conditions for production.

Genome sequencing of *Streptomyces* sp. FERM BP-7812 isolates. Should heterologous expression of the pladienolide BGC prove difficult to achieve, engineering the native producer is an alternative route for biosynthetic production of 17S-FD-895. As was just discussed, this is not ideal given the observed instability of the strain. Genome sequencing of isolates from each of the three identified phenotype groups may provide understanding of the molecular mechanisms underpinning the observed phenotypic instability, which could in turn be leveraged to improve the native producer as a chassis.

Establish an *in vitro* assay for measuring extension activity. Once recombinant PldMod3 is isolated in pure form, it can be used to establish an assay for monitoring extension activity *in vitro*. Two such assays have been reported that may be suitable. First, the NADPH consumption assay monitors the activity of reductive loop domains in PKS assembly lines by continuous UV spectrophotometry (Lowry *et al.*, 2013). This would be the simplest assay of the two options, but the construct in pTC01 lacks a chain release mechanism (TE domain), which is necessary for removal of the extension product so that the module may engage in a subsequent round of extension. One potential solution is to leverage the preserved C-terminal docking domain for “attachment” of a free-standing TE domain fused to the complementary N-terminal docking domain from the second PldPKS subunit, PldAII.

On the other hand, the PPant ejection assay uses mass spectrometry to determine the nature of chemical groups that are covalently bound to the phosphopantetheine functionality of ACP domains (Dorrestein *et al.*, 2006). While this assay can be used to simply test whether or not chimeric constructs are active, it can also provide much more detailed information. By probing the nature of substituents attached to ACP domains, the PPant ejection assay facilitates the identification of unintended biosynthetic bottlenecks created in the engineering process. These bottlenecks may stem from compromised protein-protein interactions between native and donor domains at the site of engineering, or from the inability of native downstream domains to act on a modified substrate. While a much more versatile tool, the PPant ejection assay is significantly more laborious and expensive as measurement is performed on a mass spectrometer.

Engineering chimeric modules with donated reductive loop. There are many ways to swap modules in a PKS. For the purposes of a reductive loop swap, the most well-described method involves a synthetic “polylinker” in place of the native reductive loop (Kellenberger *et al.*, 2008; AnnaVal *et al.*, 2015). This polylinker contains multiple restriction sites which can be used to insert donor reductive loops at various junctions. The donor reductive loops can be obtained synthetically, with the varying restriction sites added onto ends via PCR. Multiple constructs can be made using a single donor loop by changing which

junction, or “cut-and-paste”, site is used for insertion of the donor DNA. Each construct can then be expressed, purified, and assayed *in vitro* to determine activity. Ideally, one of these constructs would produce a polyketide intermediate with the desired stereochemical modifications effected by the donor B2-type KR. However, should activity be too low or entirely nonexistent, additional modifications could be made in an attempt to remedy this, including alteration of ACP_n and KS_{n+1}.

Notably, PldMod3 is the terminal module of its subunit, meaning that KS_{n+1} (Pld-KS4) is located on the next assembly line subunit, PldAII. It is unknown whether maintenance of native ACP_n/KS_{n+1} interaction is equally critical when the pair is located on entirely different subunits. The reductive loop swap of PldMod3 provides a good test case for whether the specificity of this protein-protein interaction is more relaxed or more strict in domains on separate polypeptides.

4.5 Experimental Methods

4.5.1 Strain propagation and isolation of phenotypically distinct colonies

Streptomyces sp. FERM BP-7812 was received as a lyophilized pellet. The pellet was rehydrated in 300μL Bacto™ Tryptic Soy Broth (BD Biosciences) and transferred in its entirety to a 250-mL Erlenmeyer flask containing a stainless-steel spring and 50-mL Tryptic Soy Broth. The culture was placed on an orbital shaker and grown at 30°C and 180 rpm for three days. The grown liquid culture was serially diluted to yield 10-fold and 100-fold dilutions. A 100-μL aliquot of each dilution was used to inoculate solid media cultures on four media types – A1F, ISP2, MS, and NZSG. The plates were incubated at 30°C for 14 days. Isolation of single colonies was performed by picking a colony with a toothpick and stabbing the toothpick into a fresh plate of the same media type as the plate the colony was picked from. Plates containing single colonies were also incubated at 30°C until full phenotypic differentiation was achieved (about 10 days).

A1F [1% BD Difco™ Soluble Starch (BD Biosciences); 0.4% Bacto™ Yeast Extract (BD Biosciences); 0.2% Bacto™ Peptone (BD Biosciences); 0.1% Calcium Carbonate (99+% ACS reagent, Acros Organics); 0.01% Potassium Bromide (99+% KBr, Alfa Aesar); 0.004% Ferrous Sulfate Heptahydrate (MP Biomedicals); and 1.5% agar], **ISP2** [1% Bacto™ Malt Extract (BD Biosciences); 0.4% Bacto™ Yeast Extract; 0.4% Dextrose (D-Glucose) Anhydrous (Fisher Chemical); and 1.5% agar], **MS** [2% Bacto™ Soytone (BD Biosciences), 2% D-Mannitol ($\geq 98\%$, Sigma-Aldrich); and 2% agar], and **NZSG** [2% BD Difco™ Soluble Starch; 1% Dextrose (D-Glucose) Anhydrous (Fisher Chemical); 0.5% N-Z-Amine® A (Sigma-Aldrich); 0.5% Bacto™ Yeast Extract; 0.3% Calcium Carbonate (99+% ACS reagent, Acros Organics); and 1.5% agar]. Difco™ Granulated Agar (BD Biosciences) was used as the solidifying agent.

Isolates selected for cryopreservation were cultivated in a 250-mL Erlenmeyer flask containing a stainless-steel spring and 25-mL Tryptic Soy Broth. The flasks were inoculated by excising a section of the agar containing a healthy colony and transferring it to the flask with a sterile loop. Following 3 days of incubation on an orbital shaker at 30°C, cryostocks were made by mixing 1 mL of culture with 1 mL of 40% glycerol.

4.5.2 Routine strain cultivation

Liquid cultures of isolates were grown in 250 mL Erlenmeyer flasks with a stainless-steel spring and incubated on an orbital shaker at 30 °C and 200 rpm. Bacto™ Tryptic Soy Broth was used for seed cultures and **SPCGB** [5% BD Difco™ Soluble Starch (BD Biosciences); 3% Pharmamedia (ADM); 2% β -cyclodextrin (Acros Organics); 1% Dextrose (D-Glucose) Anhydrous (Fisher Chemical); and 0.1% Calcium Carbonate (99+% ACS reagent, Acros Organics)] liquid medium was used for pladienolide production cultures. Solid cultures were grown on ISP2 agar plates at 30°C, and full phenotypic differentiation among the isolates is observed after approximately 9 days of incubation under these conditions.

Isolates used in the passaging study – 0N10-n3, 0M-m4, 0I-i6, and 0I10-i3 – were cultured in 50 mL Tryptic Soy Broth as described above, with successive liquid passaging every 3 days. The initial liquid culture (“Passage 1”) was inoculated with 50 μ L of cryostock, and 1 mL of culture was used to inoculate the next 50 mL passage in a serial fashion. Visualization of phenotypic diversity in each passage was achieved by spreading 200 μ L of culture (10^{-4} dilution in Tryptic Soy Broth) on ISP2 agar plates in triplicate prior to inoculating the next passage. Passage plates were grown in a dry incubator at 30°C and photographed after 9 days of growth.

4.5.3 Genomic DNA isolation and 16S analysis

Genomic DNA was isolated from 15 out of the 30 cryopreserved isolates of FERM BP-7812 using the GenElute Bacterial Genomic DNA kit (Sigma-Aldrich). The 16S rRNA gene was PCR amplified using primers 27F (5'-AGA GTT TGA TCM TGG CTC AG-3') and 1492R (5'-CGG TTA CCT TGT TAC GAC TT-3') and DreamTaq DNA polymerase (Thermo Scientific). PCR products were purified using the DNA Clean and Concentrator-5 kit (Zymo Research) and submitted for bidirectional Sanger sequencing using the same primers.

4.5.4 IDBac

Isolate cryostocks were streaked on ISP2 plates. Healthy, single colonies were transferred to a well of a 48-well plate containing 0.750mL ISP2 agar per well. The plate was incubated at 30°C for 9 days. Using a toothpick, each colony was spread in triplicate on a MALDI target plate. Remainder of protocol carried out as described in Clark *et al.* (Clark *et al.*, 2018).

4.5.5 Pladienolide B production analysis

Seed cultures were grown as described in section 4.5.2. Production cultures were generated by inoculating 60 mL SPCGB medium with 1 mL of the seed culture and grown at 30°C for 5 days. Cultures were harvested by centrifugation, and supernatants separated from cell pellets. The supernatant was

extracted 3x with eq. vol EtOAc and dried down *via* rotary evaporation. Extracts were resuspended in MeOH (preliminary extract in Figure 4.7 was resuspended in DMSO) for HPLC analysis on an Agilent 1260 Infinity HPLC system with a Kinetex 5 μ m C18 column (100Å, 150 \times 4.6mm, Phenomenex) with a SecurityGuard ULTRA Column Guard with a C18 (4.6mm ID). Spectra were obtained by monitoring 240 nm. Solvent schemes are presented in TABLE XII.

TABLE XII: HPLC SOLVENT SCHEMES USED FOR PLADIENOLIDE B ANALYSES

(A) PRELIMINARY ANALYSIS (FIGURE 4.7)

t (min)	%A (MeOH)	%B (H ₂ O + 0.1% TFA)
0	10	90
45	75	25
60	75	25

(B) PHENOTYPE ANALYSIS (FIGURE 4.8)

t (min)	%A (MeOH)	%B (H ₂ O + 0.1% TFA)
0	10	90
26	100	0
30	100	0
31	10	90
35	10	90

(C) BLACK ISOLATE ANALYSIS (FIGURE 4.9)

t (min)	%A (MeOH)	%B (H ₂ O + 0.1% TFA)
0	10	90
30	100	0
35	100	0
36	10	90

+ 5.00min post-time for column reequilibration

A standard of pladienolide B (Cayman Chemical) was dissolved in DMSO to a concentration of 1 $\mu\text{g}/10\text{ }\mu\text{L}$. Estimated titers presented in TABLE VIII and TABLE IX were calculated *via* calibration curve. The curve is shown in Figure 4.18 was constructed using data from every injection of PldB standard over many experiments (TABLE XIII).

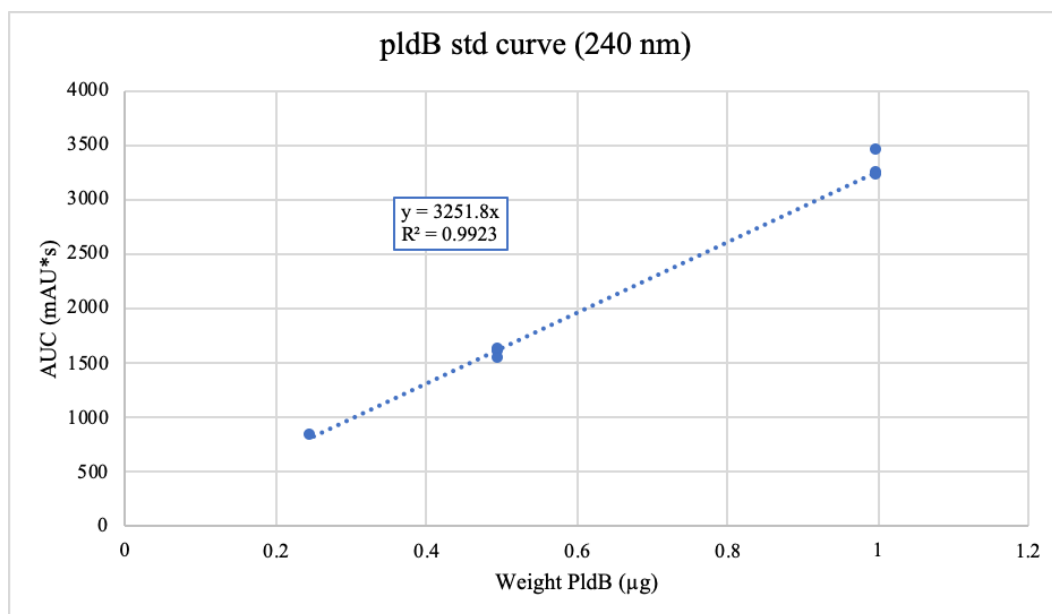


Figure 4.18 Pladienolide B calibration curve

TABLE XIII: PLADIENOLIDE B STANDARD INJECTIONS

Weight pld B std (μg)	AUC (mAU*s)
1	3228
1	3193
1	3431
0.5	1584
0.5	1513
0.5	1593
0.25	802

4.5.6 Sequence analysis of B2-type ketoreductase donor candidates

Donor candidate B2-type KR domains were found using ClusterCAD. The KR_{B2} from the first extension module of the erythromycin PKS was used as a reference sequence for a blastp search of the database. Those hits that corresponded to a confirmed B2-type KR were identified and sequences of relevant KR_n, ACP_n, and KS_{n+1} were copied from the ClusterCAD database and imported into Geneious.

