ABSTRACT

KALKREUTER, ROBERT EDWARD. Improving the Scope and Utility of Precursor-Directed Biosynthesis via Synthetic Biology. (Under the direction of Dr. Gavin J Williams).

Nature has utilized the power of evolution to create intricate biosynthetic pathways for the synthesis of various classes of natural products. Terpenes, the largest group of natural products, are primarily derived from only two substrates. Another large class, the polyketides, are known for their structural diversity and their many clinical uses. Type I polyketides, like the antibiotic Erythromycin A, are constructed via multi-megadalton polyketide synthase (PKS) 'assembly lines'. The modular nature of these pathways has led to decades of interest in harnessing these enzymes for regioselective alterations of privileged scaffolds not easily accessible via synthetic or semisynthetic chemistry.

Of particular interest for enzyme engineers, the acyltransferase (AT) domain in PKSs is responsible for the selection of the malonyl-CoA-derived extender unit substrate for incorporation by a given module. Despite rarely incorporating substrates beyond malonyl-CoA or methylmalonyl-CoA naturally, ATs arguably can provide the most added diversity to the polyketide natural product family through the selection of rare or non-natural malonyl-CoA derivatives. By shifting the selectivity of the AT, polyketide products can be diversified at every other carbon, providing chemical handles for new interactions and further semi-synthetic derivatization.

To obtain a better understanding of AT selectivity, movements, and mutations, molecular dynamics simulations were performed (Chapter 2). Insight from the simulations led to identification of novel residues that contribute to AT selectivity. Successful mutations were used

to inform swaps of AT structural motifs, leading to substrate selectivity shifts of greater than 300-fold while accounting for replacement of less than 3% of the AT's sequence. During these *in vitro* studies, the ketoreductase (KR) domain was identified as a bottleneck for non-natural extender unit incorporation for the first time (Chapter 3). This lack of promiscuity by the AT, KR and other domains were explored more fully using the final two modules of the pikromycin PKS, PikAIII and PikAIV. The native and engineered promiscuities of both modules were compared, and AT mutagenesis enabled the first incorporation of two non-natural extenders into a full-length polyketide product (Chapter 4).

To provide a robust source of natural and non-natural polyketide and terpene extender units, a directed evolution tool capable of screening or selecting for the best strain or enzyme is needed. Towards this goal, a transcription factor-based biosensor was refactored, optimized, and characterized with malonyl-CoA and the non-native methylmalonyl-CoA to provide extenders for new polyketide chassis (Chapter 5). To apply this technology towards the directed evolution of terpenes, a repressor protein recognizing linear monoterpenes was engineered and used as a test case for the development of more effective genetically-encoded biosensors (Chapter 6). Together, the described combined approaches of *in silico*, *in vitro*, and *in vivo* evolution of natural product biosynthetic enzymes provide a foundation of new knowledge and strategies for future natural product pathway engineering.

Improving the Scope and Utility of Precursor-Directed Biosynthesis via Synthetic Biology.

by
Robert Edward Kalkreuter

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Chemistry

Raleigh, North Carolina
2018

APPROVED BY:

_____	_____
Dr. Gavin J Williams Committee Chair	Dr. Reza Ghiladi
_____	_____
Dr. Jonathan S Lindsey	Dr. Wei-chen Chang

DEDICATION

I would like to dedicate this work to my parents, Bob Kalkreuter and Patrea Pabst, for all of their considerable support throughout my life. I would not have grown into the person or scientist I am today without both of you.

BIOGRAPHY

Robert Edward Kalkreuter was born in Decatur, Georgia to Bob Kalkreuter and Patrea Pabst, and he has a younger brother, Charlie. Edward attended elementary and middle school in Hart County, Georgia while also exploring the outdoors on the family's rural Georgia farm. Edward then attended and graduated from Rabun Gap-Nacoochee School. While there, he decided to attempt a Ph.D. in chemistry due in large part to support from his parents and educators. After high school, Edward attended Emory University in Atlanta, Georgia, and he graduated in 2013 with a Bachelor of Science in Chemistry. During his time there, Edward studied the effect of nitric oxide in prostate cancer cell lines under the supervision of Dr. John Petros and Dr. Rebecca Arnold. After finishing his undergraduate studies, Edward started his Ph.D. work on enzyme engineering in the lab of Dr. Gavin Williams at North Carolina State University in Raleigh, NC. Following his Ph.D., Edward will move to the Scripps Research Institute in Jupiter, Florida to begin his postdoctoral studies under the direction of Dr. Ben Shen.

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Dr. Gavin Williams, for his continual support of my research, my crazy ideas, and my growth as a scientist. He has challenged me to defend my ideas and my science, and he has pushed me to leave my academic comfort zone. None of this work would be possible without his mentorship.

I would also like to thank all members of the Williams Lab, past and present. Everyone in the lab has had a positive effect on my research and success through helpful discussions, teaching, training, support, and collaboration. I would like to specifically thank Dr. John McArthur for his help in getting me started in the world of natural products and directed evolution. I would also like to thank Kyle Bingham and Aaron Keeler for their incredibly high levels of dedication in the lab, and much of the work described herein is due to their hard work and attention to detail.

I would like to acknowledge Dr. Yaroslava Yingling and her student, Hoshin Kim, for their help in teaching me the basics of molecular dynamics simulations.

Additionally, I would like to thank my undergraduate research advisors, Dr. John Petros and Dr. Rebecca Arnold, for all of their support in preparing me for a career in research.

Finally, I would like to thank my committee members, Dr. Reza Ghiladi, Dr. Wei-chen Chang, and Dr. Jonathan Lindsey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INDEX OF ACRONYMS AND ABBREVIATIONS

10-dML 10-deoxymethynolide

6-dEB ..... 6-deoxyerythronolide B

6-PGDL 6-phospho-D-glucono-1,5-lactone

A ..... adenylation domain

ACP ..... acyl carrier protein

AGOS ..... artificial gene operon assembly system

AT ..... acyltransferase

ATP ..... adenosine triphosphate

BAC ..... bacterial artificial chromosome

BGC ..... biosynthetic gene cluster

C ..... condensation domain

Cin ..... cinnabaramide

CM ..... C-methyltransferase

CoA ..... coenzyme A

CRISPR ..... clustered regularly interspaced short palindromic repeats

DBD ..... DNA-binding domain

DEBS ..... 6-deoxyerythronolide B synthase

DH ..... dehydratase

DMAA ..... dimethylallyl alcohol

DMAPP ..... dimethylallyl pyrophosphate

EDTA ..... ethylenediaminetetraacetic acid

EGFP ..... enhanced green fluorescent protein

emCoA ..... ethylmalonyl-CoA

EMSA ..... electrophoretic mobility shift assay

ER ..... enoylreductase

Ery ..... erythromycin

FAC ..... fungal artificial chromosome

FACS ..... fluorescence-activated cell sorting

FG-MD ..... fragment-guided molecular dynamics

FPLC ..... fast protein liquid chromatography

FPP ..... farnesyl pyrophosphate

G6P glucose-6-phosphate

G6PDH ..... glucose-6-phosphate dehydrogenase

GES ..... geraniol synthase

GFP ..... green fluorescent protein

GGPP ..... geranylgeranyl pyrophosphate

GPP ..... geranyl pyrophosphate

GT ..... glycosyltransferase

Hct ..... hectochlorin

HPLC ..... high-performance liquid chromatography

HR-LCMS ..... high-resolution liquid chromatography mass spectrometry

IPP ..... isopentenyl pyrophosphate

IPTG ..... isopropyl-β-D-thiogalactopyranoside

KR ..... ketoreductase

KS ..... ketosynthase

LBD.....ligand-binding domain

mCoA.....malonyl-CoA

MD.....molecular dynamics

mmCoA.....methylmalonyl-CoA

mmSNAC.....methylmalonyl-SNAC

MT.....methyltransferase

MWCO.....molecular weight cut-off

$NADP^+$.....nicotinamide adenine dinucleotide phosphate, oxidized

NADPH.....nicotinamide adenine dinucleotide phosphate, reduced

NRPS.....non-ribosomal peptide synthetase

O-MT.....O-methyltransferase

PCP.....peptidyl carrier protein

PDB.....protein data bank

pgmCoA.....propargylmalonyl-CoA

pgmSNAC.....propargylmalonyl-SNAC

Pik.....pikromycin

PKS.....polyketide synthase

Ppant.....phosphopantetheine

RBS.....ribosome binding site

RFP.....red fluorescent protein

RMSD.....root-mean-square deviation

RMSF.....root-mean-square fluctuation

RTH.....'round-the-horn mutagenesis

SDM ..... site-directed mutagenesis

SDS-PAGE ..... sodium dodecyl sulfate-polyacrylamide gel electrophoresis

sfGFP ..... super-folder green fluorescent protein

SNAC ..... S-*N*-acetylcysteamine

T ..... thiolation domain

TCEP ..... tris(2-carboxyethyl)phosphine

TE ..... thioesterase

TFA ..... trifluoroacetic acid

Tha ..... thailandin

TIR ..... transcription initiation rate

TP ..... thiophenol

WT ..... wild-type

CHAPTER 1

Introduction to Precursor-Directed Biosynthesis of Natural Products

Adapted with permission from: Kalkreuter, E.; Carpenter, S.M.; Williams, G.J., Precursor-Directed Biosynthesis and Semi-Synthesis of Natural Products. In *Chemical and Biological Synthesis: Enabling Approaches for Understanding Biology*, Westwood, N.J.; Nelson, A., Eds. The Royal Society of Chemistry: London, UK, 2018; p 270-307.

Adapted with permission from: Kalkreuter, E.; Williams, G.J., Engineering Enzymatic Assembly Lines for the Production of New Antimicrobials. *Curr. Opin. Microbiol.* **2018**, *45*, 140-148.

1.1. Introduction

Natural products are biosynthesized by a wide variety of enzymes and pathways, each with their own unique assembly strategies. Two of the largest classes, polyketides and terpenes, are synthesized by polyketide synthases (PKSs) and terpene synthases, respectively. Together, polyketides and terpenes are wide-spread in Nature (averaging more than one gene cluster per bacterial genome) and demonstrate potent antimicrobial activities highlighted by such clinically-relevant compounds as erythromycin A, avermectin, taxol, and artemisinin.[1] Because of the significant value of these antimicrobials (> $20 billion annually for polyketides alone) and complex molecular structures, there is intense interest in engineering their producing biosynthetic pathways to access new natural products and analogues that fine-tune the biological activity and pharmacological profile.[2] In this Chapter, the biosynthesis of polyketides and terpenes will be summarized along with a review of combined chemical and biosynthetic approaches employed to diversify their structures.

1.2. Strategies for Coupling of Biosynthesis and Chemical Synthesis

Biological and chemical synthesis can be coupled in a number of ways that vary according to the number of total steps involved, the order of biological *vs*. chemical transformations, and the degree of biological manipulations or chemical synthesis required. In general, such approaches usually

capitalize on the stepwise and often modular biosynthetic assembly of natural products (**Figure 1.1A**). Most commonly, these strategies can be divided into the four approaches described below and in **Figures 1.1B-D**.

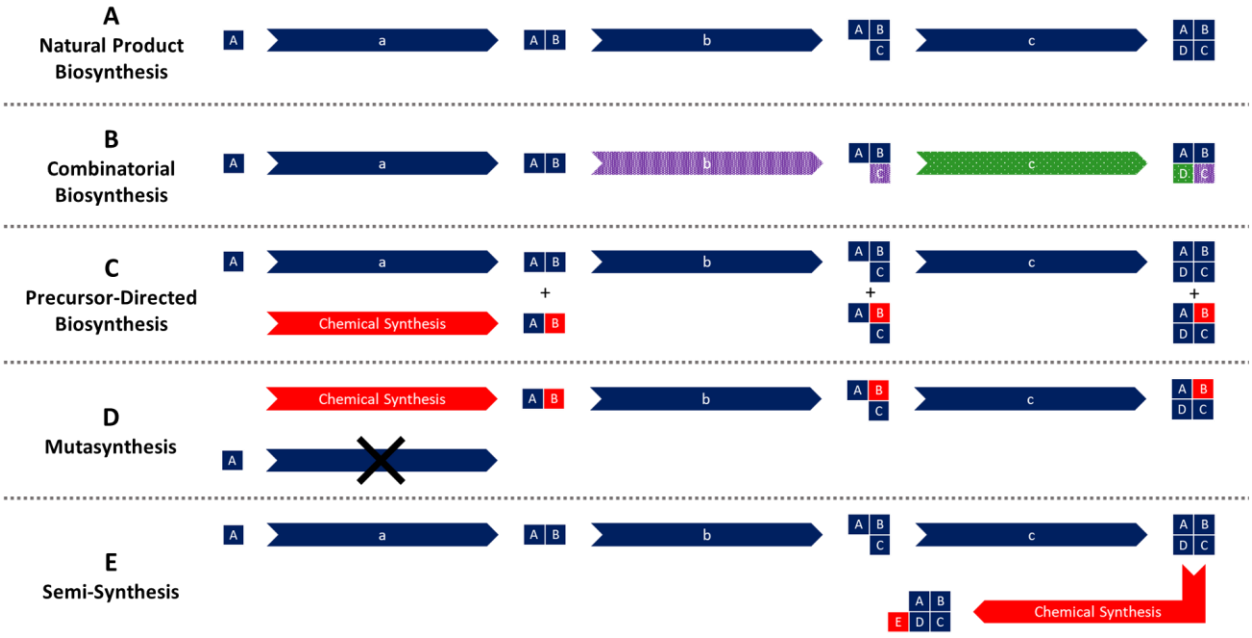


Figure 1.1. Strategies for Coupling of Biosynthesis and Chemical Synthesis. A) In the natural pathway, the native machinery synthesizes various intermediates and final products from native precursors (individual blocks). Panels **B-E** describe different chemical or biological approaches to introduce synthetic elements or additional non-native substrates (shaded blocks).

1.2.1. *Combinatorial Biosynthesis*

Broadly, combinatorial biosynthesis is the genetic manipulation of biosynthetic pathways to produce "unnatural" natural products. Formally, this approach does not involve synthetic chemistry, but is the platform for strategies that do. Combinatorial biosynthesis ranges from making relatively small changes to a biosynthetic pathway to provide a modified natural product structure, to making more comprehensive changes, such as mixing and matching genes from different pathways or organisms to make libraries of chimeric structures (**Figure 1.1B**). Although the focus of this chapter is to highlight the contribution of chemical synthesis, we will include examples of combinatorial biosynthesis by way of comparison.

1.2.2. *Precursor-Directed Biosynthesis*

Precursor-directed biosynthesis (**Figure 1.1C**) leverages organic synthesis to provide native or unnatural substrates that are fed to microbial strains and converted to a natural product or analogue by a biosynthetic pathway. This approach relies on the natural or engineered promiscuity of enzymes that utilize the artificial precursor and is especially prevalent in biosynthetic pathways that use easily diversifiable substrates like amino acids or carboxylic acids. Although this is a powerful approach to introducing diversity into the structures of complex natural products, precursor-directed biosynthesis often produces a mixture of natural and unnatural final products.[3]

1.2.3. *Mutasynthesis*

While highly similar to precursor-directed biosynthesis, mutasynthesis (**Figure 1.1D**) uses a microbial host in which a precursor biosynthetic pathway is knocked out. This process is useful if the synthetic precursor analogue is a poor substrate for the intercepted enzyme compared to its natural substrate or when a single product is desired.[4-8] A significant shortcoming of mutasynthesis is that often a precursor cannot be knocked out if it is required for primary metabolism in the cell. Chemobiosynthesis, a version of mutasynthesis used mostly for assembly line-like pathways, skips early-stage steps of a pathway through the feeding in of a synthetic precursor.[9]

1.2.4. *Semi-Synthesis*

Semi-synthesis (**Figure 1.1E**) accesses small molecules using traditional organic chemistry using starting materials isolated from Nature. These synthetic modifications may be useful to improve efficacy, bioavailability, stability, or other properties. In many cases, production of a natural product by semi-synthesis is much more practical than a total synthesis.[10]

1.3. Engineering the Biosynthesis of Polyketides

Nature has organized some of the most complex enzyme systems into pathways akin to assembly lines. These biosynthetic pathways are organized in a modular fashion, where each module acts a 'station' on the assembly line to add and tailor a portion of the final product. Possibly the most notable example of these assembly lines is the type I polyketide synthase (PKS).[11-12] These enzymes produce secondary metabolites that are diverse in structure and biological activities. A significant portion of these products are pharmaceutically relevant, as seen with many classes of natural products, and are considered to be privileged scaffolds.[13-14] As with synthetic drugs, there is a great interest in diversifying or making analogues of these natural products in the hopes of producing molecules with improved activity, decreased toxicity, etc. A variety of approaches including semi-synthesis, total synthesis, and biosynthesis, have been applied to these products with varying levels of success.[15-16] As more is discovered about the structure and enzymology of these assembly lines, targeted and rational engineering has become more feasible. Many of the first engineering successes were the results of precursor-directed biosynthesis and combinatorial biosynthesis, which have now been complemented with mutasynthesis and other hybrid strategies.

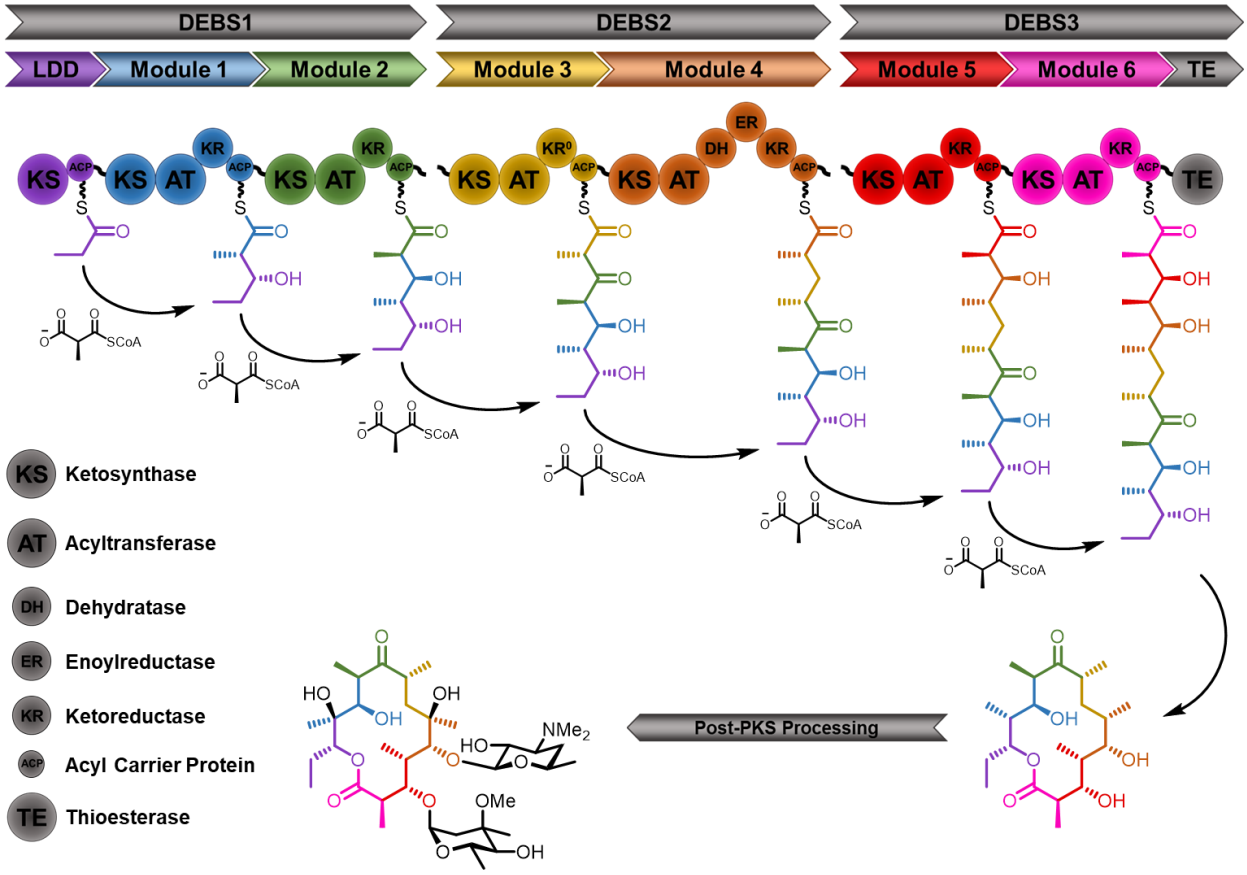


Figure 1.2. Organization of the Archetypal Assembly Line Responsible for Biosynthesis of the Erythromycin Core. The erythromycin PKS is made up of six extension modules, each responsible for incorporating a propionate group from methylmalonyl-CoA into the macrolactone core. A series of post-PKS tailoring steps yields the final natural product.

Type I PKSs are modular enzymatic systems that stitch together 'ketide' units—most commonly derived from malonyl- and methylmalonyl-CoA—in a stepwise fashion to create complex and stereo-rich partially reduced polyketides. The canonical and arguably most well-studied example of a type I PKS is the 6-deoxyerythronolide B synthase (DEBS), which synthesizes the macrolactone core of the antibiotic erythromycin A (**Figure 1.2**).[17] Other notable examples of polyketides produced by type I PKSs include rapamycin (immunosuppressant), epothilone A (anti-tumor), and lovastatin (anti-cholesterol, **Figure 1.3A**). This small pool of examples reveals the remarkable diversity in structure and biological activity of this natural product class. Indeed, these

privileged structures have an exceptionally high hit rate when compared to other natural product classes. For example, a 2005 study calculated a 0.3% hit rate from naturally occurring polyketides, compared to <0.001% of a standard pharmaceutical screen.[18]

Despite the diversity of these natural products, they are biosynthesized from a modest pool of simple small molecule building blocks (**Figure 1.3B**).[19] The intricacy of the natural product is derived from their templated biosynthesis; the proteins are organized as modules which are further divided into domains with distinct enzymatic activities. Each module is responsible for the installation of a building block and consists minimally of an acyltransferase (AT), ketosynthase (KS), and an acyl carrier protein (ACP). The AT initiates the incorporation of the thioester-activated building block at each module by acylating its active site serine residue. In most cases, the AT is specific for a single substrate and discriminates against other similar substrates available in the cell. The malonyl-derived building block is then transferred to the phosphopantetheine (Ppant) arm of the ACP. While tethered to the ACP, the KS catalyzes a Claisen-like condensation with the precursor from the upstream module to yield the newly extended product. Modules can also include up to three reductive domains: a ketoreductase (KR), enoylreductase (KR), and dehydratase (DH). These reduce the β-carbonyl to an alcohol, alkene, or fully saturated methylene, respectively. Further complexity can be introduced by modules that contains in-line tailoring domains, such as *C*- or *O*-methyltransferases (MTs). Finally, at the terminal module, the product is either hydrolyzed or cyclized by the thioesterase (TE) to yield the final polyketide product. The polyketide product is then usually further decorated by post-PKS tailoring enzymes, such as glycosyltransferases (GTs), P450s, or MTs.

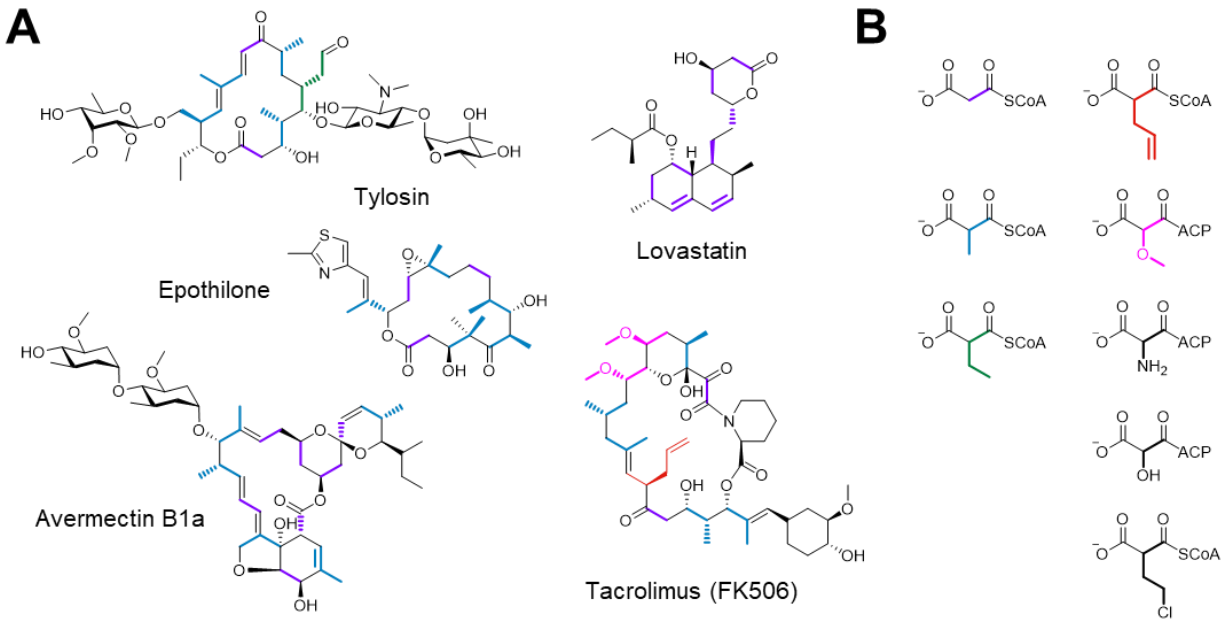


Figure 1.3. Diversity of Natural Polyketides. A) Examples of polyketides produced by Type I PKSs. **B)** Examples of building blocks utilized by type I PKSs. The three most common are malonyl-, methylmalonyl-, and ethylmalonyl-CoA.

Shortly after PKS pathways were first discovered, the concept of producing 'hybrid' antibiotics by genetic engineering was quickly translated into practice by leveraging actinorhodin biosynthesis to produce analogues, though little was known about the mechanism and specificity of these pathways.[20] A few years later, avermectin analogues were biosynthesized using precursor-directed biosynthesis approaches to yield compounds with anthelmintic and insecticidal activity.[21] These successes, among others, served as inspiration for the future of polyketide synthase engineering for the production of analogues.

1.3.1. *Challenges Associated with Engineering PKSs*

While there have been numerous successes towards reprogramming PKSs, there are still many challenges associated with engineering modular PKSs. The difficulty in accumulating structural information is at the heart of much of the fundamental research being done now; however, most

of the structure-guided engineering as yet is grounded in crystal structures of individual domains or homology models based on a small number of existing structures. For PKSs, it was not until 2014 that the Skiniotis and Sherman groups published a cryo-electron microscopy study of an intact module.[22-23] This was an important addition to the literature, but imaging of the different conformations only further demonstrated the importance of intra- and intermodular protein-protein interactions. More recently, a new method for room-temperature delivery of a crystal was used to look at an AT domain to get a better idea of its native conformational state.[24]

Our gaps in the structural and conformational knowledge of PKSs have closed significantly in the last 2-3 years, but there are still many unknowns associated with engineering PKSs and their structures. Much of the early engineering focused on domain or module swapping to give novel analogues. Unfortunately, most of these constructs were inactive.[25-26] Without appropriate, high-throughput screens, it is very difficult to sort through large libraries of pathway variants. This limits the application of "irrational" engineering approaches based on directed evolution. Additionally, the dogma has been that the AT domain of PKSs serves as the primary "gatekeeper" for selecting extender units for their respective modules; however, this view is shifting as more evidence suggests other bottleneck domains impact which extenders are incorporated into the final product.[27-29]

1.3.2. *A Model for Biosynthetic Engineering: 6-deoxyerythronolide B Synthase (DEBS)*

The discovery and decoding of the modular organization and domain functions of type I PKSs[17, 30-32] coupled with the advent of modern molecular biology techniques allowed for construction and manipulation of hybrid PKS pathways limited only by available DNA segments, quickly

accelerating engineering efforts to produce non-natural polyketide analogues. Many of the most notable early successes were achieved with DEBS.

Included in these engineering efforts were module deletions or swaps, in which an entire module or modules were either deleted to produce a smaller product or replaced with a module from another pathway to introduce an alternative building block. Cortes and co-workers discovered that the terminal thioesterase domain at the terminus of DEBS3 could be appended to the terminus of DEBS1 (which includes the loading module and modules 1 and 2, **Figure 1.1**) to create a truncated system capable of producing triketide lactones, providing evidence that the thioesterase was capable of cyclizing alternative substrates and allowing for a smaller model system to explore combinatorial biosynthesis.[33] The creation of this minimal model system combined with the newfound ability to express these enzymes in a heterologous host[34] and to use them *in vitro*[35-36] greatly accelerated combinatorial biosynthesis efforts. Khosla and co-workers determined that the TE was also active when fused to the C-terminus of module 5 to biosynthesize the predicted 12-membered macrolactone[37] and to module 3 to produce tetraketides,[38] further supporting the modularity of these systems.

The DEBS1TE system was modified by Leadlay and co-workers to create a hybrid system, replacing the natural propionate-accepting loading module with the known promiscuous loading module from the avermectin PKS, producing a panel of triketide lactones. This activity successfully transferred when the promiscuous loading module was inserted into the complete DEBS pathway in the producing strain, to synthesize a panel of erythromycin analogues.[39] This

advance provided additional evidence that unnatural precursors could be carried and processed through downstream enzymatic transformations when introduced at early biosynthetic steps.

Rowe et al. constructed a hybrid DEBS pathway capable of synthesizing expanded macrolactones by inserting modules 2 and 5 from the rapamycin synthase (Rap). The hybrid PKSs incorporated a malonyl unit at the location predicted by the unnatural module insertion.[40] However, as was seen with many hybrid systems, these suffered substantially from decreased yields of the desired product, and in some cases the major product was the result of the inserted module being skipped. The failings of these hybrid systems are multifaceted and still more complex than we fully understand, though considerable progress has been made to improve hybrid PKSs like these. For example, by assembling a panel of DEBS1TE/Rap hybrids and determining the productivity of the construct, it was determined that maintaining the natural ACP-KS domain partners between successive modules resulted in productive constructs, as it allows for the native protein-protein interactions to be maintained.[41] A panel of 154 bimodular constructs containing modules from eight different PKSs reveal flexibility in the linker region between the loading module and module 1 and that DEBS module 6, which has no natural downstream KS partner, was a universal donor in tested constructs.[42] In some cases, previously unproductive hybrid didomains could be restored by transplanting the natural ketosynthase domain for the donor module into the acceptor module, further supporting the importance of ACP/KS protein-protein interactions.[43]

In addition to success with module swapping, specific domains have been excised and replaced within a module to create a hybrid module. The truncated DEBS1TE system was used in a domain-swapping experiment in which the AT from module 1 was replaced with rapamycin AT2 to

produce a triketide lactone.[44] This concept was later expanded to synthesize full desmethyl-erythromycin analogues, demonstrating that the change in structure could be processed by downstream modules and post-PKS enzymes, albeit at low efficiencies.[45] In a similar fashion, AT4 from DEBS (the AT from module 4) was replaced with a predicted ethylmalonyl-CoA specific AT from the niddamycin PKS to successfully produce an ethyl-erythromycin derivative.[46] Such approaches have been applied to several other type I PKSs.[47]

Domain swapping has also been applied to the reductive domains, including substitution of an active KR with an inactive KR to yield the corresponding keto-analogue[48] and a swap of an active KR with an active DH-KR cassette to produce an α,β-unsaturated ketone.[49] Ketoreductase inactivation was also achieved through a targeted deletion of a portion of KR in module 5 of DEBS to produce a 6-dEB analogue.[50] High homology among reductive loops domains combined with downstream module tolerance to changes in oxidation have led to many successes in ketoreductase and dehydratase domain engineering.

The construction of hybrid PKS often yields unproductive or inefficient enzymes. In a landmark demonstration, multiple genetic modifications of the erythromycin PKS DEBS afforded a library of 61 6dEB analogues through the combinatorial replacement and insertion of various AT, KR, DH, and ER domains.[51] Yet, the yield of most of the compounds were <1% of the natural product, 6dEB, and could not be accurately determined. In recent years improvements have been made in predicting the appropriate boundaries for domain replacement and improving faulty protein-protein interactions. *In vitro* analysis of wild-type and hybrid modules revealed that the introduction of heterologous ATs significantly disrupt protein interactions between other domains

within the module.[52] To address this, an efficient chimeragenesis strategy was employed to create hybrid DEBS/pikromycin modules for the production of engineered macrolactones and successfully discovered a significant portion of active hybrids.[53] The throughput this platform, although improved over previous efforts, is still limited by the need to carry out a chromatographic separation for each chimera tested. Further investigations have reinforced the importance of the ACP/KS interaction epitope in chimeric systems,[54] as poor activities could be improved through mutagenesis.[54-55] More recently, Keasling and co-workers aimed to more systematically design boundaries for AT replacement, which may reinvigorate PKS engineering through AT swaps.[56] Other strategies have focused on identifying amino acid mutations that can reprogram the extender unit selectivity of a given AT, thus enabling production of regioselectively modified polyketides if the target extender unit is available to the host strain.[57]

The power of combinatorial biosynthesis to tailor the structures of polyketides is evident from the examples described above. Biosynthetic pathways can be predictably modified through inactivation, swapping, and deletions to yield non-natural analogues of natural products. Most of these examples take advantage of natural enzymatic and small molecule components, thus limiting the chemical space that can be accessed by combinatorial biosynthesis. For example, there is no known organism that natively produces propargylmalonyl-CoA, a potential extender unit that could be used to facilitate semi-synthesis via "click" chemistry. Accordingly, there is also no known PKS that natively utilizes propargylmalonyl-CoA as an extender unit. Subsequently, alternative methods to access non-natural chemical functionalities need to be developed. This has been the primary driving force behind the emergence of precursor-directed biosynthesis and mutasynthesis.

1.3.3. *Accessing and Installing New Substrates into Erythromycin and Other Polyketides*

As the quality and scope of structural and mechanistic data has improved, research has begun to move away from the model system of DEBS. Additionally, engineering PKS domains has become more precise, often focusing on single amino acid residues. At times, these engineering exercises occur in the traditional 'gatekeeper' domains while others have focused on eliminating bottlenecks in other domains.[29] Sherman and co-workers synthesized a series of non-natural pikromycin pentaketides to probe the substrate specificity of a module+TE fusion *in vitro*.[27] Use of molecular dynamics and quantum mechanics led to introduction of a serine to cysteine mutation in the TE that subsequently broadened the ability of the TE to macrocyclize non-natural substrates.[58]

Typically, the most common way to introduce non-natural moieties into a natural product is by taking advantage of the native promiscuity of an enzyme and supplying the desired substrate (precursor-directed biosynthesis).[59] Often this is a difficult strategy for modular systems where there may be several modules incorporating the same substrate. Weber and co-workers took advantage of a uniquely promiscuous ethylmalonyl-CoA *trans*-AT from the kirromycin pathway, KirCII, and fed in two malonate derivatives, allyl and propargyl, which were then each converted to the corresponding CoA thioesters and incorporated into the mature final products *in vivo*.[60] The propargyl-kirromycin analogue provided a handle for further modification via "click chemistry" that enabled the addition of a fluorescent tag. While this new compound was not a better antibiotic than kirromycin, the ability of KirCII to facilitate the regioselective modification of a hybrid polyketide/non-ribosomal peptide with non-natural chemical functionality is remarkable and could be leveraged to modify other natural products.

Several groups have worked with the final module of the erythromycin PKS to engineer a shift in extender unit selectivity, but to date, the desired erythromycin analogues have been produced only as the minor product *in vivo*.[61-62] Previously, Williams and co-workers had utilized enzymatically-generated methyl- and propargylmalonyl-CoA and found a relaxed substrate specificity (8% propargyl incorporation by wild-type).[63-64] Perhaps surprisingly, a tyrosine to arginine mutation, located in the canonical YASH motif of methylmalonyl-CoA-specific AT domains, switched the selectivity of the module to 88% non-natural product. It remains to be seen whether this approach can be extended to other PKSs, which would then provide a powerful strategy to reprogram substrate selection and installation by type I PKSs. These findings highlight the scope of engineering these types of systems both *in vivo* and *in vitro*.

1.3.4. *Domains, Modules, and Linkers*

In combination with the work described above, a much larger chemical space can be accessed through more highly disruptive approaches to engineering PKSs. Often portrayed as being analogous to building with Legos, the swapping of modules and domains in these enzymatic pathways has long been a holy grail for the field. Recently, there have been a number of particularly notable advances that focus on defining and re-defining the domain and module boundaries that bring us closer to making this goal a reality. One of the more provocative current proposals is that the conventional KS_n-AT_n-ACP_n module that is typically exchanged in PKSs should actually be AT_n-ACP_n-KS_{n+1}.[41, 65-66] In addition, continuing mechanistic studies are providing insight into how assembly lines support vectoral biosynthesis, with the hope of better guiding future engineering of chimeric assembly lines.[67] Additionally, efforts are being made to categorize and develop tools for the design of these chimeras.[68]

As mentioned previously, early attempts at domain swapping have been met with mixed results.[25-26] Recently, the Keasling lab has made significant progress in delineating guidelines for replacing ATs and reductive loops (KR/ER/DH) without the loss in product yields seen previously.[56, 69] The newly defined AT boundaries were predicted from structural and sequence alignments and were examined by protein yield and kinetic parameters. The reductive loop swaps were also efficient; however, there are a few more caveats due to the KR acting as a proofreader when exposed to a non-native side chain.[70] The AT boundaries in particular are potentially very powerful targets for engineering PKSs for analogue production. While Keasling, et al. only tested natural AT domains for their swaps, it could be easily adapted to work in combination with engineered ATs such as those described above.

Intermodular linkers have also been the target of engineering along with the intramolecular domain swaps described above. For example, Nielsen and co-workers demonstrated significant tolerance with the makeup of intermodular linkers between PKS and non-ribosomal peptide synthetase (NRPS) modules in the construction of novel hybrid pathways.[71] This was taken to the extreme of replacing a native PKS-NRPS linker with red fluorescent protein (RFP), which maintained production of the expected product. Interestingly, these same modules would not work together in *trans*, showing that the linker at the very least was only required to keep the modules in close proximity.

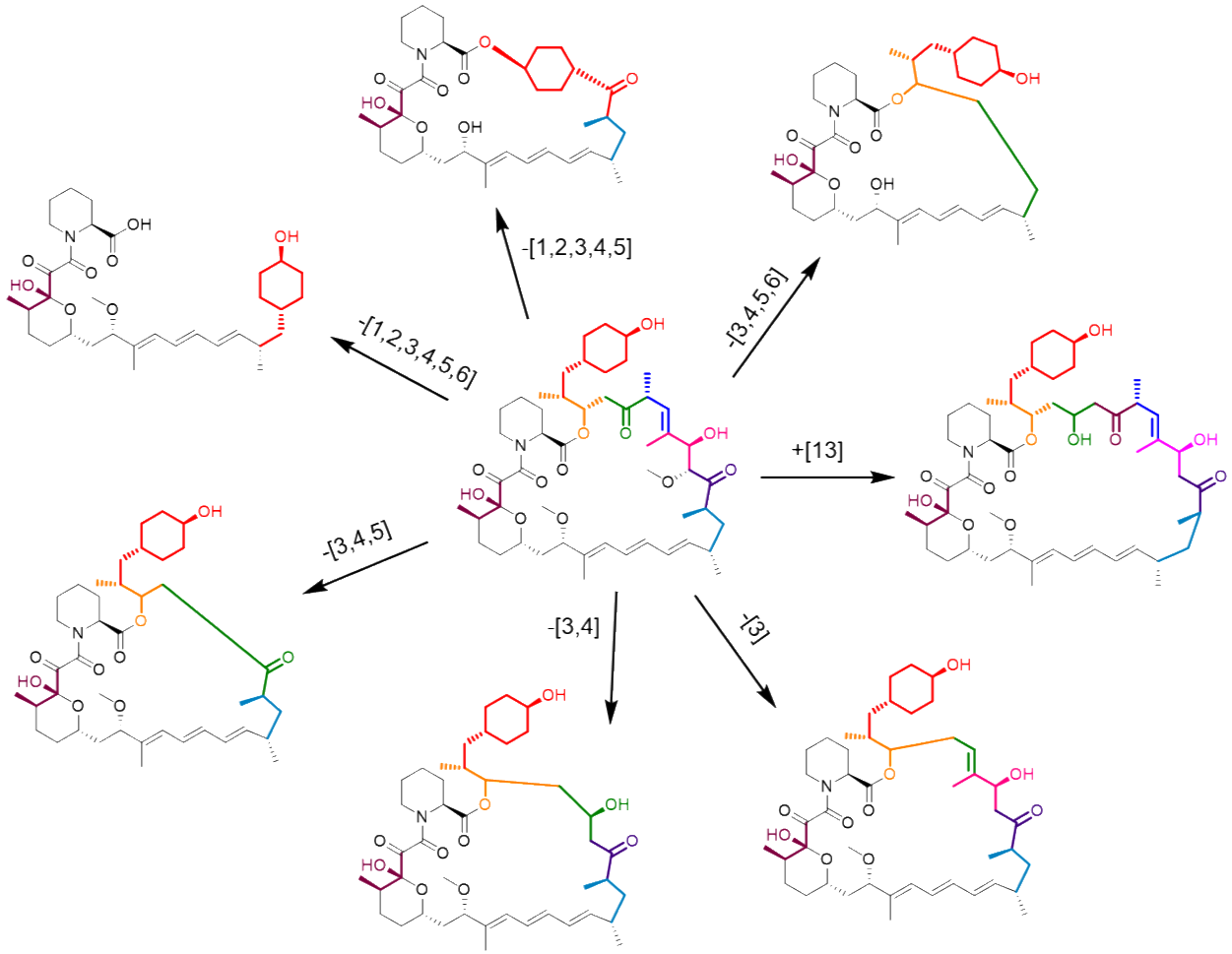


Figure 1.4. Scheme Illustrating the Diversity of Rapalog Generation via "Accelerated Evolution". Accelerated evolution of the rapamycin biosynthetic gene cluster led to recombination within the PKS modules, leading to the loss of combinations of modules 1-6 or the addition of a second copy of module 13.

In addition to mapping out boundaries, improvements in assembling and screening chimeric constructs have continued.[72] Several groups collaborated on a new method for laboratory-scale evolution of modular gene clusters they termed "accelerated evolution" or AE.[73] Their method is based on the idea that these assembly lines evolved through duplication and deletion events. Through forced recombination events of the 15-module rapamycin PKS-NRPS hybrid, new functioning constructs were created that were composed of between 9 and 16 modules (**Figure**

1.4). This approach was coupled with precursor-directed biosynthesis using 32 alternative starter units, yielding over one hundred new rapalogues.

1.3.5. *Metabolic Engineering and Synthetic Biology*

Ultimately, to produce novel polyketides and non-ribosomal peptides biosynthetically, optimized production and engineering hosts, useful precursor pathways, and tools for heterologous expression will be needed. Common strategies for polyketide production include deletion of competing pathways, overexpression of endogenous precursor pathways, or transfer to a more well-studied system.[74-75] Occasionally, heterologous genes can be added for non-natural precursor production using fed-in synthetic components.[64, 76-77] Finally, methods are required to screen constructs at faster rates than is traditionally accomplished.

Gene Cluster Engineering

Biosynthetic gene cluster (BGC) refactoring is an important aspect of metabolic engineering due to the complex balance between tailoring and assembly line genes with precursor-producing pathways. Additional tools have been developed for plug-and-play assembly of BGCs such as the use of the artificial gene operon assembly system (AGOS) to construct heterologous BGCs.[78] Others have stayed with more traditional methods like yeast homologous recombination as in Pfeifer's two-BAC erythromycin production system for *E. coli* (significantly more stable than the previous 5-plasmid system).[79-80] Additionally, genes predicted to pull metabolic flux away from the erythromycin pathway were deleted from the cell line, resulting in higher titers. On a larger scale, the Kelleher group created fungal artificial chromosomes (FACs) containing 56 complete BGCs ranging from 38-108kb in length.[81] Coupling FACs with metabolomics allowed them to

screen far more BGCs than has been typically done, resulting in the discovery of several novel fungal natural products. In contrast to heterologous gene cluster expression, Huimin Zhao and coworkers utilized CRISPR-mediated genome editing to activate silent BGCs in five *Streptomyces* species.[82] CRISPR has also been used as a genome editing tool to improve production of erythromycin A in its native host system, *Saccharopolyspora erythraea*.[83]

High-Throughput Screening

In addition to the traditional growth and pigment-based methods used to identify production of bioactive natural products, genetically-encoded biosensors offer another option.[84-85] The Williams group has developed a transcription factor-based biosensor, MphR, to detect erythromycin A, a polyketide antibiotic, and related macrolides.[85] In the presence of the metabolic product, green fluorescent protein (GFP) is expressed, allowing for high throughput screening of new biosensors as well as the biosynthetic pathways (**Figure 1.5A**). The authors hypothesized that the native macrolide promiscuity of MphR (**Figure 1.5B**) could serve as a powerful platform for creating biosensors with new properties. Indeed, directed evolution was used to furnish variant biosensors with tailored macrolide sensitivity and specificity (**Figure 1.5C**). The ability to construct biosensors with tailored properties, coupled with high-throughput methods to produce libraries of pathway variants, is expected to overcome many current limitations of PKS combinatorial biosynthesis.

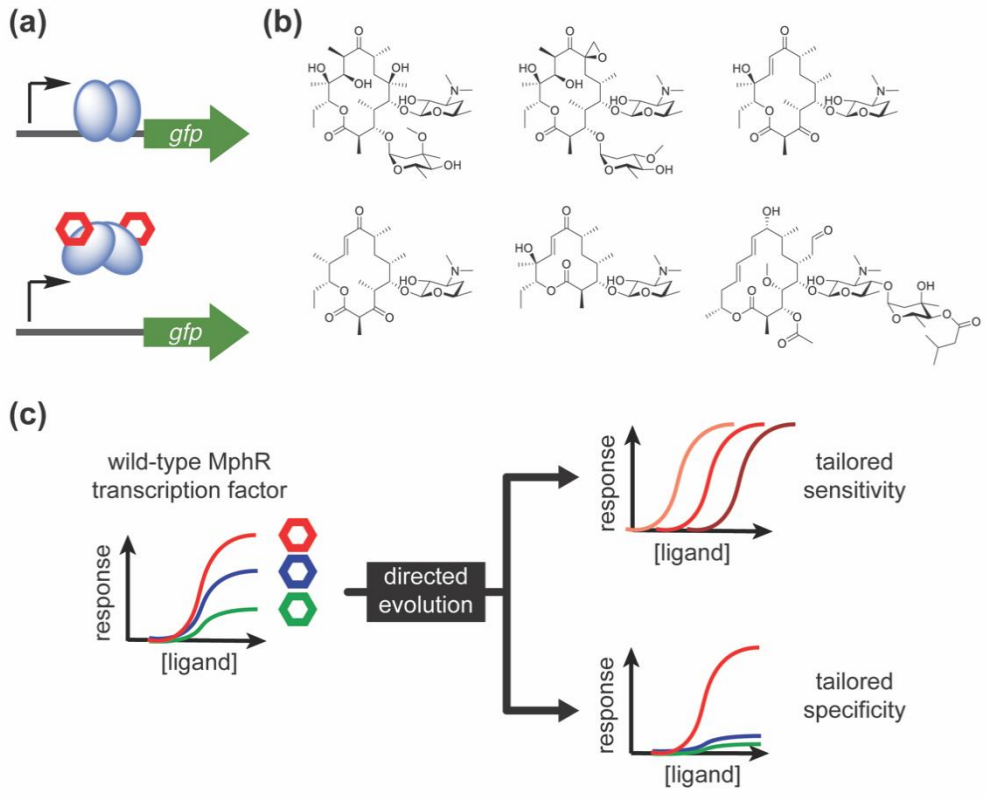


Figure 1.5. Genetically-Encoded Biosensors for High-Throughput Screening of Macrolides. A) Schematic representation of the transcription factor-based biosensor, MphR. The presence of an activating ligand (hexagon) causes a conformational change to MphR that results in the expression of GFP or another reporter protein. **B)** Some of the macrolides that are activators of the wild-type biosensor. **C)** The sensitivity and specificity of the biosensor can be tailored by directed evolution.

1.4. Terpene Diversification

1.4.1. *Terpene Biosynthesis*

Terpenes (isoprenoids, terpenoids) are produced primarily by plants and make up the largest and arguably most structurally diverse class of natural products.[86] This claim is evidenced by the wide range of sizes, cyclization patterns, and oxidation states, and their uses include pharmaceuticals, fragrances, flavors, agrochemicals, and other useful products.[87] Unlike most of the other natural product classes discussed here, terpene biosynthesis only requires two 5-carbon precursor isoprenes: a starter unit, dimethylallyl pyrophosphate (DMAPP), and an extender unit, isopentenyl

pyrophosphate (IPP). Terpene synthases will stitch together one or more IPPs onto a single DMAPP unit, resulting in the acyclic geranyl pyrophosphate (GPP, C_{10}), farnesyl pyrophosphate (FPP, C_{15}), or geranylgeranyl pyrophosphate (GGPP, C_{20}). Longer chains can be produced by the additions of GPP, FPP, or GGPP. Terpene cyclases can follow and promote cyclization of a myriad of products: mono- (C_{10}, e.g., limonene), sesqui- (C_{15}, e.g., caryophyllene), di- (C_{20}, e.g., taxadiene), tri- (C_{30}, e.g., lanosterol), tetra- (C_{40}, e.g., β-carotene), and polyterpenes (C_{50+}) (**Figure 1.6**). Even from the same cyclase, dozens of products are possible. Indeed, the γ-humulene synthase has been shown to produce 52 distinct sesquiterpenes.[88] This extreme variability is due to a carbocation-based cyclization mechanism, catalyzed by release of pyrophosphate. Post-cyclization modifications—often oxidations of the hydrocarbon scaffold—lead to the bioactive final products.

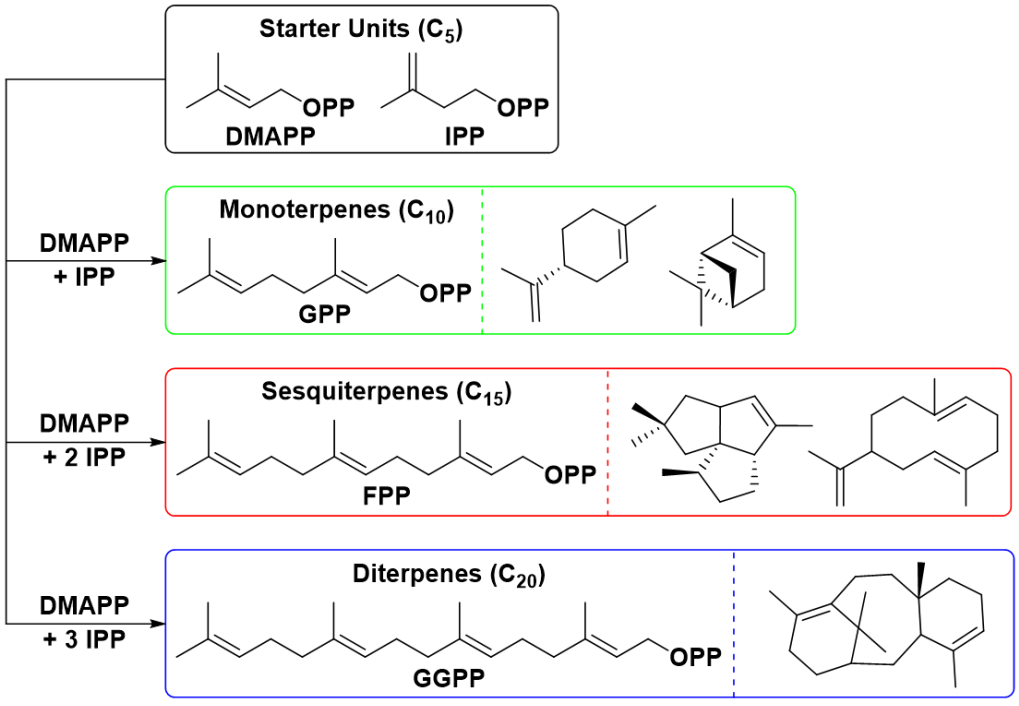


Figure 1.6. Terpene biosynthesis. Successive additions of IPP lead to linear compounds made up of 5-carbon isoprene units. Terpene cyclases can cyclize the acyclic compounds into more exotic and varied cyclic terpenes via carbocation mechanisms.