Phylogenetic trees of the KR, ACP, and KS domains were generated in Geneious® 11.1.5. The donor sequences and native pladienolide sequences of each of the three domains were first aligned using the ClustalW algorithm with the following specifications – cost matrix: GONNET; gap open cost: 3.0; and gap extend cost: 1.8. These alignments were used to build a phylogenetic tree with the following specifications – genetic distance model: Jukes-Cantor; tree build method: neighbor joining; no specified outgroup. Tree resampling was performed with the following specifications: resampling method: Bootstrap; random seed: 410,513; number of replicates: 100; create consensus tree; Support threshold: 0%.

4.5.7 Construction of pTC01

The third extension module of the pladienolide PKS was amplified from gDNA (isolate 0I10-i3) by PCR (TABLE XIV) using PrimeSTAR Max DNA Polymerase (Takara Bio) with primers P1_pldMod3_F_BamHI (5'-*CTA TCT GGA TCC CTT CTG GGT TCC CGC GAG*-3', the restriction site is underlined, and extra bases to facilitate cleavage are in italics) and P2_pldMod3_R_EcoRI (5'-*TAT ATA GAA TTC GGT CAC ACG TCC CCT ACC T*-3').

The PCR product was gel excised and purified using Zymoclean Gel DNA Recovery Kit (Zymo Research). EcoRI-HF (NEB) and BamHI-HF (NEB) were used to digest both the cleaned PCR product and vector pHis8. The two fragments were ligated with T4 DNA Ligase (Thermo Scientific) and cloned in *E. coli* 10-beta (Invitrogen) to yield pTC01. Clones were screened using HindIII-HF (NEB) and EcoRV-HF (NEB).

TABLE XIV: REACTION SETUP AND THERMOCYCLER PROGRAMMING FOR
AMPLIFICATION OF *PLDMOD3* FROM GENOMIC DNA

Reagent	μL per reaction	
0110-i3 gDNA	2.5 (50 ng)	
2X PrimeSTAR Max DNA Polymerase	25	
Forward primer (10 μM)	1.5	
Reverse primer (10 μM)	1.5	
Water	17	

PCR step	Temperature ($^{\circ}\text{C}$)	Time	
Initial Denaturation	N/A ^a	N/A ^a	
Denaturation	95	0'10"	x35
Annealing	55/58 ^b	0'05"	
Extension	72	1'06"	
Final Extension	N/A ^a	N/A ^a	
Hold	12	∞	

^aPolymerase does not require initial denaturation or final extension steps.

^bTwo different annealing temperatures were tested.

4.5.8 Purification of recombinant PldMod3

Construct pTC01 was electroporated into *E. coli* BAP1 (Pfeifer *et al.*, 2001) for protein production. A 1 L culture of *E. coli* BAP1 + pTC01 was grown in Terrific Broth (2.4% yeast extract, 1.2% tryptone, 0.4% glycerol, 1.254% K_2HPO_4 , 0.231% KH_2PO_4) with kanamycin selection (50 $\mu\text{g}/\text{mL}$) at 37°C and 200 rpm, until $\text{OD}_{600}=0.7$, at which point gene expression was induced by the addition of IPTG (250 μM final concentration). Incubation was resumed at a decreased temperature of 18°C for 4, 12, or 17 hours. Cells were harvested by centrifugation and N-terminal 8xHis-tagged PldMod3 obtained from the 4-h induction (Figure 4.16B) was purified by Ni-NTA affinity chromatography, as follows.

First, the cell pellet was resuspended in 60 mL ice cold lysis buffer (50 mM NaH₂PO₄, 300 mM NaCl, 20 mM imidazole, 10% glycerol, 1% Tween20, 1 mg/mL lysozyme, pH=7.6), and the suspension was left to incubate on ice for 30 min with occasional stirring. Cells were lysed by sonication, and cell debris was separated from the soluble lysate via centrifugation (11627 rcf, 40 min, 4°C). The soluble lysate was transferred to a beaker, on ice, containing a stir bar and 1 mL of Ni-NTA resin (Protino® Ni-NTA agarose, Macherey-Nagel), which had been pre-equilibrated with lysis buffer. The lysate-resin mixture was allowed to incubate for 1 hour while shaking on a benchtop shaker at 220 rpm. Following incubation, the lysate-resin mixture was poured into a disposable polypropylene column (Bio-Rad Econo-Pac Disposable Chromatography Column, 25 mL capacity). The lysate was allowed to flow through the column as the resin settled. The resin was washed three times with 20 mL of ice-cold wash buffer (50 mM NaH₂PO₄, 300 mM NaCl, 20 mM imidazole, 10% glycerol, pH=7.6). The protein was then displaced from the resin with 2.5 mL elution buffer (50 mM NaH₂PO₄, 300 mM NaCl, 500 mM imidazole, 10% (v/v) glycerol, pH=7.6) and eluted into a 15 mL falcon tube. Buffer exchange to storage buffer (50 mM NaH₂PO₄, 10% glycerol, pH=7.6) was performed via gel filtration chromatography using disposable PD-10 columns (GE Healthcare) containing Sephadex G-25M resin. The purified enzyme was analysed by SDS-PAGE using Mini-PROTEAN TGX Precast Gel, 4-20% (Bio-Rad) in Tris/Glycin/SDS Buffer (Bio-Rad).

For the 12- and 17-hour induction samples (Figure 4.17), purification was carried out as described above with the following modifications: Ni-NTA resin bed increased to 2 mL, a three-step wash with an increasing imidazole concentration gradient (20, 30, and 50 mM), and a three step elution with an increasing imidazole concentration gradient (100, 250, and 500 mM).

4.6 References Cited

- Abelson, J. (2017). A close-up look at the spliceosome, at last. *Proceedings of the National Academy of Sciences*, 114(17), 4288–4293. doi: 10.1073/pnas.1700390114
- Annaval, T., Paris, C., Leadlay, P. F., Jacob, C., & Weissman, K. J. (2015). Evaluating Ketoreductase Exchanges as a Means of Rationally Altering Polyketide Stereochemistry. *ChemBioChem*, 16, 1357–1364. doi: 10.1002/cbic.201500113

- Booth, T. J., Kalaitzis, J. A., Vuong, D., Crombie, A., Lacey, E., Piggott, A. M., & Wilkinson, B. (2020). Production of novel pladienolide analogues through native expression of a pathway-specific activator. *Chemical Science*. doi: 10.1039/d0sc01928c
- Chater, K. F., & Chandra, G. (2006). The evolution of development in *Streptomyces* analysed by genome comparisons. *FEMS Microbiology Reviews*, 30, 651–672. doi: 10.1111/j.1574-6976.2006.00033.x
- Clark, C. M., Costa, M. S., Sanchez, L. M., & Murphy, B. T. (2018). Coupling MALDI-TOF mass spectrometry protein and specialized metabolite analyses to rapidly discriminate bacterial function. *Proceedings of the National Academy of Sciences*, 115(19), 4981–4986. doi: 10.1073/pnas.1801247115
- Cretu, C., Agrawal, A. A., Cook, A., Will, C. L., Fekkes, P., Smith, P. G., Lührmann, R., Larsen, N., Buonamici, S., & Pena, V. (2018). Structural Basis of Splicing Modulation by Antitumor Macrolide Compounds. *Molecular Cell*, 70, 265–273. doi: 10.1016/j.molcel.2018.03.011
- Crews, L. A., Balaian, L., Delos Santos, N. P., Leu, H. S., Court, A. C., Lazzari, E., Sadarangani, A., Zipeto, M. A., La Clair, J. J., Villa, R., Kulidjian, A., Storb, R., Morris, S. R., Ball, E. D., Burkart, M. D., & Jamieson, C. H. M. (2016). RNA Splicing Modulation Selectively Impairs Leukemia Stem Cell Maintenance in Secondary Human AML. *Cell Stem Cell*, 19(5), 599–612. doi: 10.1016/j.stem.2016.08.003
- Dorrestein, P. C., Bumpus, S. B., Calderone, C. T., Garneau-Tsodikova, S., Aron, Z. D., Straight, P. D., Kolter, R., Walsh, C. T., & Kelleher, N. L. (2006). Facile Detection of Acyl and Peptidyl Intermediates on Thiotemplate Carrier Domains via Phosphopantetheinyl Elimination Reactions during Tandem Mass Spectrometry. *Biochemistry*, 45, 12756–12766. doi: 10.1021/bi061169d
- Eng, C. H., Backman, T. W. H., Bailey, C. B., Magnan, C., García Martín, H., Katz, L., Baldi, P., & Keasling, J. D. (2018). ClusterCAD: A computational platform for type I modular polyketide synthase design. *Nucleic Acids Research*, 46(D1), D509–D515. doi: 10.1093/nar/gkx893
- Eskens, F. A. L. M., Ramos, F. J., Burger, H., O'Brien, J. P., Piera, A., De Jonge, M. J. A., Mizui, Y., Wiemer, E. A. C., Carreras, M. J., Baselga, J., & Tabernero, J. (2013). Phase I pharmacokinetic and pharmacodynamic study of the first-in-class spliceosome inhibitor E7107 in patients with advanced solid tumors. *Clinical Cancer Research*, 19(22), 6296–6304. doi: 10.1158/1078-0432.CCR-13-0485
- Finci, L. I., Zhang, X., Huang, X., Zhou, Q., Tsai, J., Teng, T., Agrawal, A., Chan, B., Irwin, S., Karr, C., Cook, A., Zhu, P., Reynolds, D., Smith, P. G., Fekkes, P., Buonamici, S., & Larsen, N. A. (2018). The cryo-EM structure of the SF3b spliceosome complex bound to a splicing modulator reveals a pre-mRNA substrate competitive mechanism of action. *Genes and Development*, 32(3–4), 309–320. doi: 10.1101/gad.311043.117
- Gao, X., Wang, P., & Tang, Y. (2010). Engineered polyketide biosynthesis and biocatalysis in *Escherichia coli*. *Applied Microbiology and Biotechnology*, 88(6), 1233–1242. doi: 10.1007/s00253-010-2860-4

- H3 Biomedicine Inc.(June 2020) *A Phase I Study to evaluate H3B-8800 in participants with Myelodysplastic Syndromes, Acute Myeloid Leukemia, and Chronic Myelomonocytic Leukemia*. U.S. National Library of Medicine, National Institutes of Health. <https://clinicaltrials.gov/ct2/show/NCT02841540>
- Hong, D. S., Kurzrock, R., Naing, A., Wheler, J. J., Falchook, G. S., Schiffman, J. S., Faulkner, N., Pilat, M. J., O'Brien, J., & LoRusso, P. (2014). A phase I, open-label, single-arm, dose-escalation study of E7107, a precursor messenger ribonucleic acid (pre-mRNA) spliceosome inhibitor administered intravenously on days 1 and 8 every 21 days to patients with solid tumors. *Investigational New Drugs*, 32(3), 436–444. doi: 10.1007/s10637-013-0046-5
- Jenke-Kodama, H., Börner, T., & Dittmann, E. (2006). Natural Biocombinatorics in the Polyketide Synthase Genes of the Actinobacterium *Streptomyces avermitilis*. *PLOS Computational Biology*, 2(10), 1210–1218. doi: 10.1371/journal.pcbi.0020132
- Jez, J. M., Ferrer, J. L., Bowman, M. E., Dixon, R. A., & Noel, J. P. (2000). Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase. *Biochemistry*, 39(5), 890–902. doi: 10.1021/bi991489f
- Kashyap, M. K., Kumar, D., Villa, R., Clair, J. J. La, Benner, C., Sasik, R., Ghia, E. M., Rassenti, L. Z., Kipps, T. J., Burkart, M. D., & Castro, J. E. (2015). Targeting the spliceosome in chronic lymphocytic leukemia with the macrolides FD-895 and pladienolide-B. *Haematologica*, 100(7), 945–954. doi: 10.3324/haematol.2014.122069
- Keatinge-Clay, A. T., & Stroud, R. M. (2006). The Structure of a Ketoreductase Determines the Organization of the b-Carbon Processing Enzymes of Modular Polyketide Synthases. *Structure*, 14(April), 737–748. doi: 10.1016/j.str.2006.01.009
- Keatinge-Clay, A. T. (2007). A Tylosin Ketoreductase Reveals How Chirality Is Determined in Polyketides. *Chemistry & Biology*, 14(August), 898–908. doi: 10.1016/j.chembiol.2007.07.009
- Kellenberger, L., Galloway, I. S., Sauter, G., Böhm, G., Hanefeld, U., Cortés, J., Staunton, J., & Leadlay, P. F. (2008). A Polylinker Approach to Reductive Loop Swaps in Modular Polyketide Synthases. *ChemBioChem*, 9, 2740–2749. doi: 10.1002/cbic.200800332
- Lagiseti, C., Pourpak, A., Jiang, Q., Cui, X., Goronga, T., Morris, S. W., & Webb, T. R. (2008). Antitumor Compounds Based on a Natural Product Consensus Pharmacophore. *Journal of Medicinal Chemistry*, 41, 6220–6224. doi: 10.1021/jm8006195
- Lagiseti, C., Yermolina, M. V., Sharma, L. K., Palacios, G., Prigaro, B. J., & Webb, T. R. (2014). Pre-mRNA Splicing-Modulatory Pharmacophores: The Total Synthesis of Herboxidiene, a Pladienolide–Herboxidiene Hybrid Analog and Related Derivatives. *ACS Chemical Biology*, 9, 643–648. doi: 10.1021/cb400695j
- Lee, S. C. W., & Abdel-Wahab, O. (2016). Therapeutic targeting of splicing in cancer. *Nature Medicine*, 22(9), 976–986. doi: 10.1038/nm.4165
- León, B., Kashyap, M. K., Chan, W. C., Krug, K. A., Castro, J. E., La Clair, J. J., & Burkart, M. D. (2017). A Challenging Pie to Splice: Drugging the Spliceosome. *Angewandte Chemie - International Edition*, 56, 12052–12063. doi: 10.1002/anie.201701065

- Lowry, B., Robbins, T., Weng, C.-H., O'Brien, R. V., Cane, D. E., & Khosla, C. (2013). In Vitro Reconstitution and Analysis of the 6-Deoxyerythronolide B Synthase. *Journal of the American Chemical Society*, 135, 16809–16812. doi: 10.1021/ja409048k
- Machida, K., Arisawa, A., Takeda, S., Tsuchida, T., Aritoku, Y., Yoshida, M., & Ikeda, H. (2008). Organization of the Biosynthetic Gene Cluster for the Polyketide Antitumor Macrolide, Pladienolide, in *Streptomyces platensis* Mer-11107. *Bioscience, Biotechnology, and Biochemistry*, 72(11), 2946–2952. doi: 10.1271/bbb.80425
- Manteca, Á., & Yagüe, P. (2018). *Streptomyces* differentiation in liquid cultures as a trigger of secondary metabolism. *Antibiotics*, 7(2), 41. doi: 10.3390/antibiotics7020041
- Novoyatleva, T., Tang, Y., Rafalska, I., & Stamm, S. (2006). Pre-mRNA Missplicing as a Cause of Human Disease. In P. Jeanteur (Ed.), *Alternative Splicing and Disease* (1st ed., pp. 27–46). Heidelberg: Springer-Verlag Berlin Heidelberg.
- Østergaard, L. H., Kellenberger, L., Cortés, J., Roddis, M. P., Deacon, M., Staunton, J., & Leadlay, P. F. (2002). Stereochemistry of Catalysis by the Ketoreductase Activity in the First Extension Module of the Erythromycin Polyketide Synthase. *Biochemistry*, 41, 2719–2726. doi: 10.1021/bi0117605
- Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E., Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E., & Khosla, C. (2001). Biosynthesis of Complex Polyketides in a Metabolically Engineered Strain of *E. coli*. *Science*, 291(5509), 1790–1792.
- Sakai, T., Sameshima, T., Matsufuji, M., Kawamura, N., Dobashi, K., & Mizui, Y. (2004). Pladienolides, New Substances from Culture of *Streptomyces platensis* Mer-11107: I. Taxonomy, Fermentation, Isolation, and Screening. *The Journal of Antibiotics*, 57(3), 173–179.
- Seki-Asano, M., Okazaki, T., Yamagishi, M., Sakai, N., Takayama, Y., Hanada, K., Morimoto, S., Takatsuki, A., & Mizoue, K. (1994). Isolation and Characterization of a New 12-Membered Macrolide FD-895. *The Journal of Antibiotics*, 47(12), 1395–1401.
- Siskos, A. P., Baerga-ortiz, A., Bali, S., Stein, V., Mamdani, H., Spiteller, D., Popovic, B., Spencer, J. B., Staunton, J., Weissman, K. J., & Leadlay, P. F. (2005). Molecular Basis of Celmer's Rules: Stereochemistry of Catalysis by Isolated Ketoreductase Domains from Modular Polyketide Synthases. *Chemistry & Biology*, 12, 1145–1153. doi: 10.1016/j.chembiol.2005.08.017
- Villa, R., Mandel, A. L., Jones, B. D., La Clair, J. J., & Burkart, M. D. (2012). Structure of FD-895 revealed through total synthesis. *Organic Letters*, 14(21), 5396–5399. doi: 10.1021/ol3023006
- Weissman, K. J. (2016). Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology. *Natural Product Reports*, 33, 203–230. doi: 10.1039/c5np00109a
- Wu, G., Fan, L., Edmonson, M. N., Shaw, T., Boggs, K., Easton, J., Rusch, M. C., Webb, T. R., Zhang, J., & Potter, P. M. (2018). Inhibition of SF3B1 by molecules targeting the spliceosome results in massive aberrant exon skipping. *RNA*, 24(12), 1056–1066. doi: 10.1261/rna.068544.118
- Zheng, J., & Keatinge-Clay, A. T. (2013). The status of type I polyketide synthase ketoreductases. *Medicinal Chemistry Communications*, 4, 34–40. doi: 10.1039/c2md20191g

Appendix A: Herboxidiene

This appendix details 1) propagation of *Streptomyces chromofuscus* A7847 – the native herboxidiene producer (Miller-Widemann *et al.*, 1992) – following receipt of the strain from the American Type Culture Collection (ATCC); 2) cultivation and validation of herboxidiene production; and 3) efforts to clone the herboxidiene BGC via RecET/ExoCET (Wang *et al.*, 2016; Wang *et al.*, 2018).

A.1 Cultivation of *Streptomyces chromofuscus*

A cryostock of *Streptomyces chromofuscus* A7847 was obtained from ATCC as *Streptomyces chromofuscus* (ATCC® 49982™) and propagated per ATCC's instructions.² In short, an **ISP2** agar plate [1% Bacto™ Malt Extract (BD Biosciences); 0.4% Bacto™ Yeast Extract; 0.4% Dextrose (D-Glucose) Anhydrous (Fisher Chemical); and 1.5% Difco™ Granulated Agar (BD Biosciences)] was inoculated by dipping a sterile inoculation loop into the cryostock and streaking across the plate. The plates were sealed with parafilm and placed in a dry incubator at 26°C for four days.

After four days of incubation, significant growth was observed on the plate. Colonies appeared homogeneous and off-white to tan in color. The majority of colonies exhibited spore formation, evidenced by white specks on and around the colonies. Using a sterile inoculation loop, a single sporulating colony was scraped from the surface of the plate and transferred to a 250 mL Erlenmeyer flask with a stainless-steel spring and 50 mL of ISP2 broth. The flask was placed on a rotary shaker at 26°C, 200 rpm for six days. After six days of growth, the flask contained a thick culture with visible mycelium, which was used to make 10 cryostocks by mixing the culture 1:1 with 40% glycerol.

The remaining culture was used to inoculate several ISP2 plates, which were placed in a dry incubator at 30°C. After one week, examination of the plates revealed homogeneous growth of the strain as well as significant sporulation (Figure A.1). Thus, with no signs of contamination, the cryostocks made in-house could be used for experiments involving *S. chromofuscus*.

² <https://www.atcc.org/Products/All/49982.aspx#documentation>

Appendix A (continued)



Figure A.1: Sporulating plate of *Streptomyces chromofuscus* under a microscope. Image taken through the eyepiece of an AO Spencer Spectroscopic Microscope (c. 1950) with an iPhone camera.

A.2 Validation of herboxidiene production in liquid culture

A.2.1 Cultivation and extraction

A seed culture was prepared in a 250-mL Erlenmeyer flask with a stainless still spring by inoculating 50-mL ISP2 broth with 100- μ L of *S. chromofuscus* cryostock. The seed culture was placed on an orbital shaker at 30 °C and 200 rpm for 5 days. Two production cultures were prepared by inoculating 50 mL ISP2 broth with 1-mL of *S. chromofuscus* seed culture in a 250 mL Erlenmeyer flask with a stainless still spring. The two production cultures were placed on an orbital shaker at 30 °C, 200 rpm for 5 days. The cultures were then harvested by centrifugation, and the pooled supernatant was extracted 3x with an equivalent volume of EtOAc then dried via rotary evaporation.

Appendix A (continued)

A.2.2 HPLC analysis and method development

A herboxidiene standard was obtained from Focus Biomolecules. The 200 μg standard was dissolved in 2 mL DMSO to give a 1 $\mu\text{g}/10\ \mu\text{L}$ stock. All HPLC analysis was carried out on an Agilent 1260 Infinity using a C18 Kinetex® column (5 μm , 100 \AA , 150 x 4.6 mm, Phenomenex) with a SecurityGuard™ Cartridge Holder (10 mm id, Phenomenex) and Core-Shell C18 SecurityGuard SemiPrep Cartridge (10 x 10 mm id, Phenomenex). All experiments monitored $\lambda=238\ \text{nm}$ and used a flow rate of 1.000 mL/min. Solvent schemes changed over the course of method development, and the mobile phase employed with each experiment is described with its respective chromatograms below.