With >55,000 compounds isolated by 2007, terpene structural diversity allows for a range of bioactivities, but the chiral centers and quaternary centers of terpenes renders them difficult targets for total syntheses.[86, 89] While biomimetic approaches to terpenes has been explored, combinatorial chemistry—especially semi-synthesis—has been particularly useful to access new derivatives. Other research has focused on harnessing the inherently promiscuous product specificity of the biosynthetic enzymes. In both cases, the primary role of organic synthesis may be unique among natural product classes due to the extreme homogeneity and lack of functionality of terpene precursor molecules.

1.4.2. *Terpene Diversification*

The discovery of the potent antimalarial artemisinin led to the 2015 Nobel Prize in Physiology or Medicine after Youyou Tu isolated the endoperoxide sesquiterpenoid from the plant *Artemisia annua* (sweet wormwood). Until the past decade, artemisinin's only source was this herb, leading to supply and price fluctuations. Several combinatorial biosynthetic approaches over the past few years have helped produce artemisinin and its clinically-relevant semi-synthetic derivatives. In 2006, the Keasling group reported the production of the artemisinin precursor, artemisinic acid, in *Saccharomyces cerevisiae* (yeast), providing an efficient biosynthetic route to a semi-synthetic platform for medicinal chemists.[90] This was followed up six years later with production of dihydroartemisinic acid, a more advanced precursor, by the same group.[91] In 2013, the four-step semi-synthesis of artemisinin from microbially-produced artemisinic acid was published with a 40-45% overall yield.[92] These synthetic biology platforms have built a road map for future semi-synthetic approaches for valuable natural products and analogues.[92-94]

In an apparent blow to the combinatorial chemistry community, the new source of artemisinin had only a modest impact on the overall market, and in 2015, just one year after initial production of semi-synthetic artemisinin, Sanofi halted production.[95] Nevertheless, two years later, the Allemann group published an alternative approach to the semi-synthesis of artemisinin.[96-97] Instead of utilizing a series of enzymes to produce artemisinic acid like the work in engineered yeast or *A. annua*, chemically-synthesized 12-hydroxy-FPP was used with amorphadiene cyclase to give dihydroartemisinic aldehyde in a single biosynthetic step.[96] This intermediate bypasses a transition metal-catalyzed hydrogenation from previous work. A small-scale semi-synthetic route to artemisinin was published using this approach.[97]

Several groups in recent years have explored the use of chemically-synthesized isoprenoid precursor analogues as inhibitors and mechanistic probes for terpene cyclases.[98-105] For some analogues, like 3-bromo-FPP and GPP, the electronics are too different from the natural substrates for cyclization, leading to inhibitors that could be used for structural studies of these enzymes.[103] This is a powerful tool for a class of synthases where the catalytically-active conformation has yet to be captured as a crystal. Other synthetic precursor analogues can be cyclized. Of particular note, 6-fluoro-GGPP was incubated with taxadiene synthase to show that the previously accepted stereochemistry of an intermediate was incorrect.[102] The Allemann group and others have used deuterated, methylated, and fluorinated precursors to great effect, especially through the production of more stable versions of germacrene, a sesquiterpene that affects the olfactory response of insects.[98-100, 106-107]

Taking advantage of the natural promiscuity of terpene cyclases is a useful combinatorial tool to produce terpene analogues; however, the lack of product control can be a disadvantage as well. To combat this, researchers have mutated many putative plasticity residues in the more promiscuous cyclases, like the previously mentioned γ-humulene synthase.[88, 108-109] Yoshikuni, et al. displayed impressive control and understanding of the cyclization mechanism through rationally designed mutations.[109] The wild-type enzyme had seven major products, but a series of one to five mutations led to seven enzyme variants producing significantly shifted product profiles towards each of the major products. The authors noted specific mutations promoted predicted cyclization patterns, and this knowledge in principle could be combined with precursor-directed biosynthesis to give even more varied analogues.

Recently, biosynthetic studies of a previously poorly characterized diterpene biosynthetic pathway led to the development of an efficient and impactful semi-synthetic platform. Pleuromutilin is an antibiotic produced by *Clitopilus passeckerianus* that has served as a precursor for semi-synthetic derivatives used in veterinary and human medicine. Although several semi-synthetic derivatives of this diterpene antibiotic are being produced, large-scale production of designer semi-synthetic analogues requires an enhanced understanding of the biosynthetic pathway to pleuromutilin. Genetic characterization of the steps involved in pleuromutilin biosynthesis, through heterologous expression in *Aspergillus oryzae*, led to a panel of mutant strains that accumulate each pathway intermediate. Subsequently, *A. oryzae* was established as a platform for bio-conversion of chemically modified analogues of pleuromutilin intermediates, and was leveraged to generate a semi-synthetic pleuromutilin derivative with improved antibiotic activity.[110]

In addition to the previously described diversity, some natural products, like meroterpenoids (isoprenoid-polyketides), are synthesized by hybrid pathways. These products often involve the prenylation of small aromatic molecules, like in the case of the anti-oxidant naphterpin. Orf2, the promiscuous aromatic prenyltransferase from the naphterpin pathway, has been shown to accept a panel of 1-, 2-, and 3-ring aromatic prenyl acceptors.[111] Of note, a geranyl moiety was transferred to both aromatic carbons and exocyclic oxygens, depending on the substrate. In a similar example, *in vivo* production of an analogue of pyripyropene A was accomplished by feeding in benzoic acid instead of nicotinic acid.[112] Yet another prenyltransferase, CnqP3, has shown not only aromatic prenyl acceptor promiscuity, but also isoprenoid substrate promiscuity (DMAPP and GPP).[113] These examples of natural prenyltransferase promiscuity notwithstanding, the scope and utility of terpene diversification could be further expanded by rational design and site-directed mutagenesis of the component enzymes.[114]

1.5. Summary and Outlook

The many opportunities presented by PKS assembly lines, their associated biosynthetic machinery, and the unpredictable terpene synthases is matched only by the equally numerous challenges that face the 'Design-Build-Test' mantra of synthetic biology. Assembly line PKSs especially offer great potential for on-demand access to natural and non-natural natural products. Although structural and mechanistic insights are slowly emerging and will contribute to successes in rational reprogramming of biosynthetic gene clusters, high-throughput approaches that test the function of large libraries of assembly line and pathway variants are likely to yield engineered production strains. In particular, coupling advances in chimera design, gene cluster assembly, and high-throughput screening with systems biology approaches might provide fundamental design

principles for strain performance. A recent surge in the development of custom-made transcription factors that respond to target effector molecules provides a potentially generalizable and powerful platform towards this goal.

1.6. Scope of this Dissertation

The following work will discuss several approaches towards enabling the diversification of natural products, specifically polyketides and terpenes. In Chapter 2, molecular dynamics simulations of PKS acyltransferases are used to improve our knowledge of the mechanism of such enzymes and the roles individual residues and mutations play in substrate selection. Chapters 3 and 4 discuss the results of introducing known and novel mutations into the DEBS and pikromycin polyketide synthases, respectively. The results described therein expand our ability to introduce non-native moieties into polyketide backbones. In Chapter 5, a previously-described genetically-encoded malonyl-CoA biosensor is reengineered and utilized for detection of other malonyl-CoA analogues. This tool is especially useful for use as a screen for directed evolution of enzymes like MatB, a substrate-promiscuous malonyl-CoA synthetase. Chapter 6 describes efforts to develop a transcription factor-based biosensor for the high-throughput screening of short acyclic terpenes for use with natural and non-natural products.

CHAPTER 2

Computational Modeling of Polyketide Synthase Acyltransferases

2.1. Introduction

Acyltransferases (ATs) are the "gatekeeper" domains of polyketide biosynthesis and are responsible for selecting substrates for each condensation reaction. Accordingly, there is significant interest in manipulating the specificity of ATs in order to generate polyketide analogues. Directed evolution is a powerful strategy to engineer enzyme substrate specificity but critically requires a high-throughput screen or selection. Currently, such technologies that could be applied to AT domains have not been reported.[115] As an alternative, the mechanism and function of ATs has been extensively probed in an effort to ultimately rationally or semi-rationally redesign their substrate specificity, including attempts to identify residues responsible for selecting substrates and mutations that enable shifts in substrate selectivity. Much of the work has focused on the ATs from the DEBS PKS,[116] responsible for 6-deoxyerythronolide B biosynthesis (the core of erythromycin A), and this chapter will focus on AT6 from the same pathway.

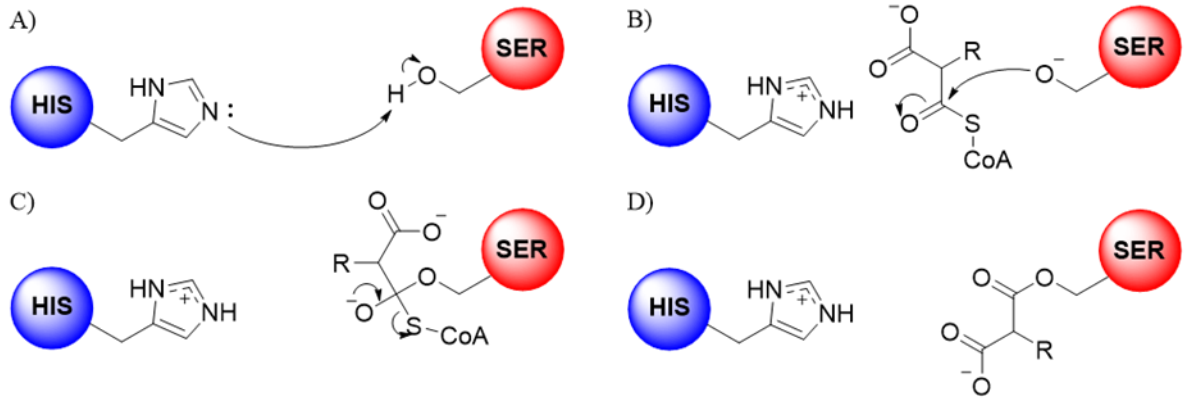


Figure 2.1. Mechanism of AT Selection of Extender Unit. A) A His residue acts as a general base to activate the catalytic Ser. **B,C)** If the extender unit is positioned correctly, the activated serine replaces the CoA. **D)** The malonate is now primed for attack by the ACP.

The AT mechanism is similar to that of the widely-studied serine protease family, utilizing a catalytic dyad of histidine and serine (**Figure 2.1**).[117-119] It has been proposed that the histidine (His747 in EryAT6) deprotonates the nucleophilic serine (Ser644 in EryAT6). After the malonate is covalently attached to the AT, the reverse mechanism occurs, with the phosphopantetheine arm of the acyl carrier protein (ACP) acting as the second nucleophile, resulting in a ping-pong bi-bi mechanism. The charged tetrahedral intermediate is likely stabilized by nearby glutamines (Gln616 and Gln645 in EryAT6).[62] Additionally, an arginine (Arg669 in EryAT6) in the back of the active site forms a stabilizing salt bridge with the free carboxylate of the malonyl extender unit (**Figure 2.2**).[117]

In addition to the mechanistic insights, recent literature has emphasized the role of conserved residues in substrate specificity. These include two motifs, the GHSxG motif containing the catalytic serine in the α/β-hydrolase large subdomain (the "right" side of the active site) and the YASH motif in the ferredoxin-like small subunit (the "left" side of the active site).[120] These motifs are often conserved within ATs that select the same malonate extender units. The better-known YASH motif includes the catalytic histidine and is so-called because of the four residues found in methylmalonyl-CoA specific ATs, including all six modular DEBS ATs. The native YASH motif of EryAT1 has been swapped with HAFH (the malonyl-CoA specific motif) and HASH (a hybrid motif).[121] Each engineered motif-swap shifted the strict methylmalonyl specificity of the enzyme toward incorporation of malonyl (40:60, methylmalonyl:malonyl) or led to relaxation of specificity (60:40, methylmalonyl:malonyl), respectively.

While this early success was promising, it was not until 2013 that two major breakthroughs in AT engineering were published. Our lab described the native promiscuity of EryAT6 *in vitro* with regard to a varied panel of malonyl-CoA derivatives.[64] Previously, a YASH motif-containing AT like EryAT6 was expected to take only methylmalonyl-CoA; however, productive activity was seen with derivatives such as allyl, methoxy, and phenylethyl. Additionally, Sundermann and coworkers described the V742A mutation in EryAT6 for incorporation of a non-natural propargyl extender unit *in vivo*, though the use of a synthetic substrate mimic likely prevented them seeing the same native promiscuity noted earlier.[62] The Sundermann study was assisted by molecular dynamics (MD) simulations and ligand docking, lending some insight into the AT active site not previously achieved through sequence analysis and crystal structures. This chapter will use MD simulations as a framework to rationalize the impact of mutations identified by the Williams group and others and use this as a platform to design second generation mutations that could impact AT specificity.

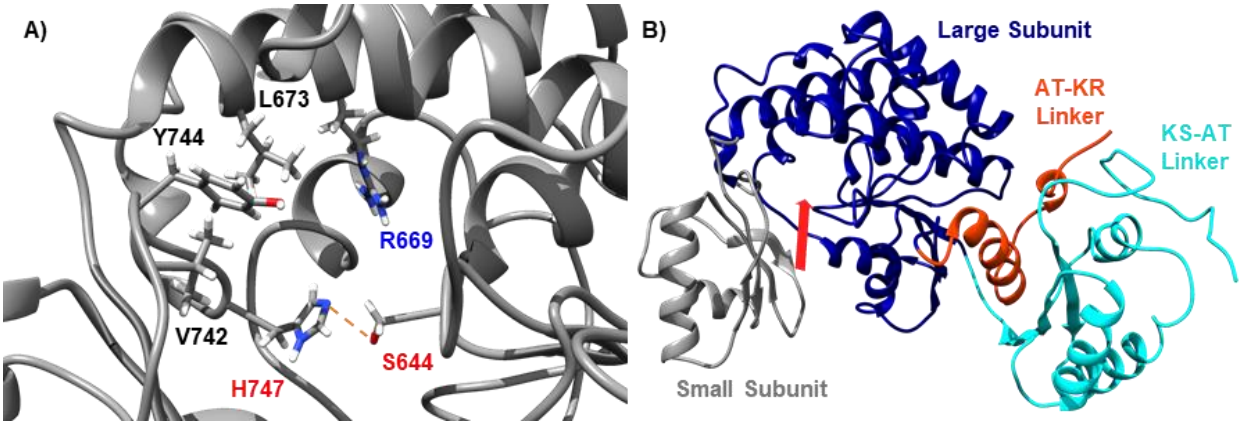


Figure 2.2. Residues and Regions of EryAT6. A) The active site architecture of EryAT6 with its catalytic residues (**red**), the required arginine (**blue**), and three residues previously targeted for mutagenesis (**black**). The distance between the catalytic dyad is indicated with an orange dashed line. **B)** The AT itself is divided into a small subunit (**grey**) and a large subunit (**dark blue**), while flanked by an N-terminal linker to the KS (**cyan**) and a C-terminal linker to the KR (**orange**). The substrate enters the active site via the path indicated by the red arrow.

Our lab has previously identified mutations at residues Leu673 and Tyr744 in EryAT6 (**Figure 2.2**), distinct from those reported by Sundermann et al., that lead to dramatic shifts in extender unit specificity. The selectivities of the wild-type (WT) and mutant DEBS module 6 (Ery6) enzymes

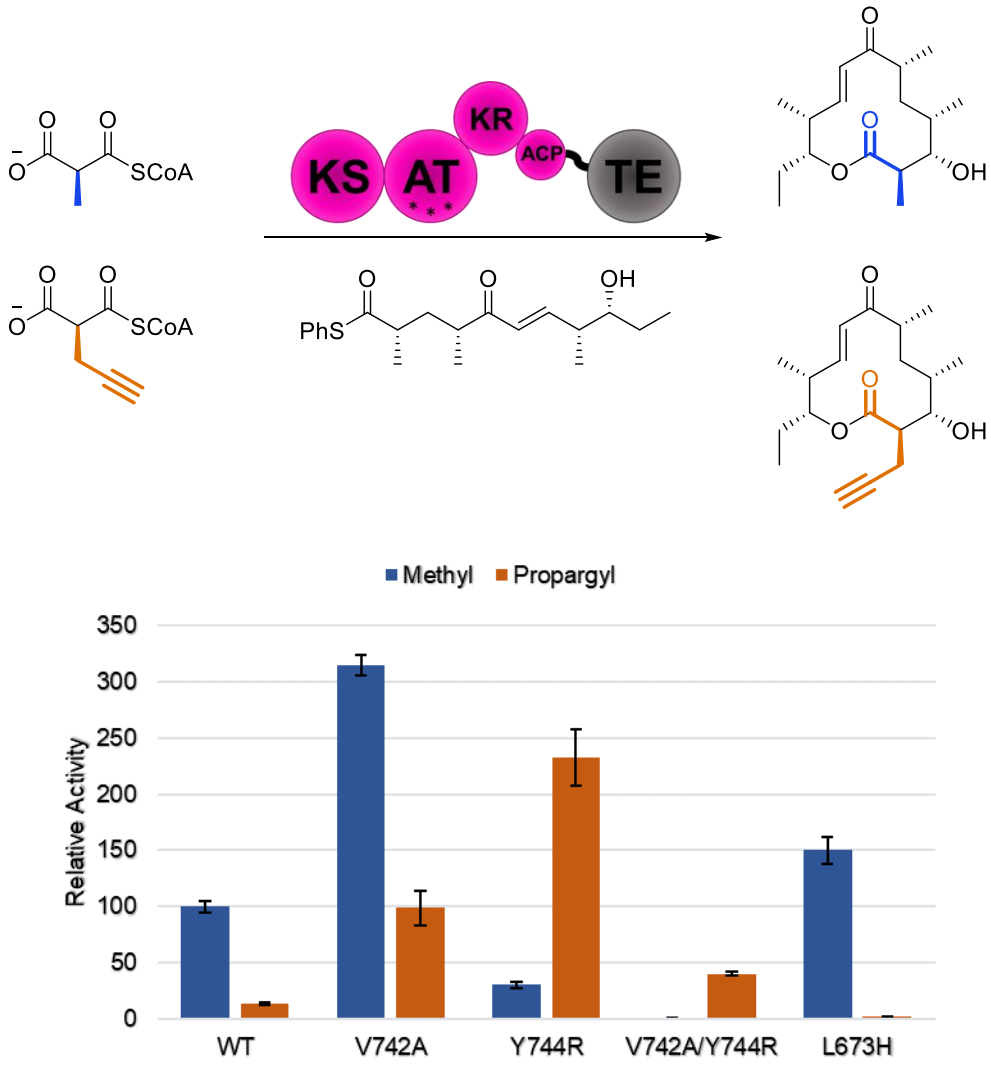


Figure 2.3. Relative Activities of Published Mutations in Ery6TE. Koryakina, et al. demonstrated the ability to shift selectivity from methylmalonyl-CoA to propargylmalonyl-CoA with one or two active site mutations. L673H nearly abolishes activity with the larger extender unit. Data shown are from competition assays with both substrates present.

were determined by a competition assay using methylmalonyl-CoA (mmCoA) and propargylmalonyl-CoA (pgmCoA). Notably, this analysis revealed that the V742A mutation described by Sundermann, et al. increases the overall efficiency and the ratio of non-natural to

natural incorporation by EryAT6, but still leads to a mixed product profile (**Figure 2.3**).[61-63] In contrast, Y744R, which is located in the YASH motif described previously, results in an 8-fold excess of propargyl incorporation over methyl incorporation. The combination of V742A and Y744R leads to the production of almost entirely one product, the propargyl analogue, albeit at the lowest conversion of any of the constructs. To create an orthogonal mutant, L673H resulted in an AT with nearly a 100-fold preference for the smaller natural mmCoA (~2% propargyl product) (**Figure 2.3)**. In this chapter, MD simulations of these mutants are used as a basis for describing their effects and for prediction of novel mutations.

2.2. Results

2.2.1. *Design and Evaluation of Models*

The goal of this project was three-fold: (1) to provide explanations for the roles of known selectivity-shifting AT mutations, (2) to identify new mutations *in silico* to complement or replace the current orthogonal mutations, and (3) to test these new mutations *in vitro*. This chapter will describe the efforts towards the first two goals, while the third will be described in Chapter 3. These results will lead to a better understanding of the molecular basis for substrate specificity.

The first step for *in silico* modeling of EryAT6 was to develop a reliable homology model, as there is no existing crystal structure. In addition to the value of having a stable model for MD simulations, appropriate boundaries for the AT needed to be determined in order to define what constitutes a functional AT *in vitro* or *in vivo*. With limited information available about PKS interdomain linkers, any new structural knowledge could greatly increase the ability to engineer functional modules. The two most similar crystal structures by sequence are of EryAT3 (PDB:

2QO3) and EryAT5 (PDB: 2HG4).[122-123] These structures were used as a basis for the online I-TASSER server to generate several models side-by-side (**Figure 2.4**).[124-126] The AT boundaries were decided by manual inspection of the protein structures and sequences. Prior kinetic work by Dunn, et al. with an excised EryAT3 included only the N-terminal linker (i.e. KS-AT linker, **Figure 2.2B**), but published studies have not compared ATs with or without their N/C-terminal linkers.[127]

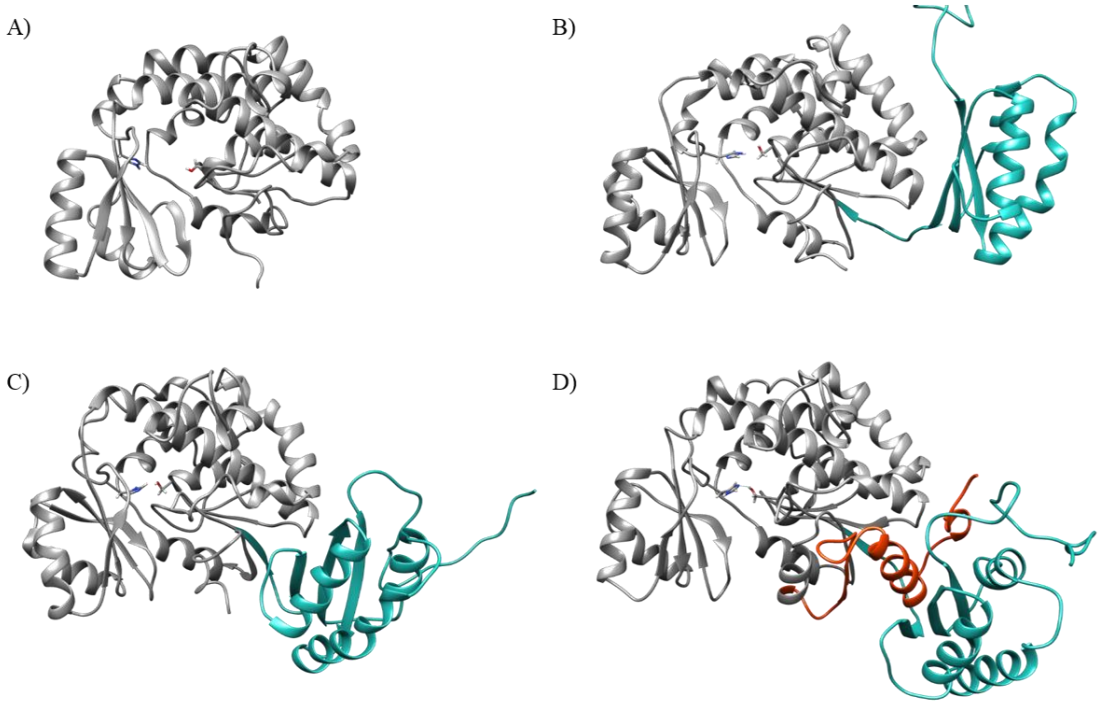


Figure 2.4. Evaluation of EryAT6 Models. EryAT6 simulated with **A)** no linkers and HIE 747 (278 residues), **B)** an N-terminal linker (**cyan**) and HIE 747 (387 residues), **C)** an N-terminal linker (**cyan**) and HID 747 (387 residues), and **D)** an N-terminal linker (**cyan**), C-terminal linker (**orange**), and HID 747 (430 residues).

The effect of the linker(s) on the stability of each AT was clear almost immediately upon initiation of MD simulations. Without any linkers present, the AT (**Figure 2.4A**) began to fall apart, with an average catalytic dyad distance of 8.24 Å. This is a far greater distance than would be expected for two residues involved with proton transfer. Notably, inclusion of the 109-residue N-terminal linker resulted in an average catalytic dyad distance of 4.13 Å **(Figure 2.4B**). There was also a noticeable

improvement in the conservation of secondary structure throughout the model; however, the active site still did not appear consistent with the accepted mechanisms. Changing the computationally-assigned protonation state of the catalytic His747 from HIE (protonated ε nitrogen) to HID (protonated δ nitrogen) in the N-terminal linker model resulted in the expected hydrogen-bonding in the catalytic dyad and a subsequent decrease in average distance to 3.03 Å (**Figure 2.4C**). This model was used primarily until a collapse of the beta sheet in the larger subunit of the AT began after tens of nanoseconds. Despite the results from Dunn, et al., a 43-amino acid C-terminal linker was added to the model (**Figure 2.4D**). This left the catalytic dyad distance essentially unchanged, but the stability of the overall structure was improved, as judged by secondary structure longevity. Due to these findings, this model, with a C-terminal linker, was selected as the template for all future MD simulations of EryAT6. Gratifyingly, a recent publication provided a comprehensive look at the boundaries of an AT domain that validated the decision to include the C-terminal linker compared to the Dunn model.[56] Furthermore, the newly developed rules regarding delineation of appropriate protein boundaries predicted that the sequence of EryAT6 should span from Gly439 to Ala880 (442 residues), which was very similar to the stable EryAT6 sequence from Ala447 to Ala876 (430 residues) predicted by the MD-guided investigation here.