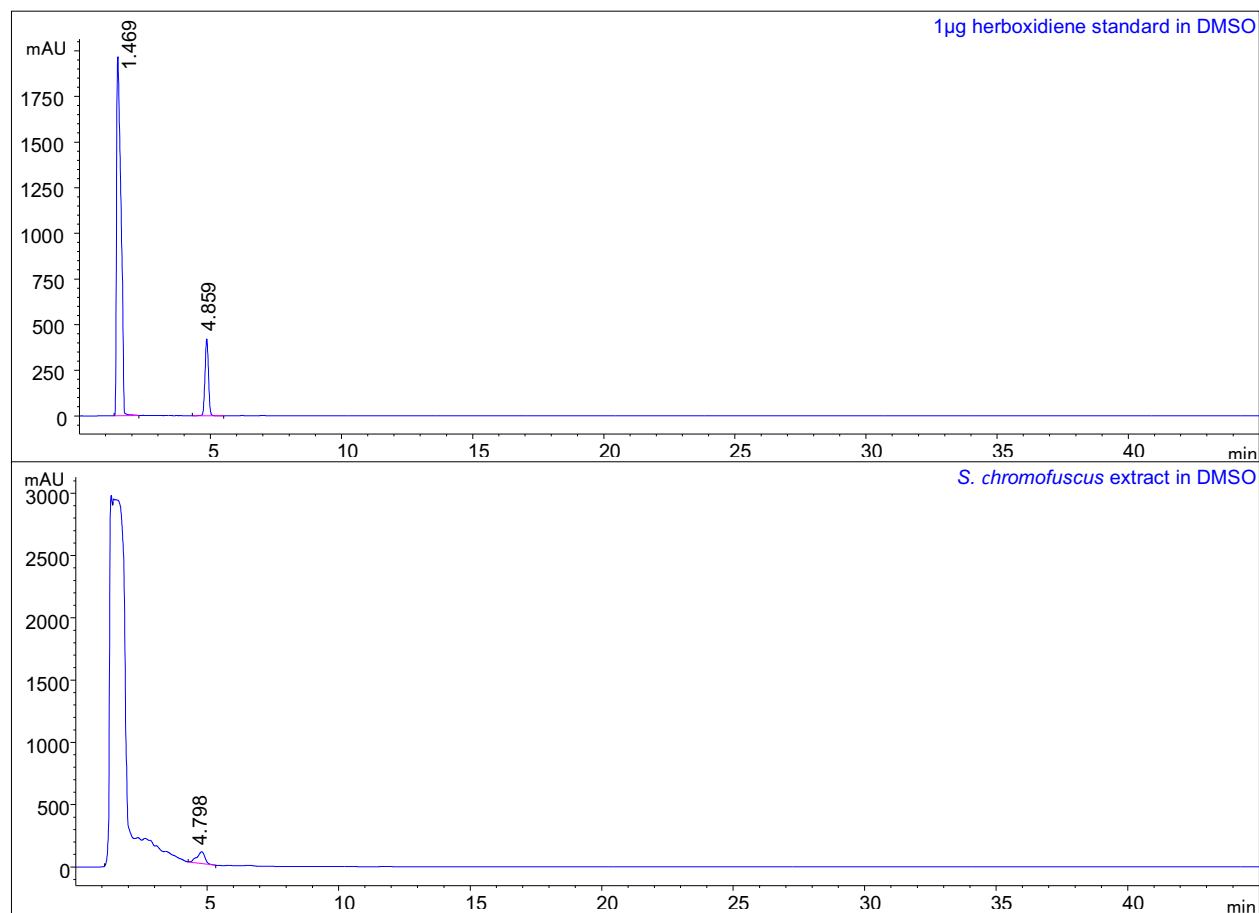
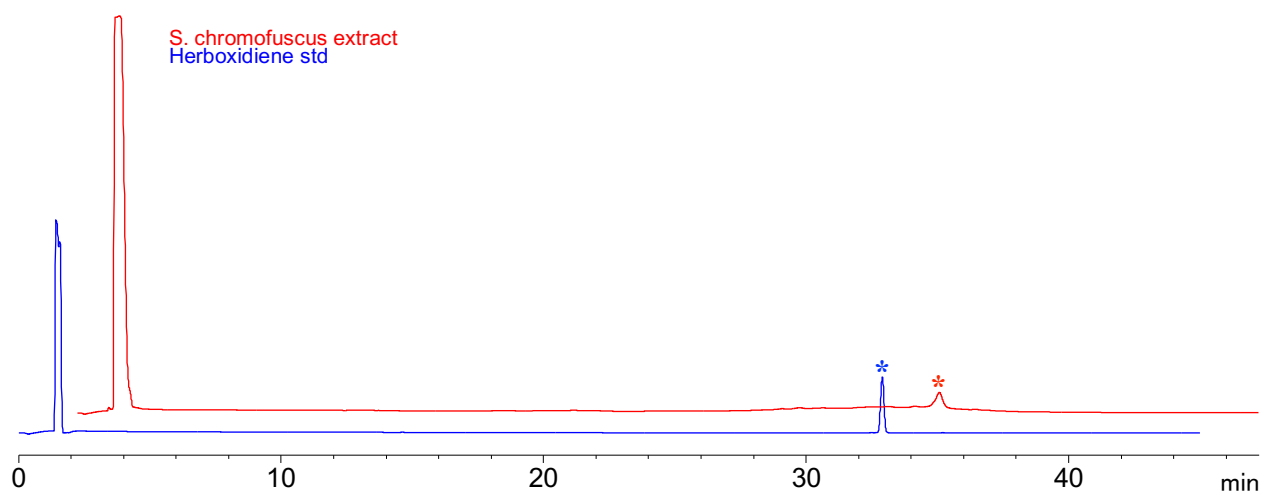


Figure A.2: Preliminary analysis of Herboxidiene production.

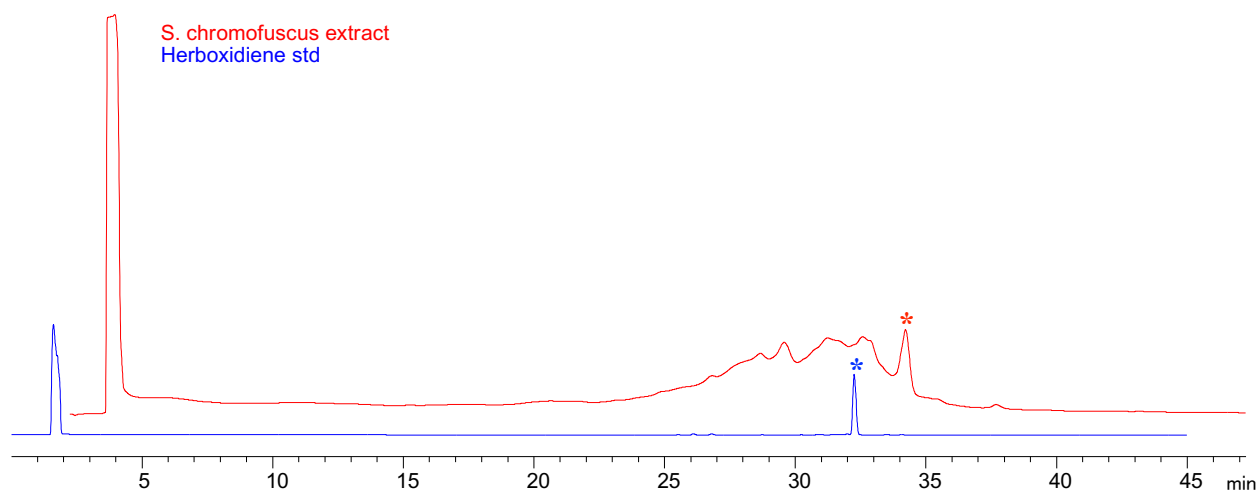
Solvent scheme: 75% MeOH/25% H₂O + 0.1% TFA, Isocratic.

Herboxidiene elutes ~4.8 min, and extract peaks are poorly resolved.

Appendix A (continued)**Figure A.3: Second analysis of herboxidiene(*) production.****FIGURE A.3/A.4 SOLVENT SCHEME**

time (min)	%A (MeOH)	%B (H ₂ O+0.1% TFA)
0	10	90
30	75	25
45	75	25

Herboxidiene elutes at ~32.9 min for both standard and extract. The extract is too dilute, making it hard to judge the resolving power of the method (Figure A.3). Thus, there were no changes to solvent scheme between Figure A.3 and Figure A.4, but extract was more concentrated.

**Figure A.4: Third analysis of herboxidiene(*) production.**

Appendix A (continued)

While the herboxidiene peak is still readily observable in Figure A.4, it is not well resolved in a more concentrated extract. There is significant overlap between the left side of the herboxidiene peak and the various, poorly resolved peaks of the compounds eluting before it. This method facilitates qualitative analysis of herboxidiene in an extract. However, if the extract is more concentrated, the poor resolution of the other compounds in the sample will likely prevent quantitative analysis of herboxidiene in an extract.

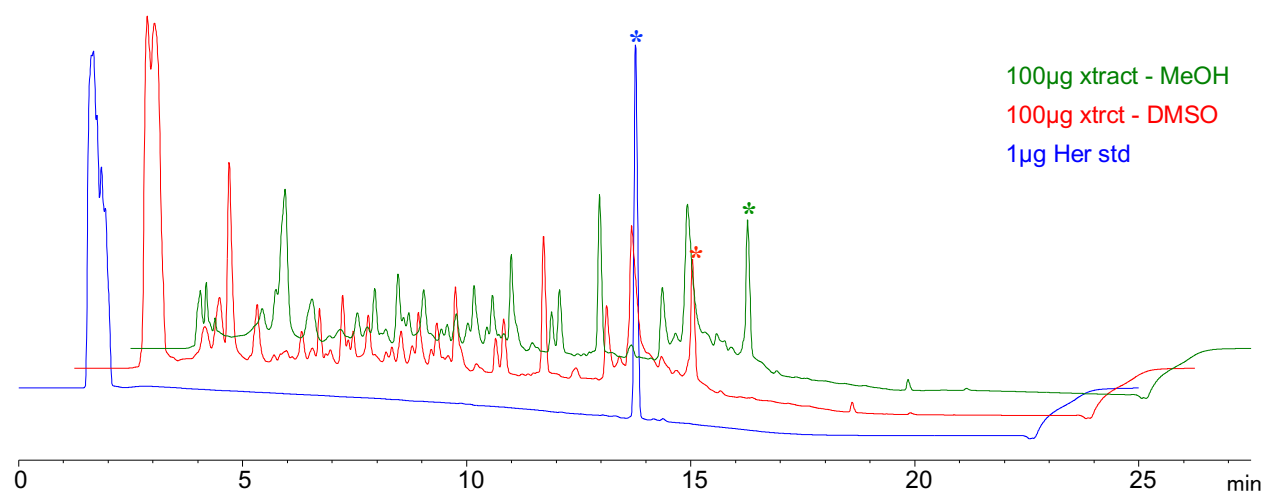


Figure A.5: Fourth analysis of herboxidiene production.

FIGURE A.5 SOLVENT SCHEME

time (min)	%A (MeOH)	%B (H ₂ O+0.1% FA)
0	10	90
10	75	25
15	100	0
20	100	0
21	10	90
25	10	90

The solvent scheme was updated by employing a steeper gradient, using formic acid (FA) instead of TFA, and the addition of a “re-equilibration” period at the end of each run. Compounds eluting upstream of herboxidiene now have well resolved peaks. It is unclear whether this is due to the steeper gradient or addition of the re-equilibration period. Achieving the same resolution of peaks over a longer time period would be ideal as the chromatogram is quite crowded.

Appendix A (continued)

Additionally, the fourth analysis (Figure A.5) was intended to determine whether extracts are best solved in DMSO or MeOH. To date, extracts have been solved in DMSO as the herboxidiene standard was solved in DMSO. The standard was dissolved in DMSO as herboxidiene is known to be poorly soluble in H₂O and alcohol. Complete solvation of the standard is essential for quantitative analyses.

The two extract solvation methods yielded nearly identical chromatograms, though the DMSO seems to provide better resolution of 3 peaks eluting over 2.5-5min. MeOH solvation allows for easy dry-down of unused extract, which is of greater value. Thus, future experiments can utilize MeOH as the solvent for resuspending extracts.

A.3 Direct cloning of herboxidiene BGC *via* RecET

RecET is a direct cloning method which utilizes homologous recombination to capture specified genomic regions into various expression plasmids (Wang *et al.*, 2016). First, restriction sites must be identified on either side of the region to be cloned. High molecular weight gDNA is then digested with the appropriate restriction enzyme to liberate this region from the genome. Finally, the digested gDNA and a linear cloning vector with homology arms to the liberated ends of the BGC are electroporated into *E.coli* GB05RedTrfA, where the two undergo linear-linear homologous recombination *via* inducible RecET.

A.3.1 Design of cloned construct

The herboxidiene BGC (Shao *et al.*, 2012) is 53 kbp (Accession #: JN671974, bases: 21,446-74,480) and consists of seven genes including a transcriptional regulator (*herA*), three PKS subunits (*herB*, *herC*, and *herD*), and three tailoring enzymes (*herE*, *herF*, and *herG*).



Figure A.6 Schematic of the 53 kb Herboxidiene BGC.

Appendix A (continued)

Because the pathway is to be heterologously expressed in *Burkholderia* sp. and promoters are to be engineered, the cloning plan described here excludes native regulator *herA*. Suitable restriction sites were identified in geneious by excluding enzymes that cut in the 50.7 kbp region (bases 23,785-74,480) between the start of *herB* and the end of *herG*. *SfcI* has the closest sites of any other commercially available enzyme to both of the desired cloning boundaries.

The upstream *SfcI* site (bp 23,160) results in the exclusion of *herA* and leaves 626 bp upstream of *herB* (Figure A.7A). The downstream *SfcI* site (bp 76,026) is located ~1.5 kbp after the end of *herG*, containing two genes outside of the *her* BGC (Figure A.7B). They are a full gene for a GCN5-like *N*-acetyltransferase and a partial gene for a hypothetical protein, however these genes are oriented in the opposite direction of the *her* cluster and will not be transcribed.