2.2.2. *Molecular Dynamics Simulations of EryAT6 Wild-Type and Mutants*

With the most stable model identified, the EryAT6 wild-type structure was neutralized and solvated in water. The resulting set-up was minimized, heated, and simulated for 60 ns with AMBER14.[128] Once structurally converged (~30 ns, based on RMSD and manual inspection), residues for each mutation were individually changed in a randomly-selected snapshot from the

wild-type trajectory. Initially, this was done for the previously described mutations V742A, Y744R, V742A/Y744R, and L673H.

Once the 30 ns simulations were run for each of the mutant ATs, the trajectories were inspected for over-time trends (e.g., flexibility) and with single-timepoint snapshots. One important factor considered was the distance between the catalytic residues Ser644 and His747. Due to these residues residing on opposite sides of the active site, this distance could give a good approximation of both the distance between the two subunits and also the fraction of time the enzyme was in a catalytically-competent conformation. As expected, there were significant differences of this distance between wild-type and mutant enzymes (**Figure 2.5**). For example, while the wild-type enzyme kept the catalytic dyad between 2-4 Å for 50% of the simulation, the Y744R mutant was in that conformation for less than 25% of time. In the remainder of the Y744R simulation, the two subunits were further apart, hinting at a larger active site that can accommodate larger substrates. In contrast, the catalytic dyad for L673H remained in the same 2-4 Å range for nearly the entirety of the simulation. The double mutant V742A/Y744R shows decreased overall activity, which fits well with the simulation showing a consistently large active site mostly outside of the 2-4 Å window.

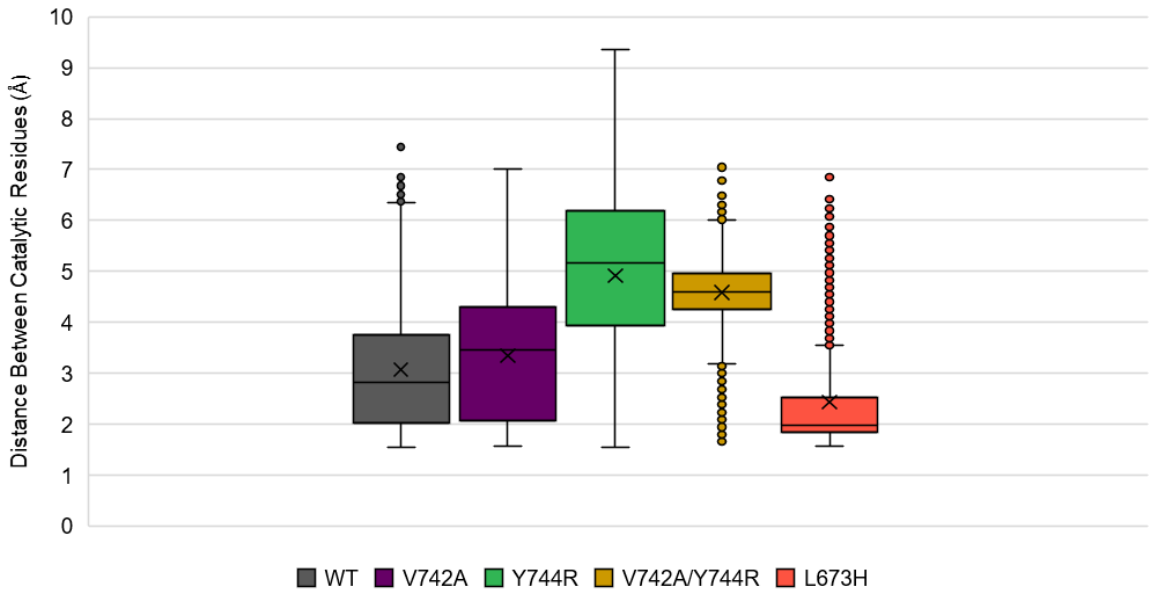


Figure 2.5. Distances Between AT Catalytic Residues Over Time by Model. The distances between the catalytic residues, Ser644 and His747, were measured for each frame and plotted. The X represents the average distance. The middle 50% of distances fall within each box. Median distances are represented by horizontal lines.

While the two catalytic residues have obvious importance to the activity of the AT, it was not immediately clear which structural changes accounted for the differences between the mutations. For the purpose of determination, the Root Mean Square Fluctuation (RMSF) was calculated for each residue as a measure of local behavior. RMSF is used to describe the displacement of atoms relative to an average, equilibrated structure over time. RMSF for the core AT structure of the wild-type and mutant EryAT6 is shown in **Figure 2.6**, with the small subunit region highlighted in blue. The linkers were excluded due to high variability. These results provide more insight into the overall motion of the macromolecule, highlighting for example the high RMSF values for the small subunit of the V742A mutant relative to the large subunit or to the wild-type enzyme (**purple**). In contrast, the increased mobility of the Y744R enzyme compared to wild-type is seen throughout the entire molecule, especially in the large subunit (**green**).

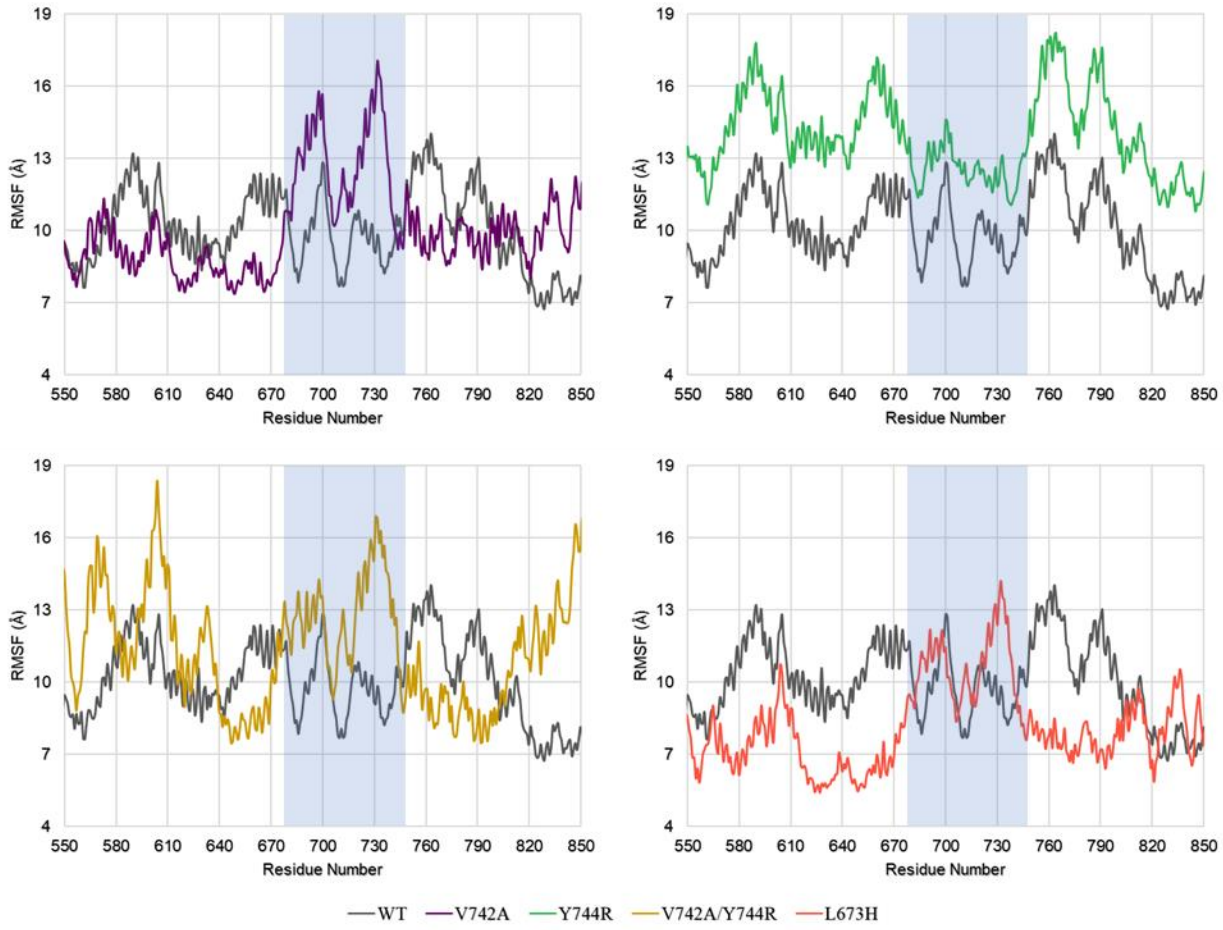


Figure 2.6. RMSF of EryAT6 Wild-Type and Mutant Models. RMSF was calculated from 60 ns (wild-type) or 30 ns (mutants) MD simulations. Higher values correspond to increased movement of a residue over the time frame. The blue box represents the residues found in the small subunit of the AT. Linkers were excluded from the figure.

Fitting with the expected trends, there is a generally lower RMSF for the methyl-specific L673H mutant, predicting a more constrained, less flexible active site. For the double mutant, a combination of effects from V742A and Y744R might be expected assuming simple additivity, and that is what appears to be the case. Higher RMSF values are seen in the small subunit relative to Y744R and in the large subunit relative to V742A. To correlate these data with structure, snapshots were randomly selected and manually inspected (data not shown).

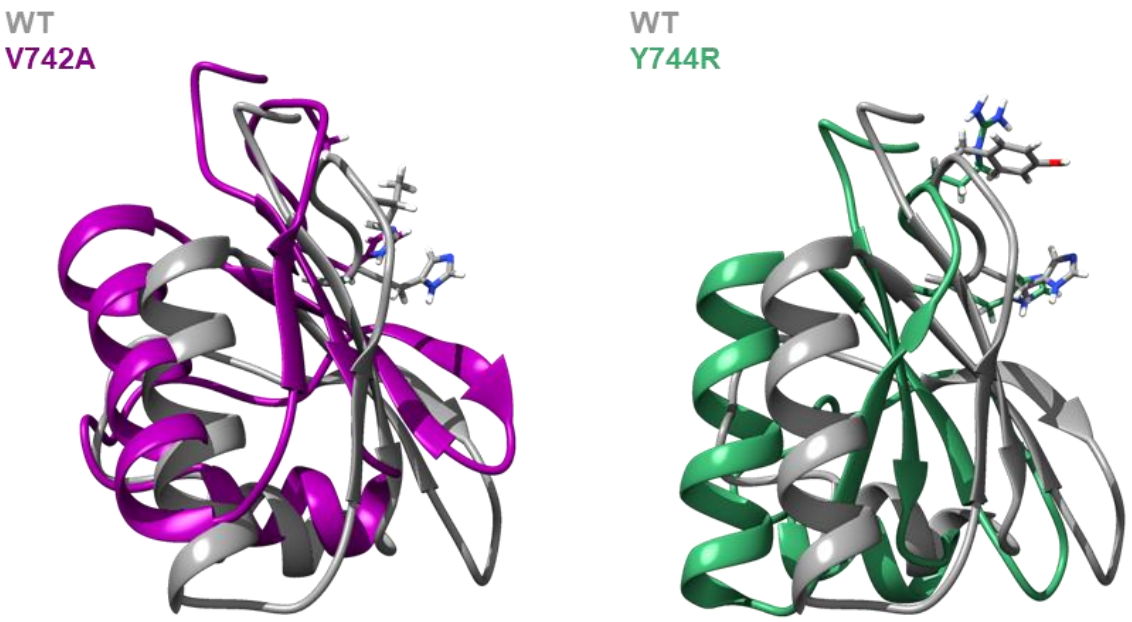


Figure 2.7. Overlays of Small Subunits. Left. The small subunits from EryAT6 wild-type and V742A were overlaid. The catalytic His747 and the residues at position 742 are shown as sticks. **Right.** The small subunits from EryAT6 wild-type and Y744R were overlaid. A significant perturbation is seen inside and outside the active site in both cases, though to different degrees. The catalytic His747 and the residues at position 744 are shown as sticks.

The V742A mutation was proposed by Sundermann, et al. to work simply by providing more space for the propargyl group to enter the active site by decreasing the size of the residue from valine (139 $Å^3$) to alanine (88.6 $Å^3$).[129] While this explanation may have some merit, it does not explain why, for example, the V742G mutant also reported in the Sundermann paper is not functional.[62]

Very early in the simulation of the V742A mutant, the small subunit of EryAT6 moved further away from the active site (**Figure 2.7**), resulting in a more accessible catalytic dyad. This is likely due to the increased flexibility of the YASH loop gained by the smaller amino acid at position 742 identified by Sundermann. In the wild-type enzyme, Val742 remains consistently projected into the active site. Upon mutation to alanine, the entire loop can rotate downward into a previously inaccessible conformation (not shown). In the inactive V742G mutant, simulations showed a small

subunit that was far too flexible, allowing the entire active site to fall apart (data not shown). By keeping the loop more rigid than in the glycine variant, the general shape of the active site is retained. Additionally, in the V742A mutant, Tyr744 of the YASH motif can rotate parallel to the active site entrance. In our proposed mechanism, the rotation of the tyrosine may be just as important for pushing selectivity towards the larger substrate as the increased movement of the small subunit. It has previously been hypothesized that the YASH motif acts as a size screen for the extender unit side chains taken by ATs. According to docking studies (Section 2.2.3), the malonate has to pass by much of the YASH motif to form a salt bridge with the stabilizing Arg669. This theory is consistent with the HAFH motif in malonyl-specific ATs (considerable steric bulk), the YASH motif in methyl-specific ATs (less bulk), and with other motifs for larger extenders, generally XAGH for ethylmalonyl-CoA, where X is usually Thr or Val.[130] As will be seen later in this dissertation (Chapter 4), active sites of ATs from other PKSs are not the same size as that of EryAT6, so a direct motif replacement like that done by Del Vecchio, et al. is unlikely to be universally effective at changing extender unit specificity.[121] The combination of the loop rotation, small subunit mobility, and the movement of the tyrosine described herein is thus a more comprehensive mechanism to explain extender unit selectivity than that proposed by Sundermann, et al.

The Y744R mutation, identified and characterized previously by our lab, results in an inversion of selectivity towards propargyl. Like the V742A mutation above, a 30 ns MD simulation showed an apparent increase in active site size. Unlike the V742A mutation though, the average distance between the catalytic dyad increases significantly, from 3.07 Å to 4.92 Å (**Figure 2.5**). In the simulation, this separation of the small subunit from the large subunit in the Y744R mutant is

caused by the disruption of a salt bridge between Arg674 (large subunit) and Asp743 (small subunit) in EryAT6 wild-type by the formation of a new salt bridge with Arg744 (**Figure 2.8**).

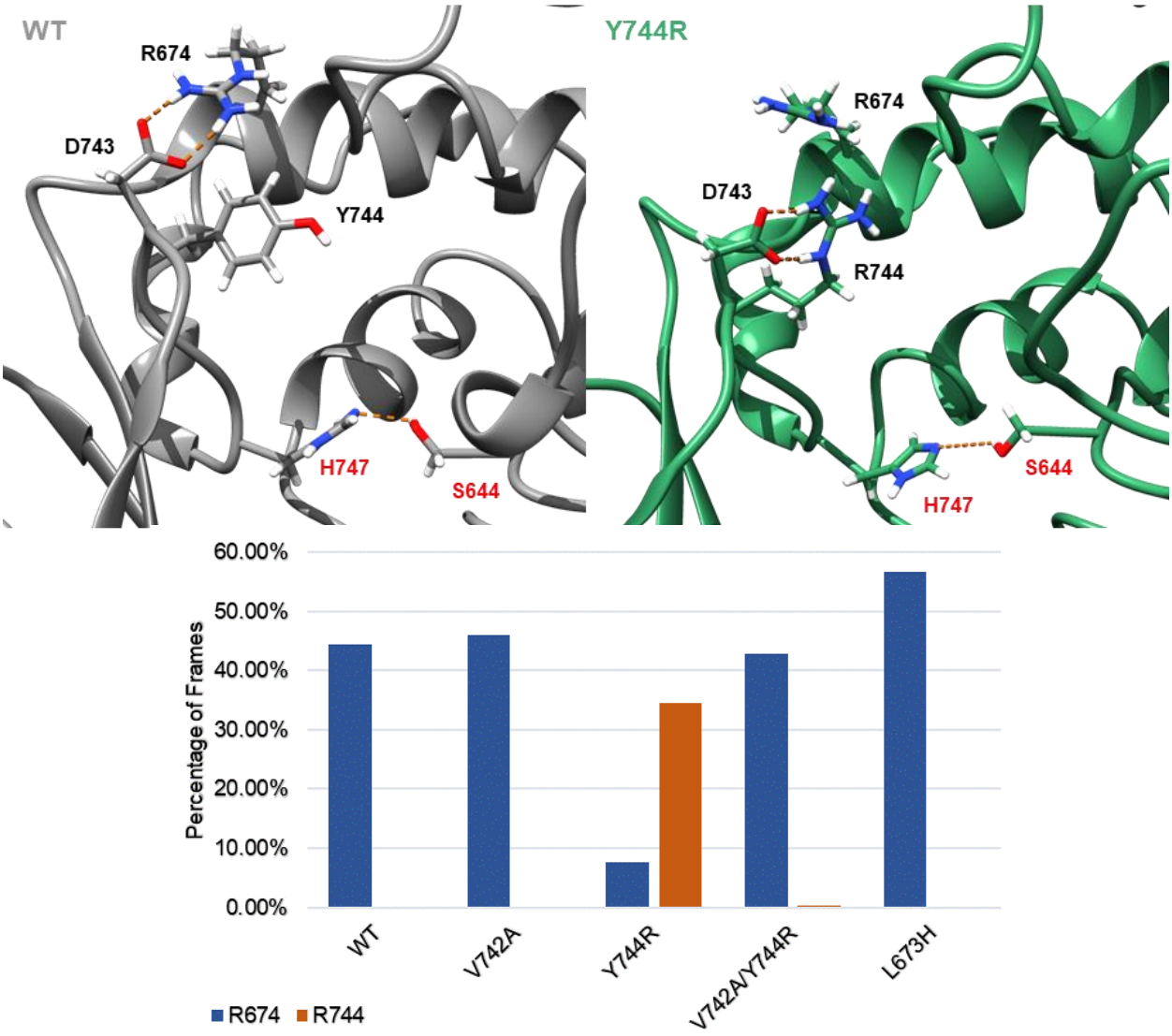


Figure 2.8. Snapshot of Y744R Mutation. Top Left. EryAT6 wild-type retains a salt bridge between Arg674 and Asp743 (orange lines) for much of the MD trajectory. **Top Right.** EryAT6 Y744R mostly eliminates the wild-type salt bridge by forming a new salt bridge between Arg744 and Asp743. **Bottom.** The percentage of frames from each simulation in which hydrogen bonding is seen with Asp743. A significant increase is also seen with the L673H mutation.

Once the new salt bridge is formed, one of the most important links between the two halves of the active site is lost, potentially allowing larger extender units to enter. The methyl group of the

natural extender would lose any stabilizing hydrophobic interactions, but a larger propargyl group may retain some of those interactions across a wider pocket.

Currently, it is difficult to determine the exact way in which the two mutations cooperate in the double mutant V742A/Y744R. It has some of the characteristics of the V742A mutant (the rotated YASH loop and flexible small subunit) and some of the Y744R mutant (the increased distance of the catalytic dyad and flexible large subunit). However, there is minimal disruption of the Arg674-Asp743 salt bridge in the double mutant, despite the presence of the arginine (**Figure 2.8**). While almost no methyl product is seen in the competition assay, this may simply be due to a lower catalytic efficiency compared with Y744R. Like in the V742G inactive mutant, this may be a case of too much flexibility hampering the ability of the enzyme to stay together long-term.

The Arg674-Asp743 salt bridge is not affected by the Y744R mutation alone. In the methyl-specific L673H mutant for example, although there was the destabilizing effect seen previously, there was an increase in the percent of simulation the natural salt bridge was in place (56.6% to 44.4%). This probably was not causal, but rather an effect of the closer proximity of the two subunits due to another effect.

In the wild-type enzyme, the two of the most important features responsible for keeping the correct active site width stable are the Arg674-Asp743 salt bridge and the hydrogen bonding of the catalytic dyad. In the L673H *in silico* model, π-stacking occurs between the two aromatic rings of Tyr744 and His673 (**Figure 2.9**). The average distance between the two ring centers is 3.4 Å, within the normal range of 3.3-3.8 Å for π-stacking. This interaction has the effect of pulling the

YASH loop (and the entire small subunit) closer to the large subunit. This also shrinks the distance between the catalytic residues to an average of only 2.43 Å. The smaller, more rigid active site thus should select for the smaller, natural mmCoA.

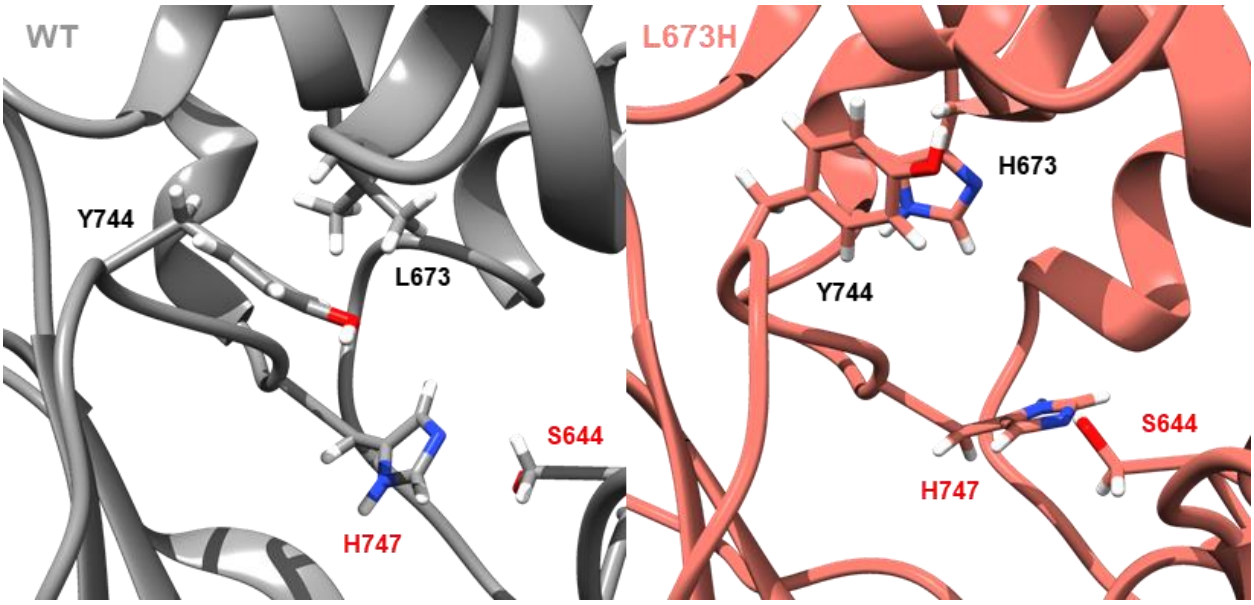


Figure 2.9. Snapshot of L673H Mutation. Left. In EryAT6 wild-type, Tyr744 only shows a slight hydrophobic attraction to Leu673. **Right.** In EryAT6 L227H, the aromatic rings of Tyr744 and His673 π-stack, resulting in an additional attraction between the two subunits and a narrower active site.

2.2.3. *Molecular Docking of Substrates in EryAT6*

Malonyl-CoA substrates were docked into the active site of EryAT6 for short MD simulations to visualize the effect on the structures upon addition of a substrate. To date, these extender units have only been visualized in the active site *in silico*, as there are no published crystal structures containing a non-covalently attached substrate. For the ping pong bi-bi mechanism to be efficient, the released CoA moiety must be able to leave the active site easily in order for the ACP to enter. When methylmalonyl-CoA was manually docked into the wild-type active site of EryAT6 and simulated in water, it was possible to observe the potential roles of various residues

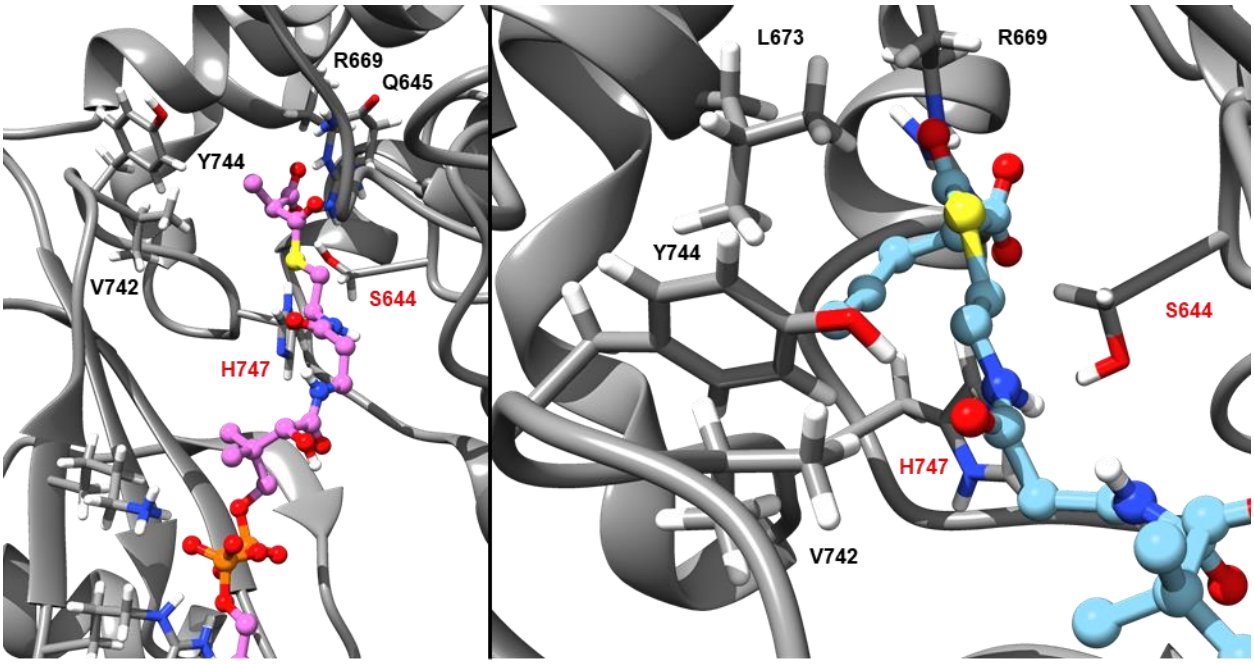


Figure 2.10. EryAT6 Wild-Type with Methylmalonyl-CoA and Propargylmalonyl-CoA. Left. Methylmalonyl-CoA was simulated for 10 ns in EryAT6 wild-type. **Right.** Propargylmalonyl-CoA was simulated for 10 ns in EryAT6 wild-type.

in catalysis, binding, and selectivity. For the CoA moiety though, there were few interactions as much of the molecule is solvent-exposed (**Figure 2.10**). The strongest interactions were between the phosphate region of CoA and the positively-charged Arg736 and Lys738. Additionally, it appears that Gln560 and His643 (not shown) are responsible for positioning the thioester directly above the catalytic serine—in disagreement with previously proposed hypotheses.[131] The importance of the salt bridge between the malonyl moiety and Arg669 is also readily seen in these models. Through much of the simulation, Tyr744 forms a hydrogen bond with the carbonyl of the thioester, likely helping to activate it for nucleophilic attack. The negatively-charged tetrahedral intermediate formed upon attack would presumably be stabilized by a combination of Gln616, Gln645, and the amide backbone.