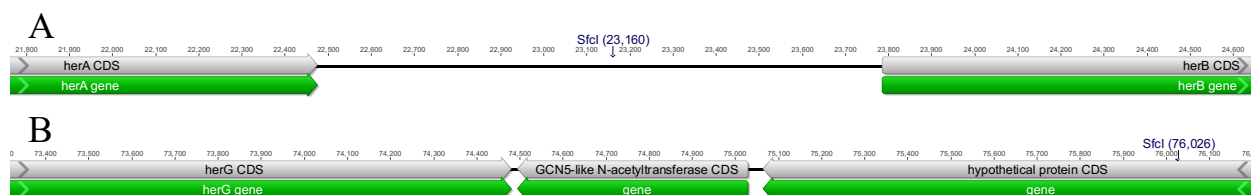


Figure A.7 *SfcI* restriction sites. (A) upstream (B) downstream

A.3.2 Preparation of linear cloning vector *via* four step process

The *her* cluster shows high sequence self-homology, as expected for a Type I PKS. Therefore, pBAC2015 has been selected from the three cloning vectors described in the RecET protocol because it is a single-copy bacterial artificial chromosome (BAC) vector and is therefore tolerant of highly repetitive sequences. Unlike the other cloning vectors described in the RecET protocol (Wang *et al.*, 2016), the size of pBAC2015 (9327 bp) precludes direct addition of homology arms *via* PCR, so the linear cloning vector is prepared through a multistep process. The process is described in full, followed by results.

Appendix A (continued)

Step 1 – **PCR 1:** Amplification of counter-selectable cassette from pBR322-amp-ccdB-rpsL with added BamHI restriction sites and BGC homology arms.

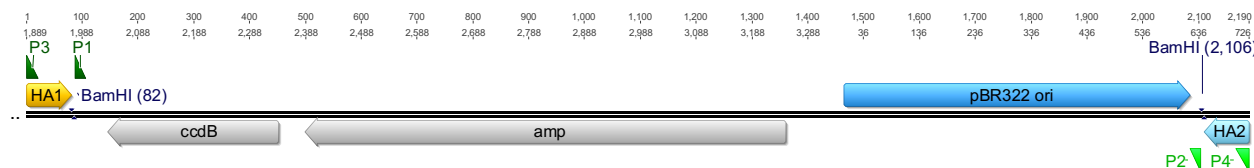


Figure A.8: Geneious map of PCR1 product. P1 and P2 are priming sites used for this amplification. HA1 and HA2 are homology arms to her BGC. The P3 and P4 sites are part of the homology arms, and they are the priming sequences used in PCR 2.

PRIMERS FOR PCR1

NAME	SEQUENCE (5'-3')	FEATURES
P_BR322ampccdB-her-1	CTTCTGGCGGACGGAGCGCCACCG GCCGCGATCGGCGCCTTGTGCGCGA GCCTGCCGCGCACCTCCTTCGCCGCC GCGGGGATCCGGTGTGGTAGCTCG CGTATT	Homology arm 1 (BGC, 3' end, reverse strand) <u>BamHI site</u> priming sequence 1 (pBR322, forward strand)
P_BR322ampccdB-her-2	TGAGAAATCGAATGACTGATGGCGG GGAGGGCGTAAGTGCCTGGTGAAGT CCTGGCGCGCAGACGCACGCGATGT TCTTCGGATCCACAAATGGCAAGGG CTAATG	Homology arm 2 (BGC, 5' end, forward strand) <u>BamHI site</u> priming sequence 2 (pBR322, reverse strand)

PCR 1 REACTION SETUP AND THERMOCYCLER PROGRAM

Reagent	μL per reaction ($V_T=50 \mu\text{L}$)	Temperature ($^{\circ}\text{C}$)	Time
Water	21	98	0'10"
pBR322-amp-ccdB-rpsL	1 (~1ng)	55	0'05"
P_BR322ampccdB-her-1 (10 μM)	1.5	72	0'11"
P_BR322ampccdB-her-2 (10 μM)	1.5		
PrimeSTAR Max Premix (2X)	25	12	∞

Appendix A (continued)

Step 2 – **PCR 2:** Amplification of PCR1 product with addition of pBAC2015 homology arms, which facilitate incorporation of the counter-selectable cassette and BGC homology arms into pBAC2015 via linear-linear homologous recombination.

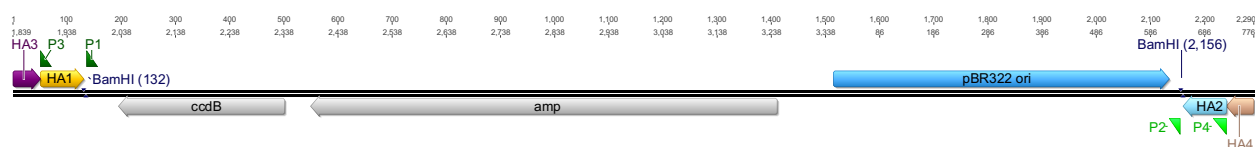


Figure A.9: Genious map of PCR2 product. P3 and P4 are priming sites used for this amplification. HA1 and HA2 are homology arms to her BGC. The P3 and P4 sites are part of the homology arms, and they are the priming sequences used in PCR 2.

PRIMERS FOR PCR2

NAME	SEQUENCE (5'-3')	FEATURES
P_BR322ampccdB-her-3	GGTAGGCCTGGCGGCCGCCTGGCCGTCG ACATTTAGGTGACACTATAGA ACTTCTG GCGGACGGAGCGCC	Homology arm 3 (pBAC2015, Sp6 end) priming sequence 3 (on Homology Arm 1)
P_BR322ampccdB-her-4	ATATTGCTCTAATAAATTTGCGGCCGCT AATACGACTCACTATAGGGAGAT GAGA AATCGAATGACTGATGGCG	Homology arm 4 (pBAC2015, T7 end) priming sequence 4 (on Homology Arm 2)

PCR 2 REACTION SETUP AND THERMOCYCLER PROGRAM

Reagent	μL per reaction (V _T =50 μL)	Temperature (°C)	Time
Water	21	98	0'10"
PCR1 product	1 (~1ng)	55	0'05"
P_BR322ampccdB-her-3 (10 μM)	1.5	72	0'12"
P_BR322ampccdB-her-4 (10 μM)	1.5	12	∞
PrimeSTAR Max Premix (2X)	25		

35x

Appendix A (continued)

Step 3 – **Linear-Linear Homologous Recombination:** Incorporation of the counter-selectable cassette with BGC homology arms into the cloning site of pBAC2015 via homologous recombination (Figure A.10). LLHR between the amplified cassette and vector backbone is facilitated by the pBAC2015 homology arms, HA3 and HA4, added during PCR 2. Here, LLHR occurs in a specially engineered strain *E. coli* GBdir-gyrA462 that can tolerate the counter selectable marker *ccdB*, a toxic protein (Wang *et al.*, 2016).

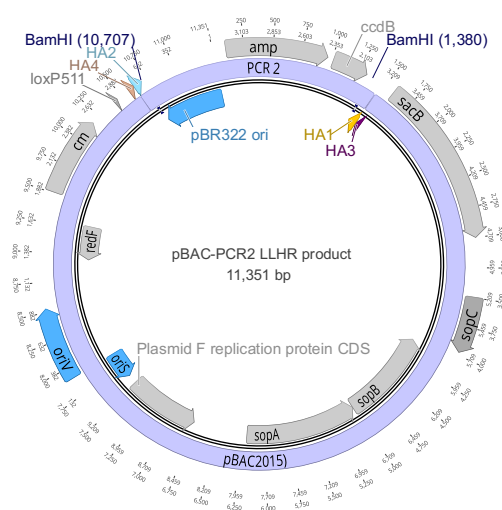


Figure A.10: Genious map of pBAC-PCR2 LLHR product.

Step 4 – **BamHI Digestion:** Release of the counter-selectable cassette from BGC, leaving behind the final linear cloning vector with BGC homology arms (Figure A.11).

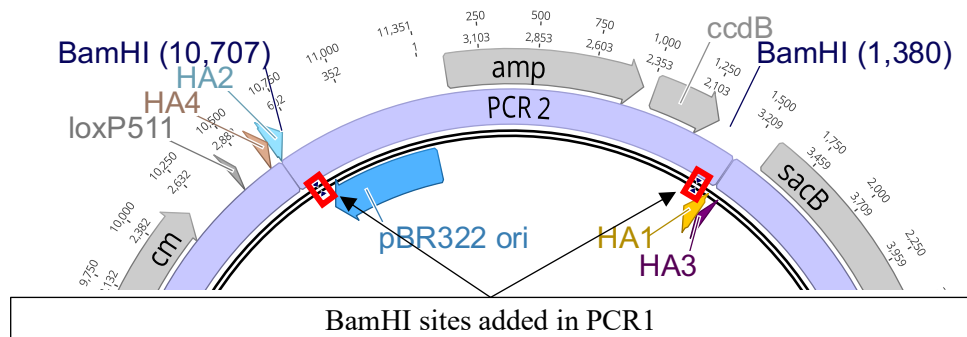


Figure A.11: Close-up view of counterselectable cassette in pBAC2015 backbone emphasizing the added BamHI sites (boxed in red) used to liberate the counterselectable cassette from the vector backbone.

Appendix A (continued)

Both PCR1 and PCR2 were carried out successfully using the protocols detailed above (Figure A.12A). The vector pBAC2015 was digested with EcoRI (linearized) prior to LLHR with the PCR2 product (Figure A.12B).

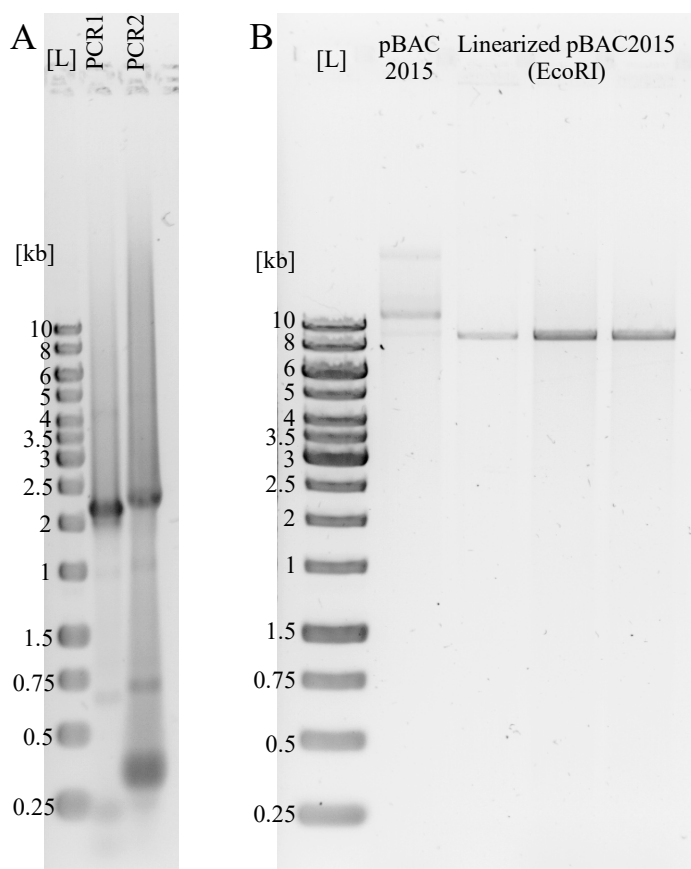


Figure A.12: Gel showing components needed for step 3, LLHR. (A) PCR1 and PCR2 products, side by side. PCR1 yields a 2202 bp amplicon while PCR2 yields a 2302 bp amplicon. **(B)** Linearized pBAC2015, which was digested with EcoRI. Three such preps were made, and compared to undigested pBAC2015.

The LLHR protocol was carried out twice, as described in Wang *et al.*, 2016. However, no positive clones were ever obtained. To troubleshoot this, the vector pBR322-amp-ccdB-rpsL was transformed into the strain which had been used for LLHR. This was done to check that the strain being used was actually *E. coli* GBdir-gyrA462, which can tolerate the counter selectable marker ccdB. Indeed, the plasmid pBR322-amp-ccdB-rpsL was successfully recovered from the strain. This indicates that the LLHR step failed for a different reason, which is at present unclear.