In addition to the interactions in **Figure 2.10**, a series of docking models with wild-type and mutant ATs with both mmCoA and pgmCoA yielded informative results. Upon initiation of the simulations, most of the substrates remained within the active site; however, mmCoA left the Y744R active site almost immediately. This result was consistent with the *in vitro* results.

2.2.4. *Modeling ATs with Altered Catalytic Residues*

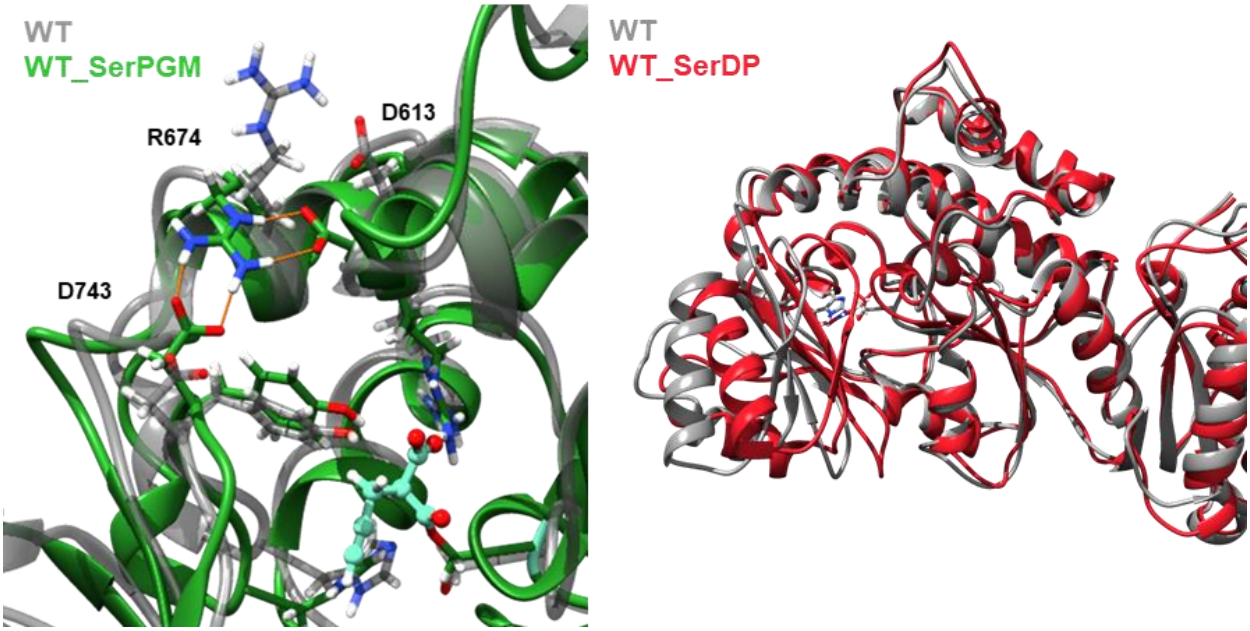


Figure 2.11. EryAT6 Wild-Type with Propargylmalonylated and Deprotonated Ser644. **Left.** EryAT6 wild-type overlaid by the wild-type enzyme with propargylmalonate (cyan) covalently bound to Ser644. **Right.** EryAT6 wild-type overlaid by the wild-type enzyme with a protonated His747 and deprotonated Ser644.

In the interest of gaining as much insight into the workings of the AT as possible, additional intermediate models were created for simulation. The purpose of this study was to determine not only the mechanisms of known specificity-shifting mutations but also to determine what aspects of the AT are important for the activity of the enzyme itself. To build these models, the catalytic residue Ser644 was replaced with a non-natural amino acid, either a propargylmalonyl-serine or a deprotonated serine, thus mimicking the expected seryl-*O*-ester intermediate formed upon self-

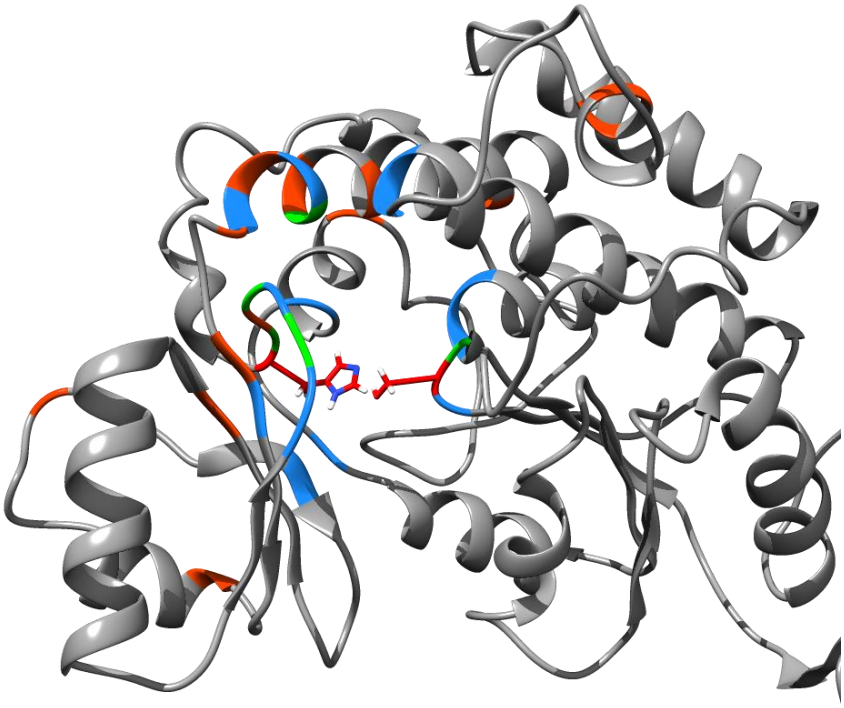
acylation of the active site Serine, and the catalytic His747 was replaced with a doubly-protonated histidine (as per the proposed mechanism in **Figure 2.1**). Each model, WT_SerPGM and WT_SerDP, respectively, was subjected to a 10 ns MD simulation. **Figure 2.11** displays the formation of a double salt bridge between Asp743-Arg674-Asp613 at the top of the active site only seen with the addition of a covalently-attached intermediate. Without the intermediate, Arg674 goes back and forth between the two aspartates, but the two subunits are not close enough together for both to form a salt bridge at the same time.

For the deprotonated serine model, the proton has been fully transferred to the histidine. While this mirrors the predicted mechanism of an AT, it is far more likely to be a very short-lived intermediate occurring simultaneously with attack of the malonyl-CoA thioester than the longer-lived intermediate simulated here (**Figure 2.11, right**).

2.2.5. *Predicting Novel EryAT6 Mutations* In Silico

The second aim of this project was to identify novel mutations that alter the AT extender unit selectivity by leveraging the insight gained from simulations of the known selectivity-impacting mutations. These mutations were chosen based on steric arguments (i.e. larger to smaller residues), disruption of hydrogen bonding networks, modulation of flexibility, or sequence alignments. To evaluate mutations, a panel of over 60 individual or combined mutations was subjected to 2 ns MD simulations. The sites and predictive origins of these mutations are shown in **Figure 2.12**. A complete list of simulated mutations with corresponding results and rationale can be found in Appendix A. From each MD simulation, a mutation was predicted to either narrow or widen the

active site. These new data could be used as a training set using feedback from *in vitro* results to better inform a second generation of substrate selectivity mutations.



Known Targets	Novel Targets Identified by MD	Novel Targets Identified by Sequence Alignment
L673; V742; Y744; S746; Q645	V612; D613; H643; E647; E648; R674; L676; M682; A707; L740; P741; V801	A596; E598; G662; A670; A672; A675; A677; G681; G689; D701; S717; A745; P790

Figure 2.12. Novel Targets for EryAT6 Substrate Selectivity. All residues targeted are highlighted in the EryAT6 structure, with some residues targeted for more than one mutation. Residues with published site-directed mutations shown to control substrate selectivity are in **green**. Novel residues that have not been previously demonstrated to control substrate selectivity in ATs were sorted into those identified by MD analysis (**blue**) or sequence alignment with ATs recognizing larger substrates (**orange**). Catalytic residues are in **red**.

The impacts of some mutations on the overall active site architecture were as expected. A subset of representative mutations is described in **Figure 2.13**. For example, V742P (instead of an alanine as previously described) yielded a narrower active site, as judged by simulation. As proline is the

most rigid canonical amino acid, the narrower active site was expected and could further strengthen the hypothesis that the flexibility of the YASH loop contributes to extender unit selectivity. Likewise, the D613E mutation opened up the active site more than in the wild-type model, as judged by simulation. The additional length of the glutamate allowed it to form a salt bridge with Arg674, competing with Asp742 without disrupting residues within the active site. With a larger active site, it could be expected to take larger substrates.

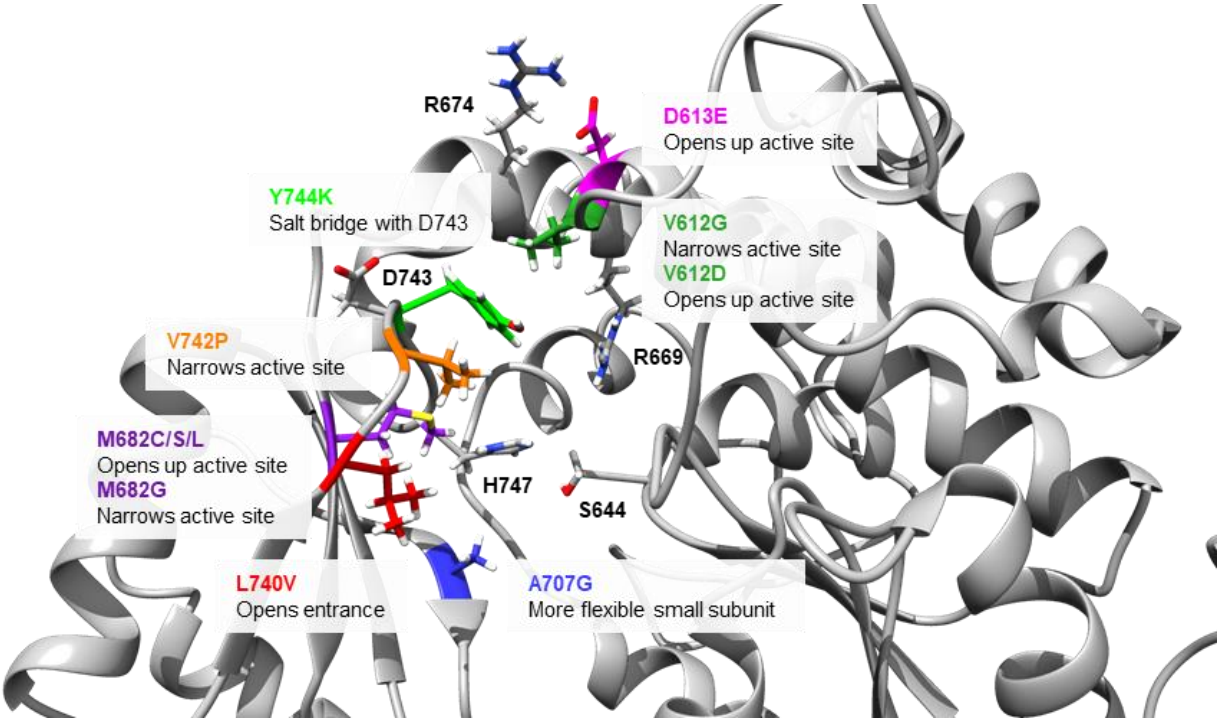


Figure 2.13. EryAT6 WT with Predicted Mutations. A sample of mutations screened in EryAT6 *in silico* with 2 ns MD simulations and the results for each.

Other mutations led to unexpected results in the simulations. For example, M682G, located near the YASH motif, resulted in a narrower active site despite such a drastic decrease in residue size. Though it is unlikely that many of these mutations act as originally predicted, being quickly able to narrow down candidate mutations *in silico* is invaluable compared with the significant cost and amount of time required for preparing and experimentally screening this number of mutant

enzymes. Many of these candidate mutations have been further screened using *in vitro* assays as described in Chapter 3.

2.3. Discussion

In this chapter, the power of computational tools for understanding the ATs of PKSs has been demonstrated. While any *in silico* model is limited in scope on its own, coupling computational predictions with *in vitro* and *in vivo* findings (e.g., Chapter 3) would provide a more accurate understanding of PKS ATs than any one approach alone. This workflow can be used to test predictions and provide a training set of mutations for further refinement of our *in silico* methodology. Plausible mechanisms that account for the impact of three mutations in ATs from two separate PKSs have been described. Prediction of additional mutations based on the structural changes in known mutants would be difficult, if not impossible, through more traditional methods such as sequence alignments. This is especially true for mutations based on a non-conserved residue, such as Arg674 of EryAT6, or when changes in large portions of the structure are involved. In addition to discovering mutations for shifting substrate specificity, new roles for other residues, such as the thioester-positioning His643 in the conserved GHSxG motif, have been proposed. Importantly, many of these results and hypotheses are easily tested via well-designed *in vitro* assays, which should greatly expand our collective knowledge on PKS ATs.

2.4. Materials and Methods

Wild-type homology models for EryAT6 were created using the I-TASSER online server.[124-126] Sequences are included in Appendix A. Mutations were introduced into structurally-converged wild-type models with Discovery Studio 4.1 from Accelrys Software, Inc. (San Diego, CA). The

ligands for docking were (2S)-methylmalonyl-coenzyme A and propargylmalonyl-coenzyme A and prepared as PDB files using ChemDraw 3D. Their corresponding force field parameters were derived using the PyRED server.[132] Ligand geometry optimization was carried out by Gaussian 09.[133] Molecular graphics and analyses of MD trajectories and PDB snapshots were performed with VMD 1.9.2 and UCSF Chimera 1.10.1.[134-136] Further analysis was performed with CPPTRAJ.[137] Images were rendered with POV-Ray.[138]

Covalent malonyl intermediates were created using non-natural amino acids in place of the catalytic serine. The PDB files were created in ChemDraw 3D as dipeptides with the structure N-acetyl-X-N'-methylamide, where X is the non-natural serine. These dipeptides were parameterized in a similar fashion to that of the free substrates using the PyRED server and Gaussian 09.

Using the AMBER14 software package, individual models' charges were neutralized with sodium ions in Xleap.[128] For docking studies, the protein models were then either manually docked with an energy-minimized methylmalonyl-CoA or propargylmalonyl-CoA. All models were then solvated with a 15 Å buffer of TIP3P water, also within Xleap. The enzymes and substrates were parameterized with ff12SB and GAFF force fields from the AMBER14 software package. Prior to production MD simulations, solvated systems were treated with four heating and seven minimization steps. Steps 2, 3, 5, and 11 heated the system to 300 K over times of 20-100 ps each. The first nine steps held the protein fixed, with the restraint constant being lowered each step. Steps 10 and 11 used no restraints. Minimization steps were completed when the change in the root mean square was below 0.01 kcal/mol•Å for the first two minimization steps and below 0.001 kcal/mol•Å for the remaining minimizations. Production simulations lasted between 30 ns and 100

ns for known mutants and wild-type models and 2 ns for predicted mutations. Step times were 2 fs. The non-bonded interaction cut-off was imposed at 9.0 Å.

CHAPTER 3

Exploring and Altering Substrate Selectivity in the DEBS Polyketide Synthase

3.1. Introduction

Type I polyketide synthases (PKSs) are giant multi-modular enzymes responsible for the biosynthesis of many members of one of the largest and most diverse classes of natural products.[2] The activities of type I polyketides range from antibacterial to anti-cancer and immunosuppressant, but each polyketide core is biosynthesized through PKS-catalyzed condensations and tailoring of individual malonyl-derived extender units. While similar in function to fatty acid synthases (FASs), PKSs increase the available diversity through varied oxidation and cyclization patterns in addition to a larger suite of extender units. Each module within the PKS is responsible for the incorporation of a single extender unit, selected by its cognate acyltransferase (AT) domain. These malonic acid thioesters are most commonly malonyl-CoA (mCoA) or methylmalonyl-CoA (mmCoA), with ethylmalonyl-CoA (emCoA) and a few rarer extenders found scattered in polyketides.[19] Through the iteration of even this small pool of extender units, a massive library of compounds is theoretically possible; however, most naturally-occurring extenders lack functionality that could otherwise dramatically alter the pharmacological properties of the corresponding polyketide and do not include chemical handles that could be leveraged for semi-synthesis.

This lack of extender unit variety is noteworthy in the widely-studied DEBS pathway responsible for the conversion of six mmCoA units into the macrolide core of erythromycin A.[139] Several PKS engineering strategies exist to alter the substrate selectivity of a given module, as described in **Figure 3.1**. With the first three of these techniques however, unfavorable or non-specific

intermodular or interdomain interactions result in low or non-existent yields. More recently,

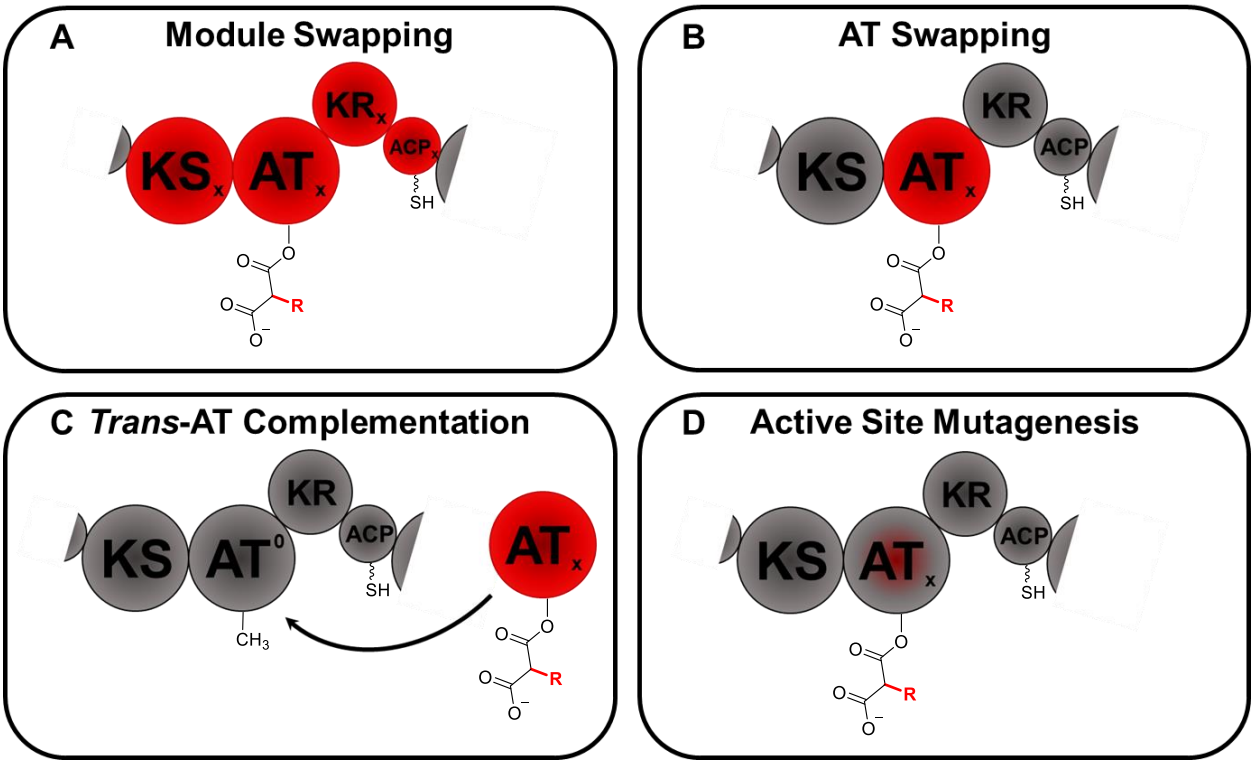


Figure 3.1. Strategies for Altering PKS Substrate Selectivity. A) Module Swapping is when a module natively recognizing a different substrate (x) or providing a different reduction loop replaces an entire native module. This early approach has resulted in a wide range of PKS activity levels without well-accepted boundaries. **B) AT Swapping** is when only the AT domain is swapped from a module with a different specificity into the module of interest. Like in **A**, these hybrid modules can suffer from lower activities. **C) *Trans*-AT Complementation** involves inactivating the native AT and supplementing with a *trans*-acting AT. However, this approach requires engineering specificity between an AT and its module. **D) Active Site Mutagenesis** can be used in conjunction with any of the first three strategies to introduce non-natural or non-native substrate selectivity in a given AT. Importantly, individual mutations are far less likely to negatively affect the modular PKS structure or its protein-protein interactions.

several labs have taken to making smaller changes in the AT active site to alter substrate selectivity that involve single-residues or motifs. The ultimate goal in this regard is to identify the exact residues responsible for substrate selection, and to use this knowledge to rationally alter the specificity of any AT by site-directed mutagenesis, in a similar way that the ten specificity-conferring residues identified in the adenylation domains of non-ribosomal peptide synthetases

has been leveraged.[140] With this is mind, it has been well-established that a small number of AT residues can predict or confer substrate selectivity, and there are a few conserved motifs—notably the YASH motif—tied to substrate selectivity.[141] It is rarely clear, however, what is the influence of each individual residue or motif. For example, in AT2 and AT3 from the epothilone PKS, there is a difference of only nine residues, but the corresponding substrate selectivity is mCoA in AT2 and is relaxed towards mCoA/mmCoA in AT3. Further investigation showed that the YASH motif alone was not responsible for this shift, even though it accounted for two of the nine amino differences between the two active sites.[121, 142] Replacement of the YASH motif in ATs of other systems with other natural motifs typically reproduces the desired, albeit often modest, change in extender unit specificity (see Chapter 4). More recently, the Schulz, Williams, and Zhou groups have each demonstrated that point mutations in or near the YASH motif of the AT6 of the DEBS pathway (EryAT6) enable incorporation of non-natural extender units, including propargylmalonyl-CoA (pgmCoA), benzylmalonyl-CoA, and others.[61-63, 143] Indeed, the YASH to RASH motif change resulted in a nearly complete inversion of selectivity (see Chapter 2 for more detail); however, it is a single mutation in the well-studied YASH motif and is not transferrable to every mmCoA-specific AT (Chapter 4). Additionally, a single residue likely cannot be used to engineer substrate orthogonality between several different natural and non-natural substrates—something not even achieved with two mutations to the YASH motif.[143] Rather, ATs have evolved to utilize a series of residues within and around the active site, and studying these residues can provide a set of guidelines or rules regarding the identities, roles, and manipulation of these residues needed to reach the potential presented by these modular PKSs.

In progress towards these guidelines, we have studied known and novel mutations affecting substrate selectivity in EryAT6. Using this model system, we have explored the structural consequences of known and predicted mutations *in silico* using molecular dynamics (MD) simulations (Chapter 2) in conjunction with the *in vitro* assays described in this chapter. The *in vitro* results were then used to provide feedback on the *in silico* model. Several residues previously unassociated with substrate selectivity have been identified in the active site and successfully mutated to shift substrate selectivity towards a panel of non-natural malonyl-CoA analogues. These newly-identified residues and motifs, in combination with those previously characterized, provide a necessary foundation for designing ATs for desired substrates. Finally, while studying the AT and the role it plays in non-natural extender unit incorporation, a previously-uncharacterized bottleneck in modular PKSs was identified.

3.2. Results

3.2.1. *Pyrone Formation by Ery6 with a Diketide Substrate*

The AT active site mutagenesis strategy has advanced significantly, as described, but there is still a need to identify more residues and specific mutations that can shift substrate selectivity. In the case of Ery6, the terminal module of the DEBS pathway, its native promiscuity allows for a baseline level of incorporation of several non-natural extender units in addition to its natural mmCoA (**1a**)—notably pgmCoA (**1b**).[64] Computational studies of existing and predicted mutations in EryAT6 were performed as described in Chapter 2, and more than two dozen of these predicted mutations (**Figure 2.12**) were introduced into an Ery6 construct lacking the TE domain (to eliminate any bias by that domain) for *in vitro* assays.

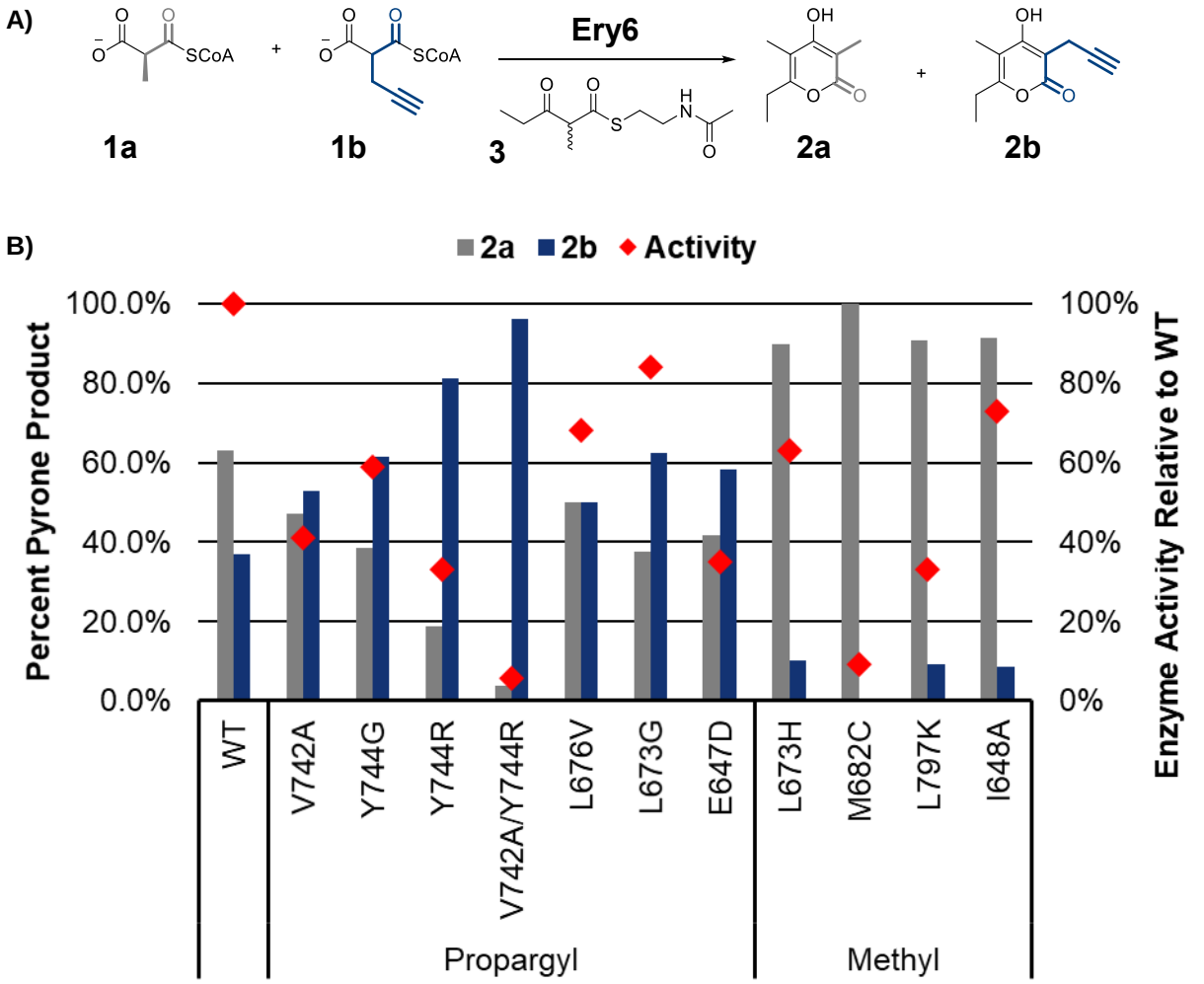


Figure 3.2. Extender Unit Competition Assay for Pyrone Production with Ery6 Variants. **A)** Scheme for competition assay between the native substrate **1a** and the non-natural substrate **1b** for Ery6. The diketide-SNAC extended chain mimic **3** allows for production of UV-visible pryone products in the absence of NADPH. **B)** Ery6 wild-type and mutants were tested for pyrone production and product distribution was determined by HPLC. The mutations shown here shifted substrate selectivity towards or away from the propargylmalonyl-CoA (**1b**). Error was ±10% of indicated values. Enzyme activity relative to wild-type Ery6 is shown with red diamonds.

When Ery6 is provided with an unreduced diketide-SNAC chain mimic (**3**) and a pair of extender units (**1a**-**b**), it produces the corresponding UV-active pyrone products (**2a**-**b**) that can be detected via HPLC (**Figure 3.2A**). The panel of mutants were also compared to some known mutants, including the propargyl-preferring V742A located near the YASH motif, the Y744R mutation

found in the YASH motif itself, the poorly-active (< 10% of wild-type) but highly-selective double mutant, and the methyl-preferring L673H. Consistent with previously-published results, wild-type Ery6 yielded nearly 40% **2b**, with the expected increases or decreases seen for each of the previously validated mutations (**Figure 3.2B**). Y744G has also been previously explored,[61, 143] but despite the smaller size and increased flexibility of glycine, it still provided 20% less **2b** than did Y744R.

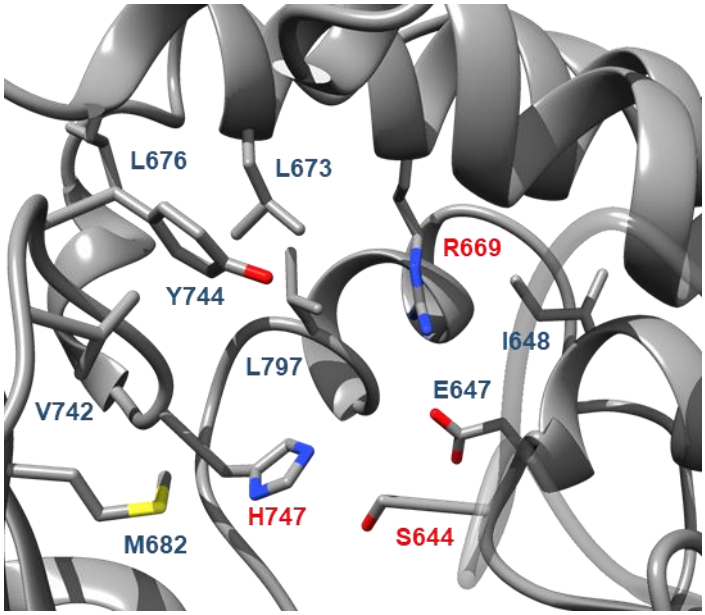


Figure 3.3. EryAT6 Substrate Selectivity Residues Implicated in Pyrone Assay. View of EryAT6 active site from opening towards rear of active site. The catalytic dyad, Ser644 and His747, and the malonate-positioning Arg669 are shown in **red**.

Of the 25 novel mutations tested, a few stood out for their improved selectivity towards either **1a** or **1b** and are shown in **Figure 3.2B**. Three new mutations showed improved activity towards the non-natural **1b**, two of which (L676V and L673G) had total activity levels near that of wild-type (~70-80%). Both Leu676 and Leu673 are located at the rear of the AT active site, potentially providing space for the inflexible propargyl moiety (**Figure 3.3**). Indeed, in MD simulations

(Chapter 2), L673G had a noticeably larger active site as the YASH motif moved into the space previously occupied by Leu673. Leu676 was targeted for much the same reason, as increased flexibility of the rear active site helix in simulations provided larger active sites in EryAT6 (**Appendix A**, **Table A1**). E647D was selected due to its highly conserved role as a hydrogen bond acceptor for the malonate-positioning Arg669 (187 of 188 aligned ATs had a Glu at this position). Based on MD results from Chapter 2, it was reasoned that removing a methylene group might force Arg669 further towards the large subunit of the AT (**Figure 3.3**, right side), providing more space for the propargyl group near the YASH motif located on the small subunit. Gratifyingly, both this mutation and L673G resulted in higher proportion of **2b** product than did the previously-published V742A or Y744G.

In the interest of identifying mutations to increase specificity both towards and away from larger extender units, a subset of mutations was compared to the previously-published L673H that showed ~90% **1a** incorporation (see Chapter 2 for mode of action).[63] Three more methyl-selective mutations were identified with M682C, L797K, and I648A (**Figure 3.2**). Of these three, M682C produced no detectable non-natural **2b** product, while the other two equaled L673H. Unfortunately, the **2a**-selectivity of M682C may be due to a decrease in overall AT activity, as the module produced only about 10% of the total wild-type activity. This loss in activity could be attributed to the importance of the hydrophobic interactions of Met682—and several of the mutations at this position screened *in silico* showed lower structural stability (**Appendix A**, **Table A1**). In contrast to the L673G and L676V mutations, L797K, located in the same region, was designed to decrease the available space around the YASH motif. Finally, I648A was selected to take advantage of a different mechanism: increasing the flexibility of the helix containing the

catalytic Ser644 (**Figure 3.3**). Based on simulations of the wild-type EryAT6, it was proposed that increasing the flexibility of this particular loop would decrease the average distance between the catalytic dyad (see **Figure 2.5**) as it does not make up part of the hinge between the two subunits (like the YASH loop, where increased flexibility showed larger active sites).

3.2.2. *10-Deoxymethynolide Analogue Production by Ery6TE with a Pentaketide Substrate*

In addition to testing modules with a poorly-accepted diketide substrate, Ery6TE mutants were assayed using the previously-described thiophenol-pentaketide substrate from the pikromycin pathway (**6**, **Figure 3.4**).[63, 144] As Ery6 normally accepts a hexaketide substrate from Ery5, pentaketide **6** was expected to be a better extended chain mimic than the diketide **3**; however, to further facilitate the acceptance of **6** and cyclization of the products (**4a-i**, **5a-i**), an N-terminal docking domain from PikAIV of the pikromycin PKS and the EryTE were appended to the Ery6 module used in the pyrone assay. Also, an NADPH recycling system was included to support KR activity, resulting in the 12-membered 10-deoxymethynolide (10-dML) products **4a-i** (**Figure 3.4**) being the possible products. As the product would no longer be detectable via an UV-vis/HPLC assay, high-resolution LC-MS was utilized instead. In contrast to the pyrone assay, the modules for the pentaketide assay were prepared as cell lysates rather than fully-purified modules due to more consistent activities seen in lysate than after subjection to a harsh purification process (data not shown).

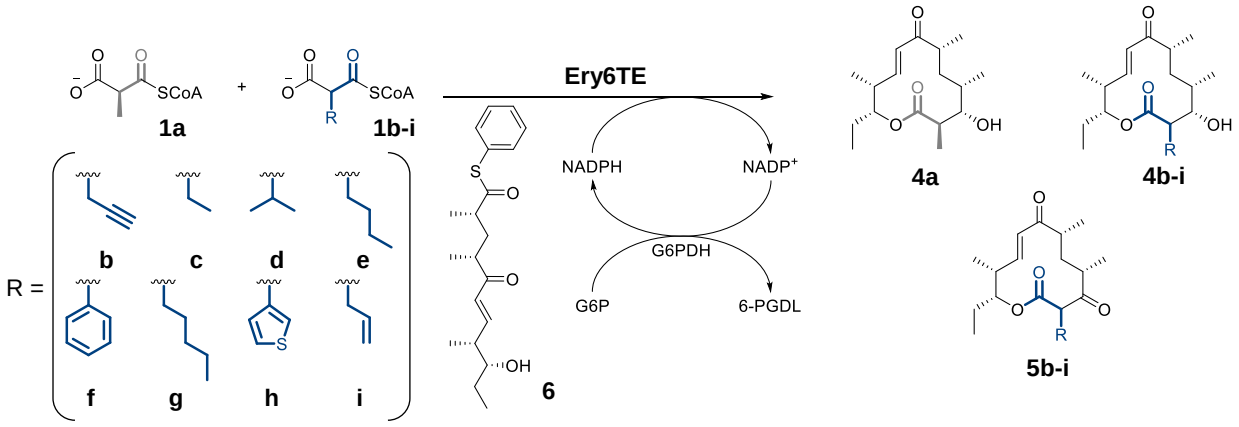


Figure 3.4. Scheme for 10-dML Production by Ery6TE. A competition assay between 2-6 malonyl-CoA substrates (**1a-i**) utilizing a pentaketide chain mimic (**6**) and a NADPH regeneration system yields 10-dML products (**4a-i**) and keto-10-dML products (**5b-i**) resulting from skipping the KR domain. NADPH was recycled by glucose-6-phosphate dehydrogenase (G6PDH) using $NADP^+$ and glucose-6-phosphate (G6P).

In similar fashion to the substrate competition pyrone assay (**Figure 3.2**), multiple extenders units were used in equimolar competition to assess substrate preference for each module; however, due to the improved sensitivity of this assay, lower concentrations of extenders could be used—providing the opportunity to compare up to six extenders in a single assay. Based on the insight into AT mutagenesis gained from MD simulations (Chapter 2) and from the pyrone assay (**Figure 3.2**), a set of 30 Ery6TE mutants including all successful mutations from the pyrone assay and others inspired by them were designed, cloned, and assayed with a mixture of 500 μM each of **1a**, **1b**, **1c**, **1d**, **1e**, and **1h** (**Figure 3.5**). For wild-type Ery6TE, products were derived from the following extender units: methyl **1a** (57.7%), propargyl **1b** (27.1%), ethyl **1c** (2.9%), isopropyl **1d** (11.4%), butyl **1e** (1.0%). Notably, when assay results were analyzed, significant proportions of mass ions consistent with the non-natural products **5b-e** were detected, indicating that the KR domain failed to reduce the C9 ketone (**Appendix B**, **Table B3**). For example, more than 45% of all products were not reduced by the KR, with 100% of the butyl product coming as unreduced **5e**.

Gratifyingly however, several of the mutants, previously-published and novel, showed altered substrate selectivity from the wild-type enzyme. Disappointingly, none of the tested enzymes accepted **1h**.

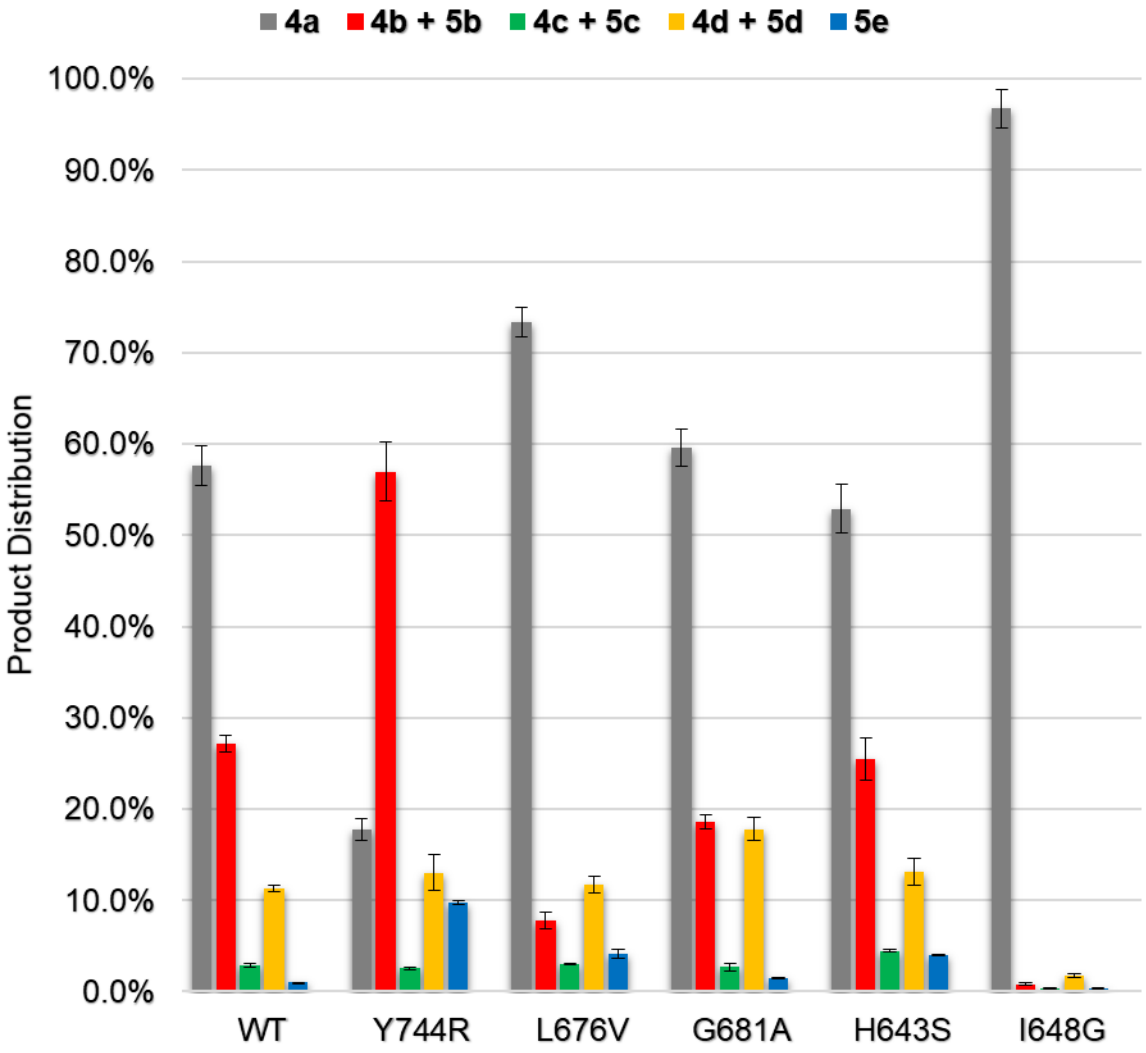


Figure 3.5. Assaying Ery6TE Mutants with Non-Natural Extender Unit Panel. Module lysates were incubated with 500 μM each of **1a**, **1b**, **1c**, **1d**, **1e**, and **1h** simultaneously. Products were analyzed by HR-LCMS and representative mutants are shown. Full results including relative activities are shown in **Figure B2** and **Table B4**.

The two previously-published mutations tested, V742A and Y744R, are located near and in the conserved YASH motif of mmCoA-specific ATs (**Figure 3.6**), respectively, and both supported significant increases in propargyl (**1b**, C_3) incorporation. Of these two mutations, Y744R supported the largest increase in butyl (**1e**, C_4) with a 10-fold change compared to wild-type, while V742A displayed the larger change in ethyl (**1c**, C_2) incorporation with a 1.5-fold improvement. Other mutations provided increased activities with various extenders, highlighted in **Figure 3.6**. Conspicuously, most of the mutations responsible for selectivity shifts towards larger extenders were located near the rear of the active site proximal to the YASH motif. This includes the L673G (2.4-fold higher butyl) and L676V (4.1-fold higher butyl) mutations identified first with MD and then again with the pyrone assay. V751A, located near Leu673 and Leu676, also showed a 2.5-fold improvement utilizing the butyl **1e**. Of the remaining mutations, G681A (~1.5-fold increases in isopropyl **1d** and butyl **1e**) is in the small subunit, as part of a β sheet directly linked to the helix containing two of the other mutations. Likewise, D613E, while on the large subunit opposite from the YASH motif, directly interacts via salt bridge with Arg674 on the helix in the rear of the active site (**Figure 3.6**) and results in a nearly 2-fold increase in butyl **1e** incorporation.

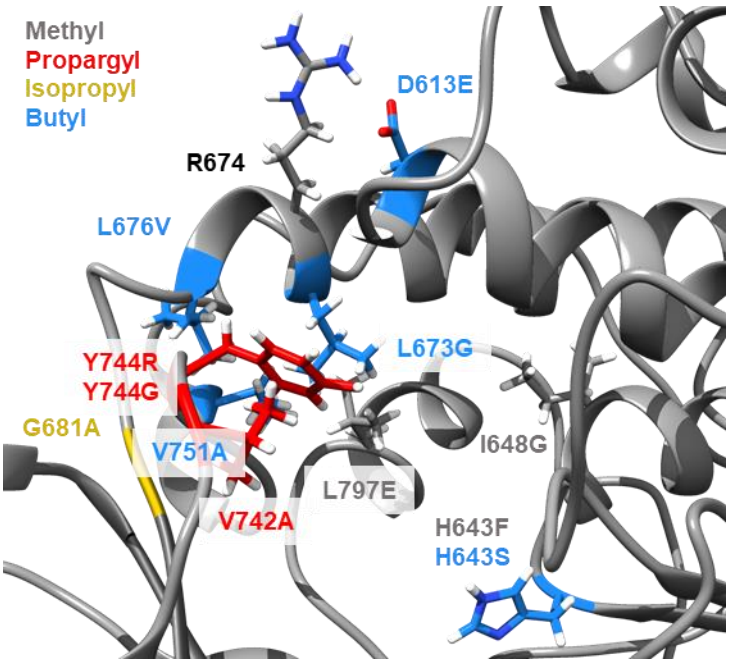


Figure 3.6. Substrate-Selectivity Mutations in EryAT6 with Most Improved Substrate Selectivity. Mutations highlighted in the non-natural extender unit panel screen are shown and highlighted in the color corresponding to the single-best improvement in non-natural production (some mutants resulted in more than one improved selectivity). Most residues are located around but not in the YASH motif.

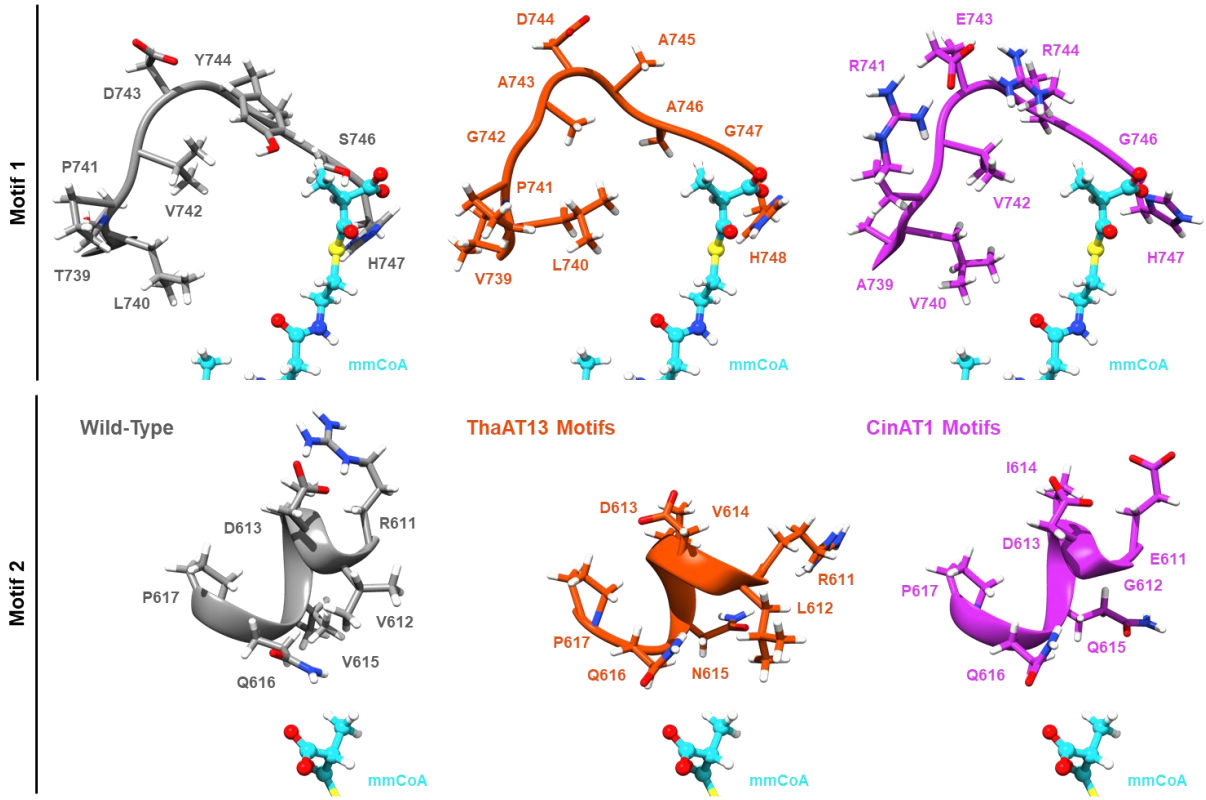
The outlier in terms of location is the improved ethyl (1.6-fold increase) and butyl (4.1-fold increase) mutant H643S. His643 is located in the conserved GHSxG motif next to the catalytic serine and has been previously proposed to either hydrogen bond with the conserved Asn709 in the small subunit or stabilize the malonyl-enzyme intermediate via hydrogen bonding with water.[131] Inspired in part by these proposals and in part by MD simulations (Chapter 2, Section 2.2.3), both H643S and H643F were tested (**Figure 3.5**)—one smaller and retaining the ability to hydrogen bond and the other bulkier and lacking any hydrogen bonding ability. Interestingly, the phenylalanine mutant demonstrated greater than 150% of wild-type activity, though the added steric bulk appears to have acted as a screen against larger extender units (95% methyl product). Surprisingly, retaining the hydrogen bonding ability through serine resulted in a lower overall activity—albeit with the improved preferences for **1c** and **1e**.

Alongside H643F, two other methyl-specific mutations were identified: I648G and L797E. Both mutants are located at positions identified during the pyrone assay (I648A and L797K, respectively). I648G resulted in 96.7% of products being the natural **4a** while L797K resulted in 98.4% **4a** (**Figure 3.5**). In all three methyl-specific mutants, much higher activity was observed relative to wild-type Ery6TE, but this is most likely due to expected low wild-type catalytic efficiency with the larger extenders that are not accepted by these mutants or unproductive hydrolysis of extenders by wild-type Ery6TE. As in the case of I648A, I648G increases flexibility of the GHSxG loop and may result in a smaller active site, while as with L797K, L797E introduces a charged residue directly behind the hydrophobic Leu673, likely contributing to less space in the rear of the active site.