Appendix A (continued)

A.4 **References Cited**

- Miller-Wideman, M., Makkar, N., Tran, M., Isaac, B., Biest, N., & Stonard, R. (1992). Herboxidiene, a new herbicidal substance from *Streptomyces chromofuscus* A7847. Taxonomy, fermentation, isolation, physico-chemical and biological properties. *The Journal of Antibiotics*, 45(6), 914–921. doi: <http://dx.doi.org/10.7164/antibiotics.45.914>
- Shao, L., Zi, J., Zeng, J., & Zhan, J. (2012). Identification of the herboxidiene biosynthetic gene cluster in *Streptomyces chromofuscus* ATCC 49982. *Applied and Environmental Microbiology*, 78(6), 2034–2038. doi: 10.1128/AEM.06904-11
- Wang, H., Li, Z., Jia, R., Hou, Y., Yin, J., Bian, X., Li, A., Müller, R., Stewart, A. F., Fu, J., & Zhang, Y. (2016). RecET direct cloning and Red $\alpha\beta$ recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nature Protocols*, 11(7), 1175–1190. doi: 10.1038/nprot.2016.054
- Wang, H., Li, Z., Jia, R., Yin, J., Li, A., Xia, L., Yin, Y., Müller, R., Fu, J., Stewart, A. F., & Zhang, Y. (2018). ExoCET: exonuclease in vitro assembly combined with RecET recombination for highly efficient direct DNA cloning from complex genomes. *Nucleic Acids Research*, 46(5), e28–e28. doi: 10.1093/nar/gkx1249

Appendix B: Examining cryptic biosynthetic gene clusters in *Micromonospora* sp. B006

Micromonospora sp. B006 is a freshwater actinomycete isolated from sediment of Lake Michigan (Braesel *et al.*, 2018). Bioinformatics analysis of the *Micromonospora* sp. B006 genome revealed at least 17 biosynthetic gene clusters (BGCs), eleven of which are orphan (i.e. the natural product is unknown). Transcript analysis indicated that four of the orphan gene clusters are transcribed under our culture conditions (A1 medium), whereas the remaining seven BGCs appear silent. This appendix describes efforts to probe two of these transcriptionally active cryptic BGC's found in its genome: 1) Cluster 13, an enediynes PKS and 2) Cluster 15, a class I lantipeptide.

B.1 Does *Micromonospora* sp. B006 produce DNA damaging enediynes?

Enediynes are extremely cytotoxic natural products. (Shen *et al.*, 2015) In order to test for the presence of enediynes in the fermentation broth of *Micromonospora* sp. B006, we utilized the colorimetric lysogenic induction assay – also called the biochemical induction assay (BIA) described in Elespuru & Yarmolinsky, 1979.

In short, the assay is carried out on solid media plates, as in Figure B.1. An agar overlay containing *E. coli* BR513 (ATCC) is placed over the base agar layer. *E. coli* BR513 synthesizes the enzyme β -galactosidase (*lacZ*) as a consequence of DNA damage. Atop agar overlay 1, the test solution being analyzed is spotted. Then, a second agar overlay containing Xgal, the substrate of β -galactosidase, is poured over the top. If there are DNA damaging agents, such as enediynes, present in the test solution, then blue zones of inhibition will appear.

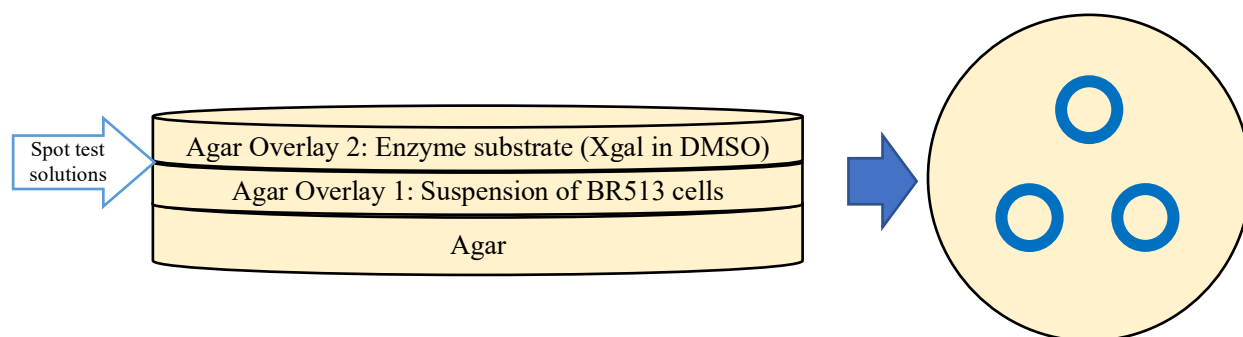


Figure B.1: Schematic of lysogenic induction assay (Elespuru & Yarmolinsky, 1979). The appearance of a blue zone of inhibition following the second overlay indicates presence of DNA damaging agents.

Appendix B (continued)

B.1.1 Assay 1.0

Two 50 mL cultures of *Micromonospora* sp. B006 were grown in A1 medium for 10 days – one culture at room temperature (RT) and one culture at 30°C. Duplicate assay plates will be prepared for each condition. Following incubation, a 20 mL aliquot of each culture was harvested via centrifugation. The supernatant was extracted three times with an equivalent volume of EtOAc. The cell pellet was resuspended in 20 mL MeOH and extracted under agitations for 30 min. Both extracts were dried down via rotary evaporation.

Four test solutions were prepared for the assay: 1) Positive control (+), Bleomycin 2 mg/mL in water; 2) Negative control (-), 1:1 A1F:50%MeOH; 3) Cell pellet extract (CPE); and 4) Supernatant extract (SE). To make the CPE and SE solutions, the dried extracts were resuspended in 1 mL MeOH. Each test solution was spotted (10µL) over agar overlay 1, and the test solutions were left to incubate for 3 hours at 37°C. Then, agar overlay 2 was poured over the top. After 30 minutes of development, only the bleomycin positive control showed a blue zone of inhibition, meaning no DNA damaging agents were detected in the extracts (Figure B.2).

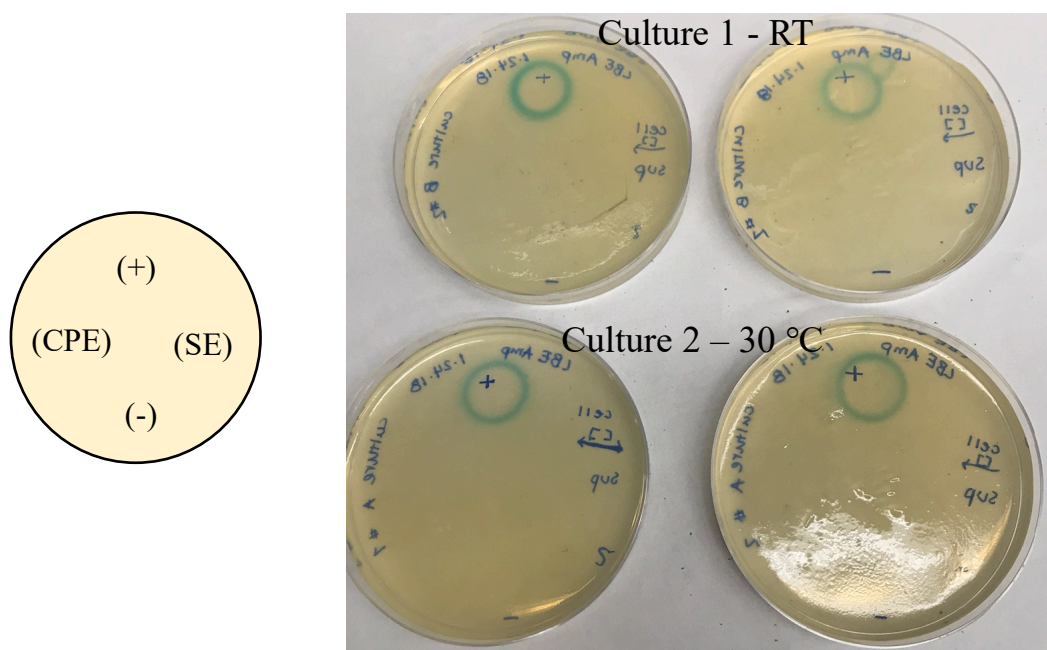


Figure B.2: Plate layout diagram and Assay 1.0 plates following development. No DNA damaging agents were detected in the B006 extracts.

Appendix B (continued)

B.1.2 Assay 2.0

A second attempt was made to detect DNA damaging enediynes. In addition to A1, two other media types were used that have been reported for production of enediynes in other strains – ISP4 (Davies *et al.*, 2005) and “Molasses” (Maiese *et al.*, 1989). Cultures were scaled up to 1L and grown at 30°C for 10 days. Still, no DNA damaging agents were detected in the extracts (Figure B.3).

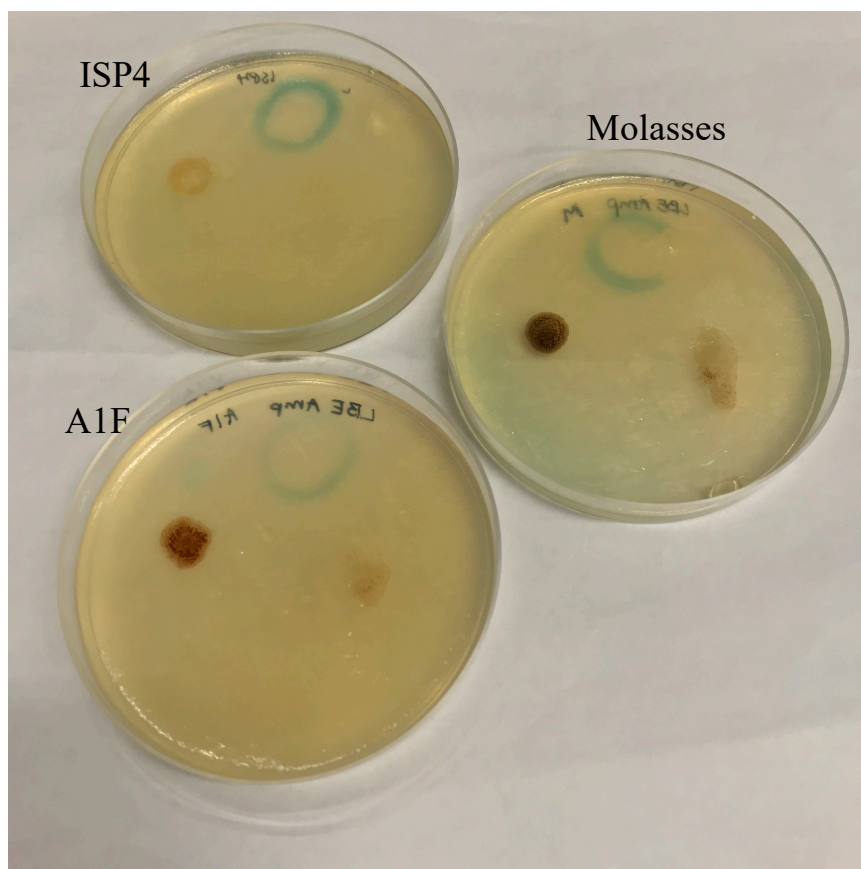


Figure B.3: Assay 2.0 plates following development. No DNA damaging agents were detected in the B006 extracts.

Appendix B (continued)

B.2. Cloning cryptic lantipeptide cluster for heterologous expression

B.2.1 Cloning plan for pTEK01

The cryptic lantipeptide BGC (cluster #15) will be cloned in the expression vector pUCP_neo and placed under control of the inducible P_bad/araC operon. The construct, pTEK01 (Figure B.4), will be assembled using NEBuilder, an *in vitro* assembly system from New England Biolabs which utilizes homology overhangs on PCR amplified fragments for recombination.