As **Figures 3.2**, **3.3**, **3.5**, and **3.6** demonstrate, the AT's substrate selectivity is contributed to by many more residues than have been previously identified—with seven novel residues for larger extenders and four more for methyl selectivity described here. As shown with MD (Chapter 2), even single mutations can affect changes far from the residue of interest, but it was often how the residues surrounding each mutation reacted (e.g., increased flexibility of the small subunit or movement of the rear active site helix) that determined whether selectivity shifted. Because many of these mutations reside outside the traditionally-defined substrate motifs, there is the possibility of combining mutations at several sites within the active site to realize more drastic selectivity shifts. Just as ATs have evolved to have different substrate preferences in nature, an artificially-engineered AT might also require multiple changes throughout the enzyme to achieve the desired substrate orthogonality. Towards this goal, multi-amino acid swaps can be performed based not on conservation alignments but rather on structural motifs.

3.2.3. *Use of Active Site Motif Swaps to Alter Substrate Selectivity*

To complement the single-site amino acid mutations and to explore how different sets of selectivity-conferring residues might work together in a mmCoA-specific AT, sequences from natural ATs with unusual malonyl-CoA specificities were inspected. Due to the clustered locations of the mutations described above and the large conformational changes seen in MD (Chapter 2), structural motifs within these ATs were predicted to discriminate between substrates more effectively than any single mutation. It was proposed that two of these motifs together could confer near-total flips in substrate selectivity. The small subunit loop spanning from Thr739 to Ser746 in Ery6TE was affected the most by the mutations at residues Gly681, Val751, Leu673, and Leu676 *in silico*, and the loop includes the previously-implicated YASH motif and Val742. This structural motif was defined as Motif 1. The helix directly above the catalytic serine in the large subunit spanning from Arg611 to Pro617 and containing Asp613 was defined as Motif 2 (**Figure 3.7**). Motif 1 was selected due to its close interactions with any malonyl-CoA C2 moieties, and Motif 2 was chosen to help position the substrates and due to its interactions with Motif 1 in simulations of the enzyme-malonyl intermediate (Chapter 2, Section 2.2.4). A third motif comprising of the helix in the rear of the active site from Ala672 to Leu676 containing several of the mutations from the previous section was considered but ultimately not tested due to the effectiveness of Motifs 1 and 2 alone.



EryAT6	TLPVD-YASH	RVDVVQP
CinAT1	ALRVE-RAGH	EGDIQQP
ThaAT13	VLPGADAAGH	RLDVNQP

Figure 3.7. Comparison of Motif Swaps in EryAT6. Models highlighting the proposed substrate selectivity-conferring motifs on the small subunit (Motif 1) and large subunit (Motif 2) from Ery6 wild-type (**grey**), ThaAT13 (**orange**), and CinAT1 (**magenta**). Numbering is from the Ery6TE hybrid constructs. MmCoA (**1b**, **cyan**) is shown docked into the AT active site. Sequence alignment shows Motif 1 and Motif 2 for the three ATs. Conserved residues are bolded.

In the interest of verifying if these two motifs (14 residues total) could flip substrate selectivity, the equivalent stretches of residues from two ATs were introduced into EryAT6: the butylmalonyl-CoA-selective AT13 from the thailandin PKS (ThaAT13) and the hexylmalonyl-CoA-selective AT1 from the cinnabaramide PKS (CinAT1).[145-146] To the best of our knowledge, CinAT1 is one

of only a couple of natural ATs containing the arginine at the first position of the YASH motif that has proven to be so effective in EryAT6 with propargyl. Despite assertions that the arginine prevents efficient use of extenders much larger than propargyl,[143] CinAT1 contains two arginines in Motif 1 (**Figure 3.7**) and takes larger extenders naturally.[145] In contrast, CinAT1 Motif 2 contains two negatively-charged residues, Glu611 and Asp613 that likely help off-set the positive charge contributed by the two arginines. ThaAT13 shares nearly all of its Motif 2 with EryAT6, but its Motif 1 provides significant increases in flexibility, especially from the extra Gly742 residue added in the middle of the loop.

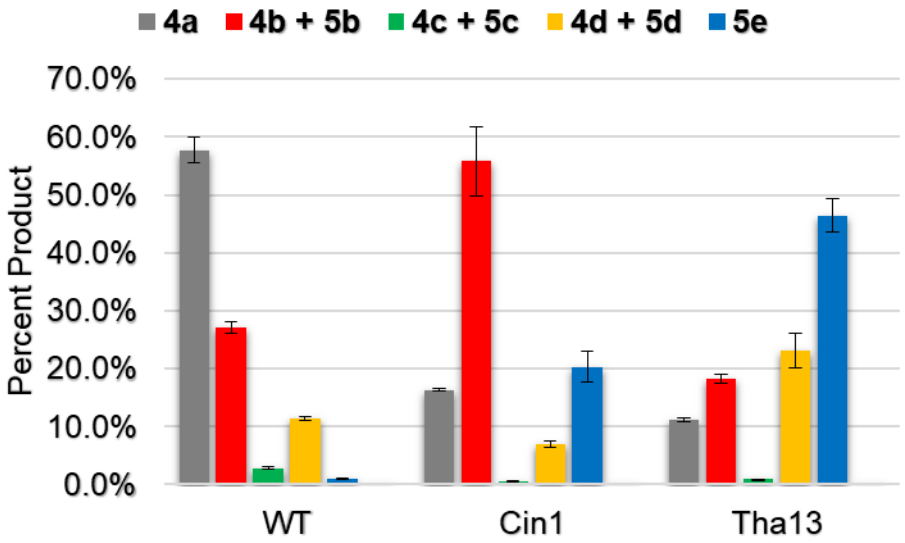


Figure 3.8. Assaying Ery6TE Motif Chimeras with Non-Natural Extender Unit Panel. Module lysates were incubated with 500 μM each of **1a**, **1b**, **1c**, **1d**, **1e**, and **1h** simultaneously. Products were analyzed by HR-LCMS, and the product distributions are shown. Full results including relative activities are shown in **Table B3**.

When the two motif-swapped Ery6TE constructs were assayed in the same fashion as the individual mutations with six equimolar extenders, the results were gratifying. In comparison to the wild-type enzyme, the Tha13 motifs' methyl (**4a**), ethyl (**4c**/**5c**), and propargyl (**4b**/**5b**)

products decreased to 11.2%, 0.9%, and 18.3%, respectively, while isopropyl (**4d**/**5d**) and butyl (**4e**/**5e**) products increased to 23.2% and 46.5%, respectively. For Cin1, the methyl (16.3%), ethyl (0.6%), and isopropyl (6.9%) product ratios decreased, and the propargyl (55.8%) and butyl (20.3%) ratios increased (**Figure 3.8**). Again, the thiophene products (**4h**/**5h**) were not detected. Activities for both chimeras were robust at over 30% of wild-type total activity. While the dominance by butyl derivatives was expected and desired for the Tha13 chimera, it was surprising to observe Cin1's propargyl percentage nearly identical to Y744R (55.8% to 57.0%), though with double the butyl percentage (20.3% to 9.7%).

To better mimic expected *in vivo* conditions in an engineered strain,[147] the two motif chimeras, wild-type Ery6TE, and several of the mutant enzymes were assayed at higher concentrations of extenders—with only one non-natural extender unit in each assay along with the native substrate for Ery6TE, mmCoA (**1a**). Two mutations at Tyr744, Y744G and Y744R, were tested along with the methyl-specific I648G mutant. Each mutant was assayed with 1.5 mM **1a** and 1.5 mM of either **1i** (allyl) **1b** (propargyl), **1e** (butyl), or **1g** (pentyl) (**Figure 3.9**). Each mutant was also assayed with a 500 μM mixture of mmCoA (**1a**) and phenylmalonyl-CoA (**1f**), but no phenyl product was detected (data not shown). Wild-type Ery6TE products contained allyl (**4i**/**5i**) at less than 3%, propargyl (**4b**/**5b**) at under 20%, and both butyl (**5e**) and pentyl (**5g**) at under 1% in competition with mmCoA. Like with previously-assayed non-natural extenders, the KR did not reduce most of the allyl products or any of the pentyl products (**Table B3**). As expected, both single mutations flipped the selectivity of the AT to propargyl, but neither mutation alone could shift the substrate preference from methyl to any of the other three substrates. In contrast, the Tha13 and Cin1 motifs both enabled Ery6TE to prefer propargyl, butyl, and pentyl to methyl, with butyl and pentyl nearing

complete flips at 80% or greater each. For Tha13, pentyl (**5g**) made up 93.5% of products in direct competition with methyl (**4a**), a greater than 300-fold increase in substrate selectivity. Interestingly, the product profiles were similar for both motif swaps when competing only with methyl, unlike when competing with multiple non-natural products (**Figure 3.8**).

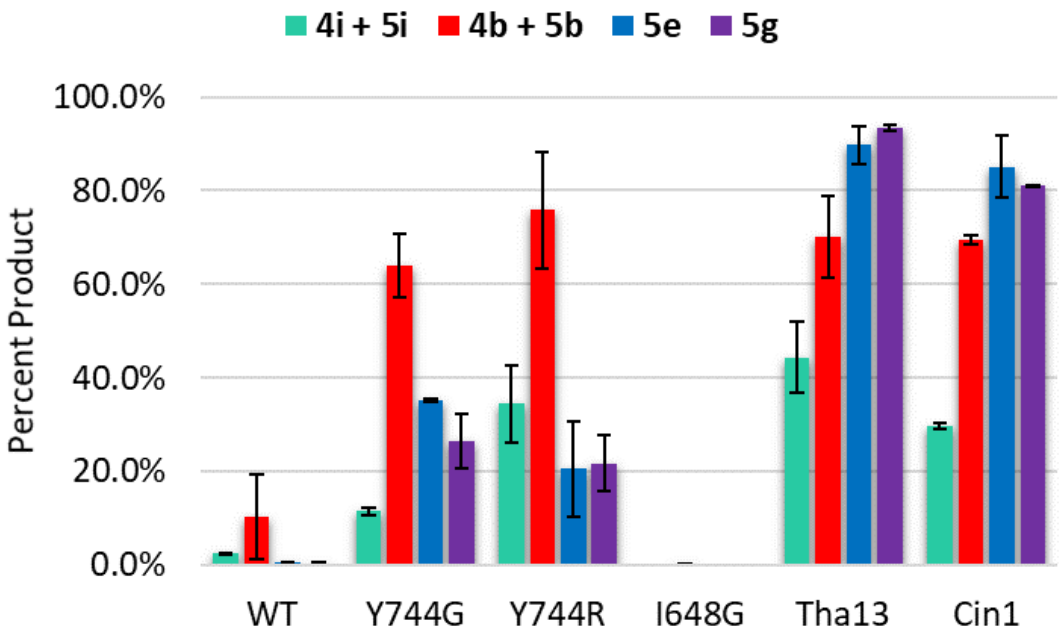


Figure 3.9. Extender Unit Competition for Ery6TE Individual Mutations and Motif Swaps. Module lysates were incubated with 1.5 mM **1a** and 1.5 mM each of **1i**, **1b**, **1e**, or **1g** individually. Products were analyzed by HR-LCMS, and the non-natural product distributions are shown.

3.2.4. *Varying the Levels of Natural and Non-Natural Extender Units for Selectivity Shifts*

Finally, while testing the module lysates with varying concentrations of extenders during assay development, it was observed that the ratio of natural to non-natural extender incorporation changed significantly even when extenders were at equimolar concentrations. While this effect is likely due to low non-natural k_{cat} values and a finite amount of substrate available in an *in vitro* reaction, it is still yet another avenue to target for creating orthogonal sets of PKS modules that

utilize different extender units. To better quantify this concentration dependence, wild-type Ery6TE was used in the pentaketide assay as before but with concentrations of **1a** (methyl) and **1b** (propargyl) ranging from 25 μM to 1500 μM (**Figure 3.10**). As shown, at low concentrations, the non-natural products **4b** and **5b** made up about 60% of the total product, but at the highest concentration tested, the non-natural share of products had fallen to one-third of the value at low concentrations. In addition to using this feature for metabolic engineering, this result helps explain the significant difference between results from the pyrone assay (~40% **2b**) and the pentaketide assay (~10-20% **4b**+**5b**), as the diketide substrate **3** is such a poor substrate for Ery6 that any k_{cat} differences between extenders likely becomes negligible. As the AT step is not normally the rate-limiting step for a PKS,[148-150] the Y744R mutant was assayed in the same fashion as the wild-type module. Little if any decrease in product ratio was seen (~75-90% **4b**+**5b**) over the same concentration range for Y744R (**Figure 3.10**), confirming that an AT with an improved activity towards the non-natural propargyl substrate does not suffer from this decrease at higher concentrations.

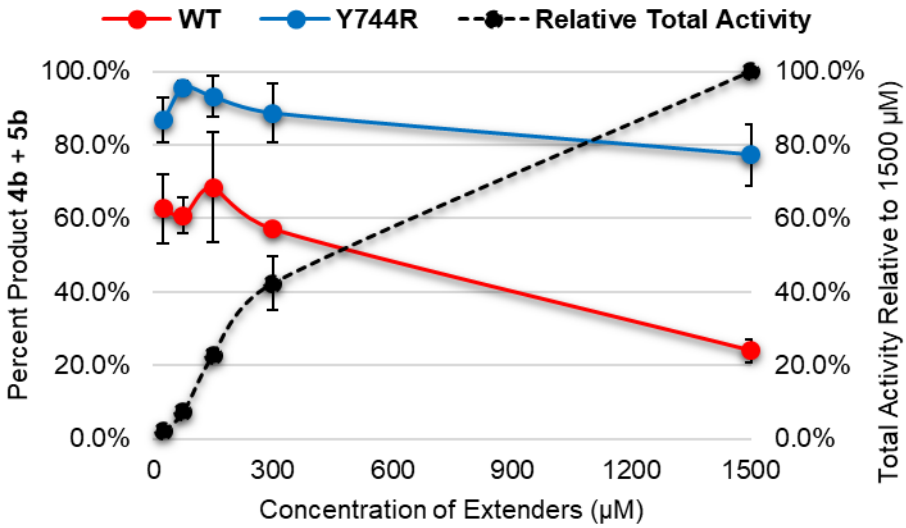


Figure 3.10. Substrate Preference Dependence on Substrate Concentrations. Wild-type Ery6TE lysate was assayed with competing equimolar amounts of **1a** and **1b**, and 10-dML products were detected using HR-LCMS. As extender concentration and total wild-type product (black dashed line) increased, the ratio of **4b**+**5b** to **4a** decreased significantly (red line). Ery6TE Y744R lysate was assayed in the same way and the product ratio is shown to decrease only slightly over the same concentration range (blue line).

3.3. Discussion

Engineering PKSs to accept non-native extender units is a necessary step towards the eventual goal of diversifying any of the many bioactive polyketide core structures. To date, there have been several advances towards introduction of these non-natural moieties using native promiscuity,[64, 151] *trans*-ATs,[60, 77, 147] full AT swaps,[56] and particularly in the last 5 years, AT active site mutagenesis.[61-63, 143, 152] Despite these developments, our lack of insight into the molecular and structural basis for substrate selectivity has left researchers targeting the same small handful of residues without exploring the remainder of the ~430 amino acids in the AT. While engineering selectivity between substrates often differing in only a single carbon unit is a unique protein engineering challenge, we have demonstrated that it can be tackled through a better understanding

of the AT's internal dynamics and an openness to explore novel residues either individually or as part of structural motifs.

After utilizing MD simulations in Chapter 2 to identify key residues and movements within the AT that lead to changes in active site accessibility (Chapter 2), *in vitro* assays were used to describe specific mutations that shifted substrate selectivity either towards or away from the native methyl extender. With fewer than 50 individual mutations screened, 13 (at 10 residues) were found to result in a 1.5-10-fold increase in selectivity towards a provided extender. These novel selectivity-affecting residues were identified throughout EryAT6's active site, particularly in the areas proximal to the YASH loop (Motif 1) such as the mmCoA-conserved VDVVQP motif (Motif 2) or a poorly conserved but highly relevant third motif corresponding to the residues Ala672 to Leu676 in EryAT6. Notably, the mutations did not always simply increase overall substrate promiscuity but rather shifted selectivity towards specific extenders (e.g., L676V increased butyl incorporation at the expense of propargyl).

Gratifyingly, the same strategy of targeting individual residues outside of the canonical YASH motif also worked well with larger structural motif swaps. The use of multiple motifs that interact together rather than clashing with other domains avoids the issues poised by full domain swaps, leaving more than 95% of the AT unchanged. As such, both chimeras presented here retained relatively high overall activities while shifting substrate selectivity nearly completely with larger substrates (e.g., Tha13 realized a >300-fold change in pentyl (**1g**) incorporation compared to wild-type). This motif swapping strategy needs to be explored further, with the above-mentioned third

motif and with motifs from other ATs recognizing unusual malonyl-CoAs. However, the proof-of-principle for this concept presented here shows promise for future AT engineering efforts.

Finally, while successful AT engineering enables the use of these more exotic substrates, it also provides new research directions to consider. For example, when the mass spectra of the full-length products were examined, there were the expected masses for non-natural incorporation (**4b**-**i**), but there were also the masses of the unreduced 10-dML products (**5b**-**i**). This lack of promiscuity by the KR towards different groups at the C2 position, never previously reported, played a major role, especially with the larger substrates, as the butyl and pentyl products were not reduced at all. This bottleneck highlights the long road remaining to achieving polyketide derivatization at-will, and it is discussed further in Chapter 4. An additional challenge that arises as increasingly promiscuous ATs are designed is the evolution of ATs capable of recognizing ever more structurally-diverse extenders. Neither of the inflexible and bulky extenders, phenyl (**1f**) or thiophene (**1h**) were accepted by any EryAT6 variant. Cumulatively however, it has been shown here that AT substrate selectivity can be modified through the use of individual and grouped mutations throughout the active site and through balancing of extender unit levels, and it is anticipated that the number of implicated residues and incorporated substrates will only grow as more are explored.

3.4. Materials and Methods

3.4.1. *General*

Materials and reagents were purchased from Sigma Aldrich (St. Louis, MO) unless otherwise noted. Isopropyl β-D-thioglactoside (IPTG) was purchased from Calbiochem (Gibbstown, NJ). The TP-pentaketide (**6**) was kindly provided by Dr. David Sherman at the University of Michigan.

Construction of the Ery6-pET28a and Ery6TE-pET28 plasmids were described previously.[63] All sequences are listed in **Appendix B**, **Table B1**. Oligos were purchased from Integrated DNA Technologies (Coralville, IA). All *holo* proteins were expressed in *E. coli* K207-3 cells containing Sfp, and all *apo* proteins were expressed in *E. coli* BL21 (DE3) cells.[153] All oligos are listed in **Appendix B**, **Table B2**.

3.4.2. *Mutagenesis of Modules*

Phusion polymerase (New England Biolabs, NEB, Ipswich, MA) was used for round-the-horn mutagenesis of all full module constructs.[154] Motif swaps of Ery6TE were constructed using Gibson assembly (NEB) with synthesized gene fragments containing the motifs of interest.[155] Correct mutants were transformed into K207-3 competent cells and plated onto LB agar plates (50 μg/mL kanamycin).

3.4.3. *Expression and Purification of Ery6 Modules*

Cells were grown in 300 mL of LB media with 50 μg/mL kanamycin at 37 °C and 250 rpm until OD_{600} reached 0.6, and then the culture was induced with 1 mM IPTG and allowed to express at 18 °C for 20 h. Proteins were purified on a nickel column with a Bio-Rad BioLogic LP FPLC (Wash buffer: 50 mM sodium phosphate, pH 7.2, 300 mM NaCl, 20 mM imidazole, and EDTA-free protease inhibitor cocktail (Roche, Basel, CH); Elution buffer: 50 mM sodium phosphate, pH 7.2, 300 mM NaCl, 200 mM imidazole, and EDTA-free protease inhibitor cocktail), concentrated with a 10 kDa MWCO filter (EMD Millipore, Burlington, MA), and stored as single-use aliquots in module storage buffer (100 mM sodium phosphate, pH 7.4, 1 mM EDTA, 1 mM tris(2-

carboxyethyl)phosphine (TCEP), 20% v/v glycerol, 0.1 μL Benzonase (NEB), and EDTA-free protease inhibitor cocktail) at -80 °C.

3.4.4. *MatB Reactions and Acyl-CoA Preparation*

MatB wild-type and MatB T207G/M306I were purified and 8mM malonyl-CoA (**1a-i**) stocks were set up as previously described.[156] A representative HPLC trace is shown in **Appendix B**, **Figure B1**.

3.4.5. *Pyrone Production Competition Assay for Ery6*

Reactions were set up as follows in 50 μL: 5 mM diketide-SNAC (**3**), 1 mM each extender unit (**1a-i**), 100 mM sodium phosphate, pH 7, and 4 μM Ery6 purified protein. Reactions were run at room temperature for 20 h and then quenched with 50 μL -20 °C methanol. Quenched reactions were centrifuged for 30 min at 13,300 rpm to pellet protein. Supernatant was analyzed on a Varian ProStar HPLC using a 290 nm light and the following gradient: 1 mL/min 0.1% TFA in water (Buffer A) and 0.1% TFA in acetonitrile (Buffer B) at 0% Buffer B (0-2 min), 0-50% Buffer B (2-25 min), 100% Buffer B (25-30 min), and 0% Buffer B (30-35 min). A module with a catalytically-inactive AT (S644A) was used as a negative control.

3.4.6. *Lysate Preparation for Ery6TE Modules*

Modules were expressed overnight at 18 °C in 300 mL cultures of LB media with 50 μg/mL kanamycin. Protein production was induced with 1 mM IPTG at OD_{600} of 0.6. After overnight expression, the culture was centrifuged at 4,700 rpm for 20 min, and the supernatant was discarded. The cells were resuspended in 1 mL module storage buffer (100 mM sodium phosphate, pH 7.4, 1

mM EDTA, 1 mM TCEP, 20% v/v glycerol, 0.1 μL Benzonase, and EDTA-free protease inhibitor cocktail) and sonicated using 51% amplitude, 10 s on, 20 s off for 10 min. After sonication, the lysed cells were centrifuged at 18,000 rpm for 1 h. The lysates were aliquoted and stored at -80 °C.

3.4.7. *Pentaketide Assay*

The pentaketide assay was set up with a total volume of 80 μL in 100 mM sodium phosphate, pH 7.0, and 20 mM $MgCl_2$. The reaction conditions included 1 mM TP-pentaketide (**6**), an NADPH regeneration system (5 mM glucose-6-phosphate, 500 μM $NADP^+$, and 0.008 U/mL glucose-6-phosphate dehydrogenase), and 34.4 μL lysate containing the module of interest. Module concentrations were calculated by Bradford assay and SDS-PAGE gel analysis. Malonyl-CoA extender units (**1a**-**i**) were added to the reaction as a set of two (1.5 mM each) or as a set of six (500 μM each). Reactions proceeded at room temperature for 16 h and were quenched with an equal volume of -20 °C methanol. After quenching, all reactions were centrifuged at 13,300 rpm three times for 3 h total, and the supernatant was filtered through a nylon 0.2 μm filter. Analysis was carried out on a high-resolution mass spectrometer – a ThermoFisher Scientific Exactive Plus MS, a benchtop full-scan Orbitrap™ mass spectrometer – using Heated Electrospray Ionization (HESI). The sample was analyzed via LC-MS injection into the mass spectrometer at a flow rate of 225 μL/min. The mobile phase B was acetonitrile with 0.1% formic acid and mobile phase A was water with 0.1% formic acid (see **Appendix B**, **Table B5** for gradient and scan parameters). The mass spectrometer was operated in positive ion mode. The LC column was a Thermo Hypersil Gold 50 x 2.1 mm, 1.9 μm particle size. This assay produces 10-dML products that can be seen as their $[M+H]^+$, $[M+H-H_2O]^+$, and $[M+Na]^+$ ions and keto-10-dML products that can be seen as their

$[M+H]^+$ and $[M+Na]^+$ ions. Extracted ions for each listed ion were summed for comparison purposes. For retention times, calculated masses, observed masses, representative extracted ion counts, and representative chromatograms, see **Appendix B**, **Tables B6**, **B7**, **B8**, and **B9** and **Figure B3**. A module with a catalytically-inactive AT (S644A) was used as a negative control.

3.4.8. *Construction of Homology Models*

Wild-type homology models for EryAT6 with ThaAT13 and CinAT1 motifs were created using the I-TASSER online server.[124-126] Homology models were then treated with the online FG-MD server to show interactions between nearby residues.[157] Molecular graphics and PDB snapshots were performed with VMD 1.9.2 and UCSF Chimera 1.10.1.[134-136] Images were rendered with POV-Ray.[138]

CHAPTER 4

Comparing and Controlling Substrate Selectivities of Acyltransferase Domains in the Pikromycin Polyketide Synthase

4.1. Introduction

Type I polyketide synthases (PKSs) are responsible for the biosynthesis of some of the most clinically important bioactive compounds in Nature, including the blockbuster drugs erythromycin A (antibiotic), rapamycin (immunosuppressant/anti-cancer), and avermectin (anthelminthic).[1] These PKSs are giant assembly line pathways that can be broken down into individual modules (**Figure 4.1**). Each module is responsible for incorporation of a single extender unit, often a coenzyme A (CoA)-linked malonate derivative. Within each module, there are a series of conserved enzymatic domains responsible for specific catalytic activities such as reduction (KR), dehydration (DH), or condensation (KS). The acyltransferase (AT) acts as the "gatekeeper" domain due to its innate ability to select a specific extender unit for priming of its cognate acyl carrier protein (ACP).