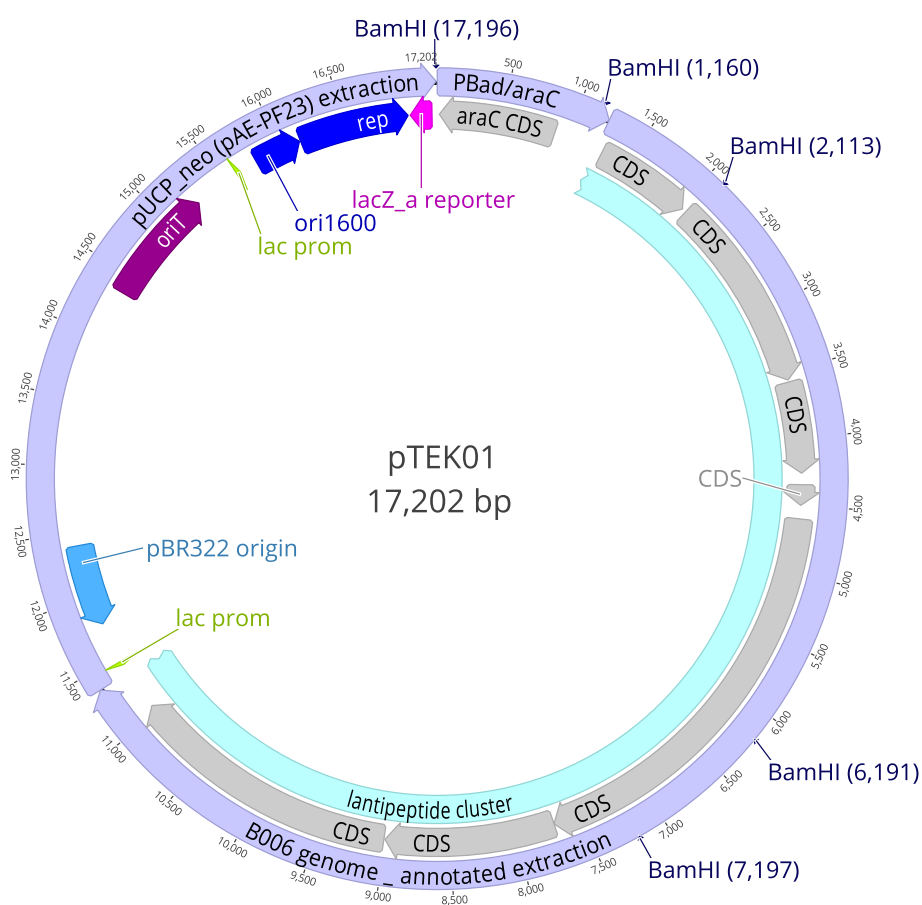


Figure B.4: Map of pTEK01 with annotated BamHI restriction sites (screening).

Appendix B (continued)

B.2.2 Isolation of B006 gDNA

Micromonospora sp. B006 gDNA was isolated using the GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich). Four gDNA preps were generated (Figure B.5).

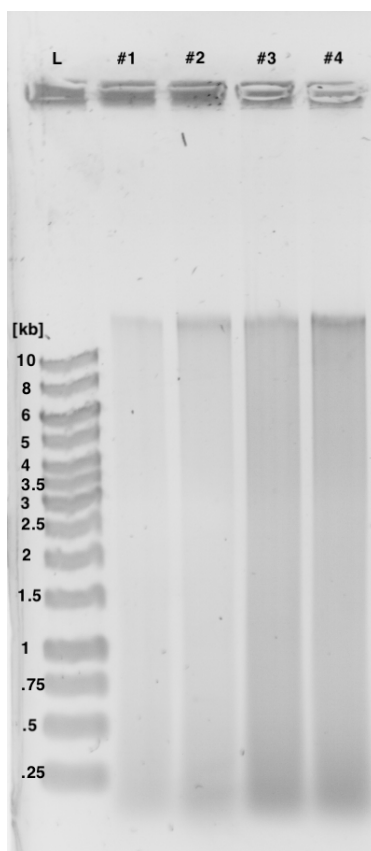


Figure B.5: Gel of *Micromonospora* sp. B006 gDNA.

NANODROP VALUES:

Sample	A260/280	A260/230	ng/μL
#1	1.98	1.84	9.3
#2	2.27	2.46	12.45
#3	1.98	2.04	33.5
#4	2.11	2.67	31.7

Appendix B (continued)**B.2.3 PCR Amplification of pTEK01 components with overhangs for NEBuilder**

PRIMERS FOR AMPLIFICATION OF PTEK01 FRAGMENTS WITH HOMOLOGY ARM
OVERHANGS FOR NEBUILDER

Primer (homology arm)	Seq (5'-3')	primer length (nt)	T_mQ5 (°C)	overlap length (nt)
P1_PBAD F (<i>neoR</i>)	<u>CTCTAGAGGATCCCC</u> ACCTGCATCGATTATTATGACAAC	25	63	15
P2_PBAD R (<i>lantipepF</i>)	<u>AGCTGGAGCTGTACGA</u> AATTCCCAAAAAACGGGTATGG	23	64	16
P3_LANTIPEP F (<i>PbadR</i>)	<u>TTTTTTGGGAATT</u> TCGTACAGCTCCAGCTTCT	19	66	14
P4_LANTIPEP R (<i>neoF</i>)	<u>GAATTCGAGCTC</u> CCGAGCGAGATTGACCAA	18	65	12
P5_NEO F (<i>lantipepR</i>)	<u>TTGGTCAATCTCGCTCGG</u> GAGCTCGAATTCGTAATCATGGT	23	64	18
P6_NEO R (<i>PbadF</i>)	<u>TAAATCGATGCAGGT</u> GGGGATCCTCTAGAGTCGA	19	65	65

AMPLIFICATION OF P_{BAD}/ARAC

Reagent	μL per reaction (V_r=50 μL)	Temperature (°C)	Time
5x Q5 reaction buffer	10	98	1'00"
5x Q5 GC enhancer	10	98	0'10"
10 mM dNTPs	1	64	0'30"
P1_PBAD F (<i>neoR</i>) (10 μM)	2.5	72	0'45"
P2_PBAD R (<i>lantipepF</i>) (10 μM)	2.5	72	2'00"
Q5 high-fidelity DNA polymerase	0.5	12	∞
Template – pIJ790	1 (~1ng)		
Water	22.5		

30x

Appendix B (continued)**AMPLIFICATION OF LANTYPEPTIDE BGC**

Reagent	μL per reaction ($V_{\text{T}}=50 \mu\text{L}$)	Temperature ($^{\circ}\text{C}$)	Time
5x Q5 reaction buffer	10	98	2'00"
5x Q5 GC enhancer	10	98	0'10"
10 mM dNTPs	1	66	0'30"
P3_LANTYPEP F (<i>PbadR</i>) (10 μM)	2.5	72	5'20"
P4_LANTYPEP R (<i>neoF</i>) (10 μM)	2.5	72	10'00"
Q5 high-fidelity DNA polymerase	0.5	12	∞
Template – B006 gDNA	2 (~50ng)		
Water	21.5		

AMPLIFICATION OF VECTOR BACKBONE – pUCP_NEO

Reagent	μL per reaction ($V_{\text{T}}=50 \mu\text{L}$)	Temperature ($^{\circ}\text{C}$)	Time
5x Q5 reaction buffer	10	98	1'00"
5x Q5 GC enhancer	10	98	0'10"
10 mM dNTPs	1	65	0'30"
Forward primer (10 μM)	2.5	72	3'00"
Reverse primer (10 μM)	2.5	72	2'00"
Q5 high-fidelity DNA polymerase	0.5	12	∞
Template – pUCP_neo	1 (~1ng)		
Water	22.5		

All PCR reactions were cleaned with the Zymo Research DNA Clean and Concentrator kit. The pUCP_neo amplicon was digested with DpnI to reduce template background. The reaction was cleaned again with the same kit.

Appendix B (continued)

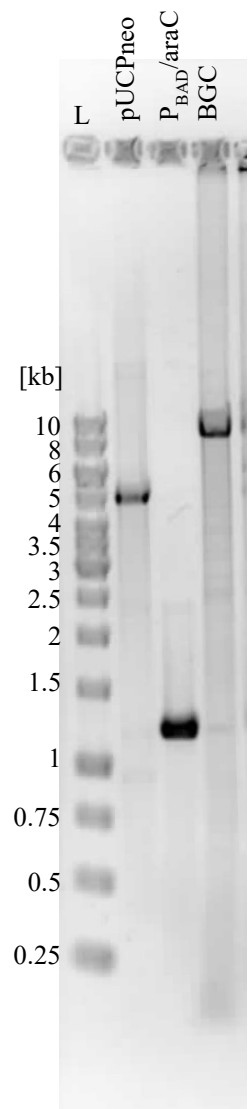


Figure B.6: Gel of purified PCR products prior to *in vitro* assembly via NEBuilder.

ANALYSIS OF NEBUILDER FRAGMENTS

Fragment	Length (bp)	ng/ μ L	A260/280	A260/230
pUCPneo (digested)	5,863	22.7	1.81	0.87
P _{BAD} /araC	1,255	75.1	1.86	2.15
BGC	10,174	33.0	1.84	0.75

Appendix B (continued)

B.2.4 NEBuilder for *in vitro* assembly of fragments

The NEBuilder HiFi DNA Assembly Cloning Kit was used for *in vitro* assembly of the fragments, following the protocol³ from NEB, which recommends 50-100ng of vector and 1:2 vector:insert for assembly of three fragments. The following reaction setup was calculated using the NEBioCalculator⁴:

CALCULATION OF FRAGMENT RATIOS

Fragment	Concentration (ng/μL)	length (bp)	mass (ng)	theoretical pmol	volume (μL)	actual pmol
pUCP_neo	22.7	5863	87.71	0.02421	3.86	0.02418
BGC	33.0	10174	304.4	0.04842	9.22	0.04840
PBad	75.1	1255	37.55	0.04842	0.50	0.04842

REACTION SETUP

Reagent	Volume (μL)
NEBuilder HiFi DNA Assembly Master Mix	13.6
pUCP_neo amplicon(DpnI digested)	3.86
Lantipeptide BGC amplicon	9.22
P _{BAD} /araC amplicon	0.50
V_T	27.18

Reaction incubated on dry block at 50°C for 1 hour. The reaction mixture was then transformed into *E. coli* 10-beta cells. The transformed cells were split 20:80 between two LB plates with kanamycin (50 μg/mL) selection and placed in incubator at 37°C. Growth was observed on both plates. Clones were grown in 5mL LB broth with kanamycin selection, and plasmid isolated with the ZR Plasmid Miniprep kit (Zymo Research).

³<https://www.neb.com/-/media/nebus/files/manuals/manuale2621-e5520.pdf>

⁴ <https://nebiocalculator.neb.com/#!/dsdnaamt>

Appendix B (continued)

B.2.5 Clone screening

Restriction digest screen of isolated plasmids carried out with BamHI.

Predicted fragment lengths (bp) for pTEK01: 9999, 4078, 1166, 1006, 953

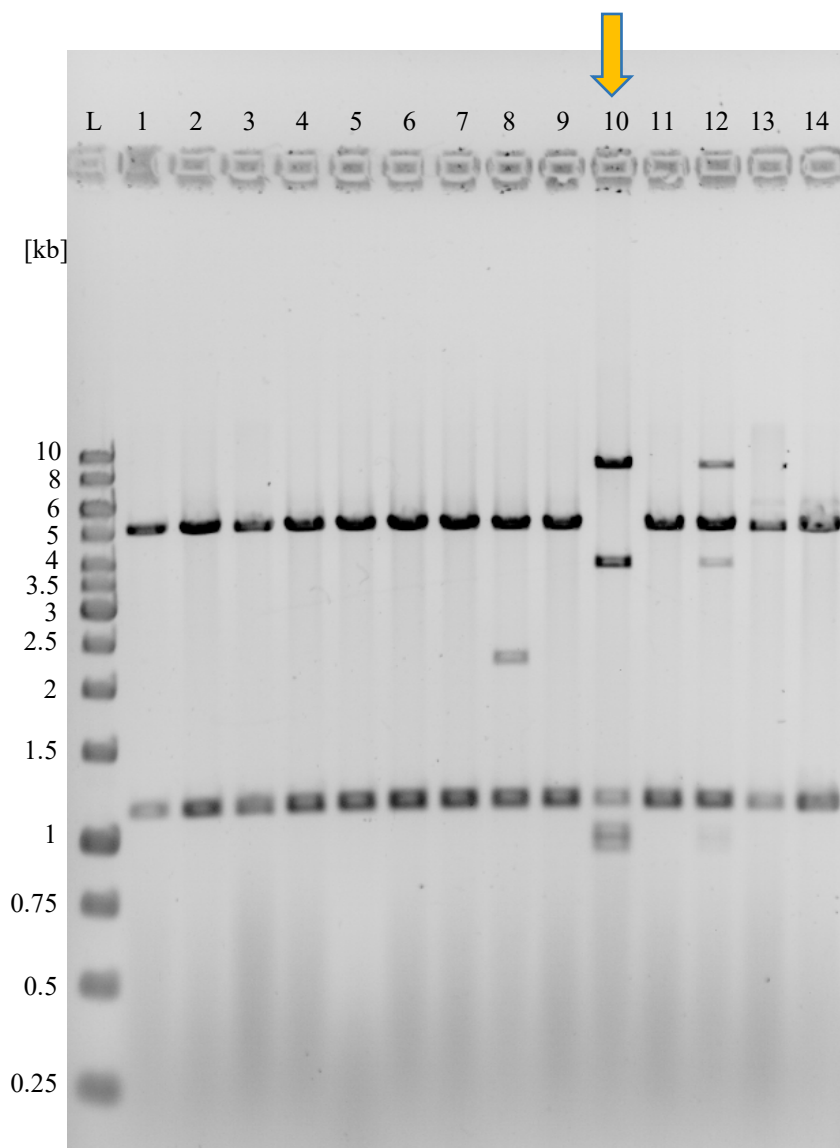


Figure B.7: Digest screen (BamHI) of *E. coli* 10-beta clones transformed with NEBuilder reaction.

One putative positive clone was identified in lane 10 of Figure B.7. The clone was cryopreserved, and the isolated plasmid subjected to bidirectional Sanger sequencing, which confirmed the fidelity of the clone.

Appendix B (continued)**PRIMERS USED FOR BIDIRECTIONAL SANGER SEQUENCING OF pTEK01**

Primer Name	Sequence (5' to 3')
P1_araC_f	TATAACCTTTCATTCCCAGCG
P2_araC_R	ACATCAGCGATCACCTGG
P4_Pbad_R	CCCAAAAAACGGGTATGG
PsTK1F1	TGCTCGTCGAGCTCTACTACGGCG
PsTK1F2	TGTTCTCGGTTCGGGTTCTGTG
PsTK1F3	TCCTGTTTCATCACCAGCCTGTT
PsTK1F4	TACTTCACCCTCGTCCTGCGT
PsTK1F5	AACGTCGCCTGACCGACGAA
PsTK1F6	ATCCGGGTGGAAGCACTGT
PsTK1F7	AGGTTCACCGTCGAGGTCACCT
PsTK1F8	CATGAGGTCATCCTGACGCT
PsTK1F9	CATCAACAGCGACGACCTCCTC
PsTK1F10	AGGTGTTACCAGGCTGGTGGA
PsTK1F11	AGTACACCGTCGAGTTCCCCGA
PsTK1F12	AGCACCGTGTCGATCGCAGAA
PsTK1F13	TGCACAAGGAACAGGACGTC
PsTK1R1	GCCACGACTGCTCGACCAA
PsTK1R2	GAAACCCGTGATGGCCGTGA
PsTK1R3	ACGCCGAGGAGGATGTGGAAC
PsTK1R4	TCCGCTGATGGGCAGGATCA
PsTK1R5	TAGGCGCTGCGTGTATCGGG
PsTK1R6	AGAGGCTCGTGGACGGAGAT
PsTK1R7	GTGCAGCGAGGTACATGCGT
PsTK1R8	GAACCGGAGGAACCACCAGT
PsTK1R9	GTCGTGGCTCTCGTACGGAA
PsTK1R10	TTGCGGTCTGGTGATCCAGTA
PsTK1R11	CTGTCCTGATGAAACAGCTC
PsTK1R12	CGAGTTGTTTCGAGCATCATT
PsTK1R13	ATGACGCCTACCTCGTACAG
PsTK1R14	GCTGGCCTTTTGCTCACATGTTC

Appendix B (continued)

B.3 References Cited

- Braesel, J., Crnkovic, C. M., Kunstman, K. J., Green, S. J., Maienschein-Cline, M., Orjala, J., Murphy, B. T., & Eustáquio, A. S. (2018). Complete Genome of *Micromonospora* sp. Strain B006 Reveals Biosynthetic Potential of a Lake Michigan Actinomycete. *Journal of Natural Products*, 81(9), 2057–2068. doi: 10.1021/acs.jnatprod.8b00394
- Davies, J., Wang, H., Taylor, T., Warabi, K., Huang, X. H., & Andersen, R. J. (2005). Uncialamycin, a new enediyne antibiotic. *Organic Letters*, 7(23), 5233–5236. doi: 10.1021/ol052081f
- Elespuru, R. K., & Yarmolinsky, M. B. (1979). A colorimetric assay of lysogenic induction designed for screening potential carcinogenic and carcinostatic agents. *Environmental Mutagenesis*, 1(1), 65–78. doi: 10.1002/em.2860010113
- Maiese, W. M., Lechevalier, M. P., Lechevalier, H. A., Korshalla, J., Kuck, N., Fantini, A., Wildey, M. J., Thomas, J., & Greenstein, M. (1989). Calicheamicins, a novel family of antitumor antibiotics: taxonomy, fermentation, and biological properties. *The Journal of Antibiotics*, 42(4), 558–563. doi: 10.7164/antibiotics.43.253
- Shen, B., Hindra, Yan, X., Huang, T., Ge, H., Yang, D., Teng, Q., Rudolf, J. D., & Lohman, J. R. (2015). Enediynes: Exploration of microbial genomics to discover new anticancer drug leads. *Bioorganic and Medicinal Chemistry Letters*, 25(1), 9–15. doi: 10.1016/j.bmcl.2014.11.019

Appendix C: Wiley Permission for Reproduction of Chapter 2

Rights DE <RIGHTS-and-LICENCES@wiley-vch.de>
 To: "tkornf2@uic.edu" <tkornf2@uic.edu>
 Cc: Rights DE <RIGHTS-and-LICENCES@wiley-vch.de>

Mon, Sep 21, 2020 at 4:45 AM

Dear Author,

We hereby grant permission for the requested use expected that due credit is given to the original source.

If material appears within our work with credit to another source, authorisation from that source must be obtained.

Credit must include the following components:

**- Journals: Author(s) Name(s): Title of the Article. Name of the Journal.
 Publication year. Volume. Page(s) Copyright Wiley-VCH GmbH/accepted not y.
 Reproduced with permission.**

If you also wish to publish your thesis in electronic format, you may use the article according to the Copyright transfer agreement:

3. Final Published Version.

Wiley-VCH hereby licenses back to the Contributor the following rights with respect to the final published version of the Contribution:

- a. [...]
- b. Re-use in other publications. The right to re-use the final Contribution or parts thereof for any publication authored or edited by the Contributor (excluding journal articles) where such re-used material constitutes less than half of the total material in such publication. In such case, any modifications should be accurately noted.

Kind regards

Heike Weller
Rights Manager
Rights & Licenses
Wiley-VCH Verlag GmbH
Boschstraße 12
69469 Weinheim
Germany
www.wiley-vch.de
T + (49) 6201 606-585
F + (49) 6201 606-332
rightsDE@wiley.com

Appendix C (continued)

Data Provided in Form

First name

Taylor

Last name

Kornfuehrer

Country

United States

State or Province

IL

Company name

University of Illinois at Chicago

I am, or I am working on behalf of one of the following:

the author of this work

I want to re-use:

Journal Content

Text permissions: number of words

All

Image permissions: name or number of figure

All

journal title

ChemBioChem

doi#

10.1002/cbic.202000349

journal Comments:

I would like to reproduce the full text of this article (and SI) as a chapter for my thesis. Thank you for your consideration.

Please provide information on the product(s) for which you are seeking permission.

Usage type

Thesis/dissertation

Appendix D: RSC Permission for Reproduction of Chapter 3

Dear Taylor

The Royal Society of Chemistry (RSC) hereby grants permission for the use of your paper(s) specified below in the printed and microfilm version of your thesis. You may also make available the PDF version of your paper(s) that the RSC sent to the corresponding author(s) of your paper(s) upon publication of the paper(s) in the following ways: in your thesis via any website that your university may have for the deposition of theses, via your university's Intranet or via your own personal website. We are however unable to grant you permission to include the PDF version of the paper(s) on its own in your institutional repository. The Royal Society of Chemistry is a signatory to the STM Guidelines on Permissions (available on request).

Please note that if the material specified below or any part of it appears with credit or acknowledgement to a third party then you must also secure permission from that third party before reproducing that material.

Please ensure that the thesis includes the correct acknowledgement (see <http://rsc.li/permissions> for details) and a link is included to the paper on the Royal Society of Chemistry's website.

Please also ensure that your co-authors are aware that you are including the paper in your thesis.

Regards
Gill Cockhead
Contracts & Copyright Executive

Gill Cockhead
Contracts & Copyright Executive
Royal Society of Chemistry,
Thomas Graham House,
Science Park, Milton Road,
Cambridge, CB4 0WF, UK
Tel +44 (0) 1223 432134

Appendix D (continued)

From: tkornf2@uic.edu
Sent: 20 July 2020 18:23
To: CONTRACTS-COPYRIGHT (shared) <Contracts-Copyright@rsc.org>
Subject: Permission Request Form: Taylor Kornfuehrer

Name: Taylor Kornfuehrer

Institution: University of Illinois at Chicago

Address:

900 S. Ashland
Chicago
60607
United States

I am preparing the following work for publication:

Article/chapter title: Stereochemistry in biology and synthetic biology
Journal/book title: Thesis
Editor/author(s): Taylor Kornfuehrer
Publisher: University of Illinois Chicago
Is this request for a thesis?: Yes

I would very much appreciate your permission to use the following material:

Journal/book title: MedChemComm
Editor/author(s): Taylor Kornfuehrer and Alessandra S. Eustáquio
ISBN/DOI: 10.1039/C9MD00141G
Year of publication: 2019
Page(s): 1256-1272
Type of material: Full paper - all text and figures
Figure/image number (if relevant):

Any additional comments:

All reproduced materials are to be used for my Thesis. Thank you for your consideration.

Agree to terms: I agree

VITA

TAYLOR KORNFUEHRER

EDUCATION

M.S., Pharmacognosy, The University of Illinois at Chicago	2020
B.S., Chemistry with High Honors, The University of Texas at Austin	2017

SCHOLARSHIPS AND HONORS

NIH T32 Training Fellowship, The University of Illinois at Chicago	2019-2020
Chemistry and Biochemistry Author's Scholarship, Department of Chemistry, The University of Texas at Austin	2016
College Scholar, College of Natural Sciences, The University of Texas at Austin	2016
Dean's List The University of Texas at Dallas	2012
Academic Excellence Scholarship The University of Texas at Dallas	2011-2013

TEACHING EXPERIENCE

Teaching Assistant – Department of Pharmacy Practice College of Pharmacy, University of Illinois at Chicago	2017-2018
Teaching Assistant – Organic Chemistry Laboratory Department of Chemistry, The University of Texas at Austin	2017
Peer Led Team Learning Leader – General Chemistry Student Success Center, The University of Texas at Dallas	2012-2013

PEER REVIEWED PUBLICATIONS

Kornfuehrer, T., Romanowski, S., de Crécy-Lagard, V., Hanson, A. D., & Eustáquio, A. S. An Enzyme Containing the Conserved Domain of Unknown Function DUF62 Acts as a Stereoselective (*R_S*,*S_C*)-*S*-Adenosylmethionine Hydrolase. *ChemBioChem* (2020, in press)

Kornfuehrer, T., & Eustáquio, A. S. Diversification of polyketide structures via synthase engineering. *MedChemComm*, 10, 1256–1272 (2019)

SELECTED POSTER PRESENTATIONS

Kornfuehrer, T. and Eustáquio, A.S. “Defining the role of an enzyme of unknown function.” American Society of Pharmacognosy 60th Annual Meeting. Madison, WI. **July 2019**

Kornfuehrer, T. and Eustáquio, A.S. “Defining the role of an enzyme of unknown function.” UIC College of Pharmacy Research Day. Chicago, IL. **February 2019**

ORAL PRESENTATIONS

Kornfuehrer, T., “Synthetic biology of polyketide natural products from *Streptomyces*”, Specialized Metabolite Community, Chicago, IL. **October 2019**