Despite the structural diversity of polyketides, the AT domains responsible for selecting the extender units for each module almost universally select from only three substrates: malonyl-CoA, methylmalonyl-CoA, and to a lesser degree, ethylmalonyl-CoA.[19] Thus, except in a few rare cases, the selected substrates account for little to no chemical diversity.[158-162] Instead, much of the diversity seen in polyketides comes from varying oxidations, cyclization patterns, or post-PKS modifications. This is nicely highlighted by the four final products that result from the pikromycin (Pik) PKS (**Figure 4.1**). However, with the development of various approaches to chemo-enzymatically produce non-natural malonyl-CoA analogues, it may be possible to increase the

structural diversity of polyketides by orders of magnitude if strategies to install them into polyketides can be developed by engineering PKSs.[76, 152, 156]

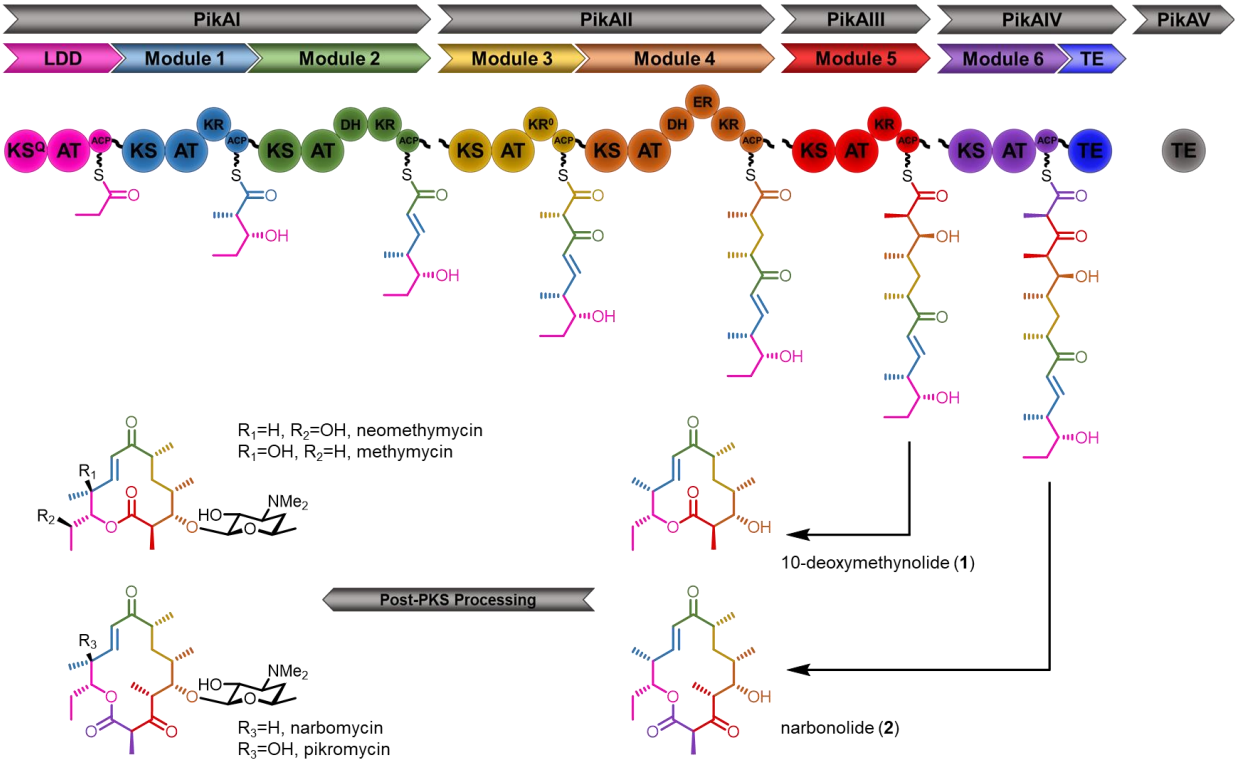


Figure 4.1. The pikromycin polyketide synthase and its products. ACP = acyl carrier protein; AT = acyltransferase; DH = dehydratase; ER = enoylreductase; KR = ketoreductase; KS = ketosynthase; KS^Q = ketosynthase-like decarboxylase; TE = thioesterase.

Traditionally, PKS engineering has been focused on swapping modules or domains to alter the final product structure, but in both cases, there are three critical limitations: (1) most PKS modules incorporate natural extenders that lack useful chemical handles, (2) non-native protein-protein interactions often result in low-efficiency chimeras,[19, 163] and (3) to achieve the site-selective installation of a given non-natural extender unit into a polyketide, the specificity of the domain/module-swapped chimera must be orthogonal to that of the native, intact extension modules. In order to produce non-natural extenders, our group and others have utilized and engineered malonyl-CoA synthetases or similar enzymes to create a panel of PKS substrates that

include a variety of useful chemical moieties.[59-60, 76, 151-152, 156] Several labs have addressed the second issue through introduction of active site mutations for the AT of interest.[61-63, 143, 152] For example, much of the early work involved swapping out the YASH motif, a conserved motif found in methylmalonyl-specific ATs, with motifs from other natural ATs (e.g., HAFH from malonyl-CoA specific ATs); however, these changes alone could not completely invert AT specificity and therefore do not provide the requisite orthogonality for site-selective modification of the polyketide structure.[121]

While site-directed mutagenesis had proven less effective than desired for switching selectivities between natural substrates,[121] our group and others have demonstrated that inherent extender unit promiscuity of some ATs provides a platform for creating new substrate specificities via site-directed mutagenesis. For example, Williams and co-workers reported that the methylmalonyl-CoA-utilizing EryAT6 and corresponding extension module (Ery6) of the 6-deoxyerythronolide B synthase (DEBS) from erythromycin A biosynthesis has significant promiscuity towards larger non-natural extender units.[64] These non-natural substrates were recently utilized by an engineered Ery6 module wherein the mutation of the YASH motif to RASH led to a switch from 92% methylmalonyl-CoA incorporation by the wild-type enzyme to 88% propargylmalonyl-CoA (non-natural) incorporation by the engineered single mutant.[63] Vögeli and coworkers recently demonstrated the ability to manipulate the substrate preference of another AT from DEBS (EryAT2) towards longer alkyl chains via a VASH motif (found in some natural ethylmalonyl-CoA-specific ATs), albeit to a lesser extent.[152]

These shifts in substrate selectivity are impressive and highly desired yet ostensibly rely on the inherent promiscuity of the chosen ATs as a toehold for redesigning their substrate specificity. Additionally, most AT engineering is accomplished with extension modules that are terminal to the assembly line and involves installation of one non-natural extender unit into the final product structure. In this chapter, the ability of active site mutagenesis to manipulate the extender unit specificity of ATs that do not display inherent promiscuity was explored. To this end, the Pik PKS, responsible for the biosynthesis of two core macrolactones, a 12-membered 10-deoxymethynolide (10-dML, **1**) and a 14-membered narbonolide (**2**) in *Streptomyces venezuelae* ATCC 15439, was carefully selected as a target for mutagenesis.[164] It was proposed that the extension modules of this pathway would be less promiscuous towards larger extenders than the prototypical DEBS modules due to its evolution in a host that also produces ethylmalonyl-CoA and to documented hydrolytic proof-reading by the AT.[165-166] Using the final two modules from this PKS, the native promiscuity of each module was first compared with a panel of natural and non-natural extender units and a series of domain swaps. Next, the substrate selectivity of each module was successfully engineered towards non-natural extender units via site-directed mutagenesis of active site residues. Finally, a new bottleneck in PKS engineering was highlighted. To our knowledge, this is the first example of successful substrate selectivity inversion in an AT that does not belong to the prototypical DEBS assembly line. Moreover, and to the best of our knowledge, it represents the first known example of two non-natural extender units being incorporated into a single full-length polyketide product.

4.2. Results

4.2.1. *Characterization of the PikAIII/PikAIV System*

To date, the majority of AT substrate selectivity engineering work has been done in Ery6, a terminal module chosen at least in part because the fully-extended non-natural chains do not need to be passed through other modules.[61-64, 116] Unfortunately, in most PKSs, two modules from one system cannot be compared equally because any discrepancy in promiscuity could be due to the gatekeeping ability of the upstream ATs or that of downstream domains or some combination thereof; however, the Pik PKS provides a unique opportunity to do so due to having not one, but two terminal modules that can give different product profiles by virtue of the cyclization specificity and lack of KR in the last module.

Of the six extension modules in the Pik PKS, only the final two are monomodular enzymes: PikAIII and PikAIV (**Figure 4.1**). These two enzymes are certainly evolutionarily related, with 74% amino acid identity over two-thirds of their sequences (PikAIV lacks a KR domain), and both AT domains (88% identical) utilize methylmalonyl-CoA. In contrast, the two ATs are quite dissimilar to the highly promiscuous EryAT2 (48% PikAT5, 47% PikAT6) and EryAT6 (46% PikAT5, 45% PikAT6) (**Appendix C, Figure C1**).[64, 152] To better understand substrate promiscuity outside of the DEBS ATs and to determine a baseline level of promiscuity for the Pik ATs, the modules were tested as lysates *in vitro* with a synthetic pentaketide chain mimic **10** and equimolar amounts of a natural (**8**) and a non-natural extender unit (e.g., **9a**, **Figure 4.2**). The reactions were run as lysates as a harsh purification process has been shown in our lab and others to negatively affect module stability and reaction consistency.[72, 167-168] In this assay, the chain mimic **10** would be loaded by PikKS5 for extension by one or both modules. As seen in the wild-type system

(**Figure 4.2**, **Table C5**), after extension by PikAIII in the presence of two competing extender units, the hexaketide can be passed on for extension and cyclization as one of four possible narbonolide products by PikAIV or shuttled directly to the TE for cyclization as one of two 10-dML products, thus 'skipping' PikAIV (PikAIV does not accept the pentaketide substrate, **Table 4.1**, R5). In this way, the 10-dML products must be derived from extension by PikAIII, while the 14-membered narbonolide products must arise by sequential extensions by PikAIII and PikAIV. Notably, a third 12-membered product (**4a-d**) is also possible in the presence of non-natural extender units due to poor recognition of the corresponding elongated intermediate by the PikAIII ketoreductase (KR). These features collectively enable a precise assessment of the ability of each module to process non-natural extender units on the basis of the corresponding product distribution, which can then be easily detected and quantified by high-resolution LC-MS.[72, 144]

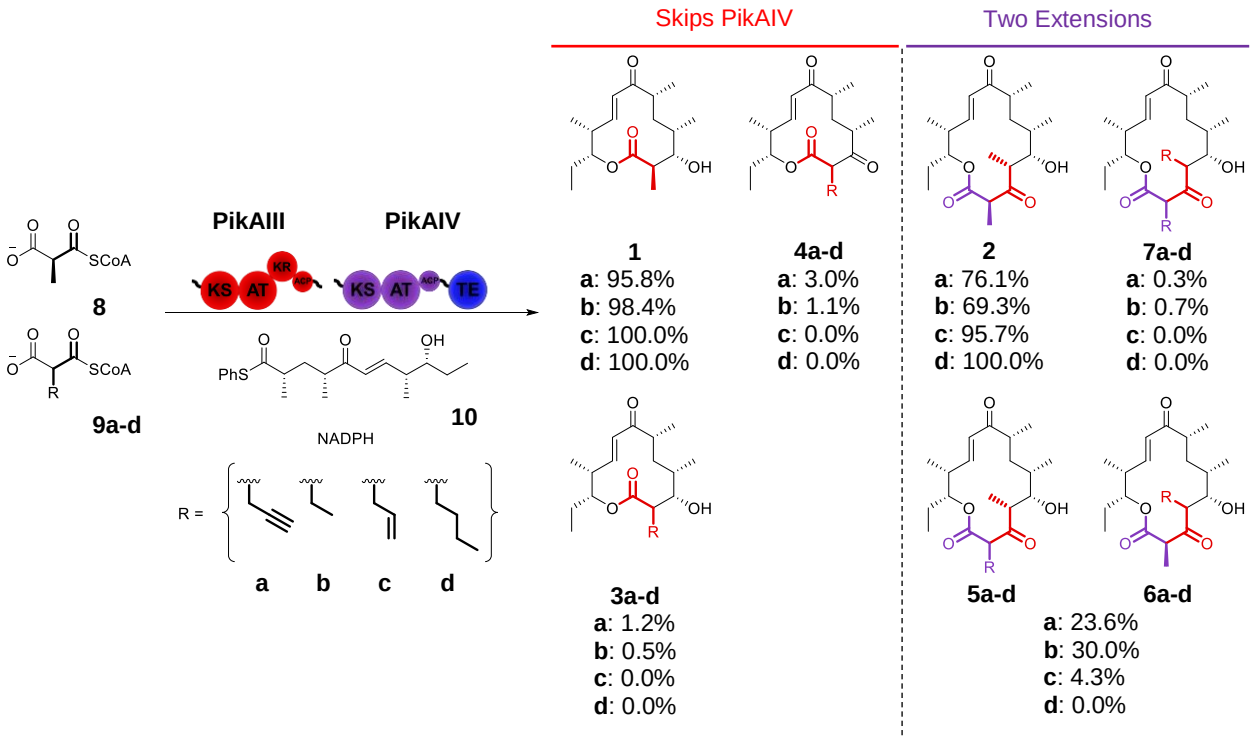


Figure 4.2. Bimodular Extender Unit Competition Assay. The two final Pik modules are incubated with the synthetic pentaketide chain mimic **10** and a mixture of the native extender **8** and an equimolar amount of one of **9a-d** *in vitro*. Products **1**, **3a-d**, and **4a-d** are produced when the PikAIII-extended chain bypasses module 6 and is cyclized by the TE. Products **4a-d** bypass the KR domain in PikAIII. Product distributions shown are for the wild-type system and are calculated separately for one- and two-extension products. Error was ± 5% of indicated

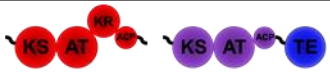
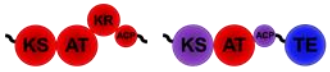
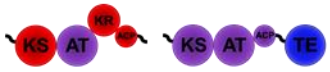
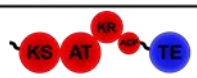
In the native *in vivo* system, both modules only take methylmalonyl-CoA (**8**), but upon incubation of the cell lysate with **10** and each of **9a-d**, only the propargyl- (**9a**), ethyl- (**9b**), and allyl- (**9c**) malonyl-CoA extender units were accepted in the presence of **8**. Notably though, in all cases, the non-natural single-extension products (**3a-c** and **4a-c**) made up 5% or less of the total 10-dML products while no more than 31% of the total narbonolide product was derived from the non-natural extender unit (e.g., **5a**+6**a**+**7a**; **Figure 4.2**). Regardless of the distribution of non-natural incorporation between PikAIII and PikAIV, together both modules are far less promiscuous than Ery6 as characterized by Koryakina and coworkers.[64] This narrow substrate scope provides an

opportunity for expansion through mutagenesis and to explore whether inherent extensive promiscuity is initially required for expansion via single amino acid mutations. Additionally, the discrepancy between the proportion of single and double extension products derived from the non-natural extender unit introduced new questions about promiscuity of individual modules within the same pathway, while detection of the previously unknown keto-10-dML products (**4a**-**4b**) raised the issue of processing efficiency by downstream domains.

4.2.2. *Comparison of the Pikromycin PKS Acyltransferases*

There are three gatekeeping domains potentially responsible for the low production of **3a-d**/**4a-d** by the modular Pik system: (1) the PikAT5, (2) the downstream PikKS6, or (3) the PikTE. Several chimeric modules were subsequently designed to probe the role of the AT domains in determining the observed product profile. The native AT of PikAIV was substituted with PikAT5 (R2, **Table 4.1**, **Table C6**), and the native AT of PikAIII was substituted with PikAT6 (R3, **Table 4.1**). In both of these chimeras, boundaries defined by the Keasling group were used but resulted in less total product as compared to the wild-type system,[56] with the R2 chimera less than a tenth as active as the wild-type system. As expected, based upon the assumption that the terminal AT (PikAT6) was more promiscuous than those upstream, the chimera that contained two PikAT5 domains, R2, produced almost 25% less propargyl **5a/6a** products than the wild-type system (**Table 4.1**, R2). Consistent with these results, the system containing two PikAT6 domains, R3, produced more than 15-fold higher levels of propargyl **3a**/**4a** and 1.6-fold more **5a/6a** products than the wild-type system (**Table 4.1**, R3). Together, these data indicate that PikAT6 is significantly more promiscuous than PikAT5 and that this feature may contribute to the low production of **3a-d**/**4a-d** by the wild-type system.

Table 4.1. Using Domain Swaps to Compare Native Promiscuities of ATs. The product distribution for the wild-type and mutant bimodular system are shown in the chart. Error was ± 5% of indicated value. Percentages were calculated for 10-dML and narbonolide products separately. N.D. = Not Detected. All enzymes listed expressed, and activities were normalized to concentration. **3a** and **4a** were summed to better highlight the effect of the AT.

Modules	10-dML **3a + 4a**	Narbonolide **5a/6a**	Narbonolide **7a**	Relative Total Activity
R1 	4.2%	23.6%	0.3%	100.0
R2 	11.8%	17.9%	0.2%	9.0
R3 	66.7%	37.2%	0.6%	47.5
R4 	0.8%	--	--	100.0
R5 	N.D.	--	--	0.0
R6 	7.4%	--	--	586.0
R7 	7.2%	--	--	180.3
R8 	N.D.	--	--	0.0

Interestingly, only 0.6% of the double-extension product from R3 contained two propargyls (**7a**) even though 66% of the single-extension products (**3a**+**4a**) are derived from the propargyl extender. However, upon closer examination, 99% of the single-extension product was the unreduced propargyl 10-dML product **4a** (data not shown). This discrepancy could be due to either PikKR5's native screening ability[169] or to perturbations in the KR due to the somewhat impaired chimera, but this lack of KR activity was only seen when a non-natural extender was used (**Figures 4.2** and **4.3**). As mass ions consistent with a keto-narbonolide product were not found, it is likely that PikKS6 discriminates against unreduced hexaketide chains.

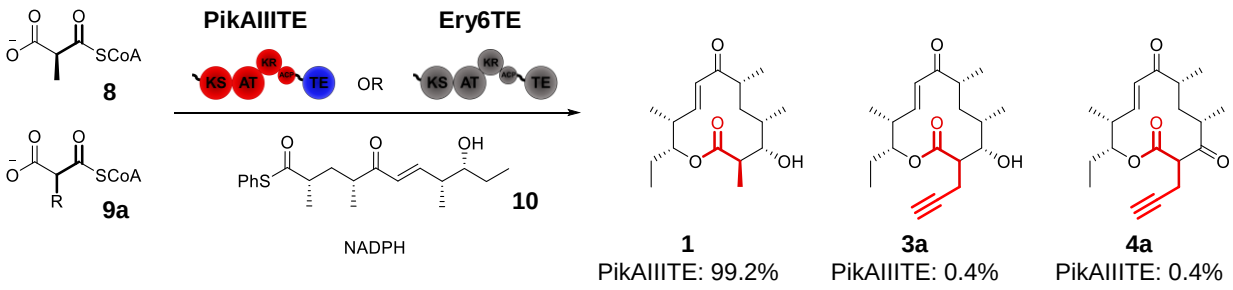


Figure 4.3. Single Module Extender Unit Competition Assay. The fusion protein PikAIIITE or the final module of the DEBS PKS, Ery6TE, are incubated with the synthetic pentaketide chain mimic **9** and a mixture of the native extender **7** and an equimolar amount of **8a** *in vitro*. Error was ± 5% of indicated value. Percentages are shown for wild-type systems.

To account for any substrate bias by domains within PikAIV that impact the product distributions, particularly the KS, PikAIII and PikAIV were decoupled by fusing PikAIII to PikTE. In addition to substrate competition assays with both modules (**Figure 4.2**), this now enables competition assays with PikAIIITE alone (**Figure 4.3**)—though not PikAIV alone as it does not accept the pentaketide substrate **10**. As expected for a single extension module, PikAIIITE produced only 10-dML-like products, and narbonolides were not detected. Notably though, of these products, only 0.8% were derived from the propargyl extender unit, with half of these comprising the keto-10-dML product (**Figure 4.3** and **Table 4.1**, R4). By comparison, 4.2% of the 10-dML products were derived from the propargyl extender unit when the wild-type bimodular system was used (**Table 4.1**, R1). Given that the ratio of propargyl product from R4 was no higher than that with the wild-type bimodular system, this result suggests that the lack of extender unit promiscuity displayed by PikAIII is likely due to PikAIII itself and not due to downstream gatekeeping by components of PikAIV—with the possible exception of the TE.

Finally, to determine the possible effect of the domains surrounding the AT within the module, each of the two Pik ATs were introduced into Ery6TE (R6, **Table 4.1**), a module and TE pair known to be capable of producing high yields of non-natural 10-dML products, yielding the chimeras R7 and R8 (**Table 4.1**).[63] Once again, issues with chimera stability prevented a direct comparison of the two ATs in these chimeras, as Ery6TE_PikAT6 (R8) was completely inactive. However, the Ery6TE_PikAT5 (R7) construct had a significantly more relaxed substrate selectivity (7.2% **3a**/**4a**) than did PikAIIITE (0.8% **3a**/**4a**) and nearly equal to that of wild-type Ery6TE (7.4% **3a**/**4a**). This 9-fold increase in propargyl incorporation by PikAT5 going from its native module to Ery6TE along with the results of the other three domain swaps establishes both the importance of AT substrate selection and of the substrate promiscuity of the other post-AT domains, especially the KS and the KR.

4.2.3. *Effects of Active Site Mutations on Extender Unit Incorporation*

In addition to the insight into substrate incorporation, the domain swaps also highlighted two other aspects important for PKS engineering. First, even with the recently updated AT boundaries,[56] domain swapping often results in inactive or poorly active enzymes. Second, some modules, regardless of AT selectivity, are naturally more substrate-permissive. To circumvent the first issue and take advantage of the second, AT active site mutagenesis should be utilized for maximum engineering efficiency.

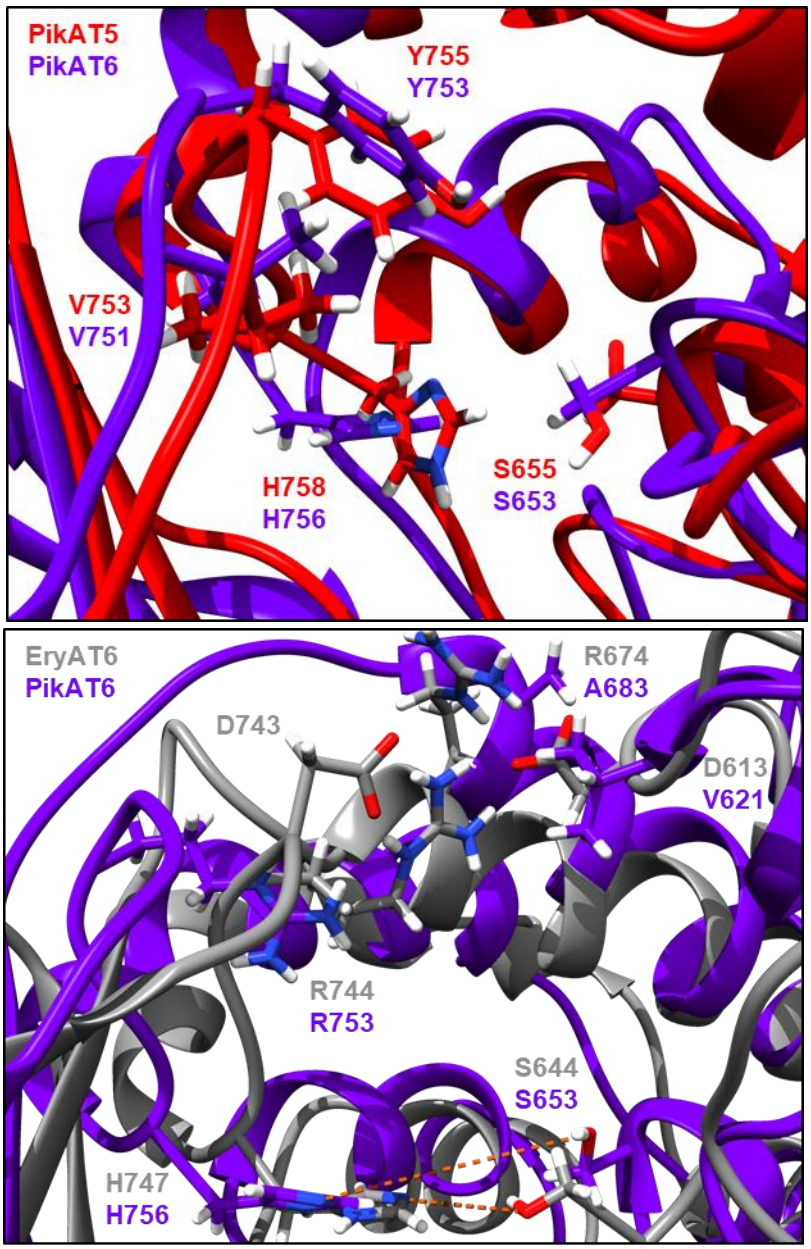


Figure 4.4. Pik and Ery AT Active Sites. Top. Homology models for PikAT5 and PikAT6 are overlaid. Minimal changes are seen between the models. Key residues shown include the catalytic Ser655/653 and His758/756 and two substrate specificity-controlling residues, Val753/751 and Tyr755/753. **Bottom.** Models of EryAT6 Y744R and PikAT6 Y753R ATs after undergoing MD simulations. Distances between catalytic residues in the mutant EryAT6 are a long but catalytically-competent distance of 4.8 Å, but the same distance in the mutant PikAT6 is 8.4 Å—characteristic of an inactive AT domain.

By inspection of homology models created for PikAT5 and PikAT6, there appears to be few if any major structural differences between the two active sites (**Figure 4.4b**, Top Panel). Both ATs also

share many of the same active site residues as EryAT6, including the YASH motif (**Appendix C, Figure C1**). Given that structural features of the PikAT5 and PikAT6 models could not explain the observed difference in extender unit promiscuity, the two best characterized mutations from EryAT6, V742A and Y744R (Ery6TE numbering), were introduced into PikAIIITE, and this single extension fusion module was tested with the pentaketide **10** and a mixture of extender units, **8**/**9a** (**Figure 4.3**, R9-R13).[62-63] In this single module system, **3a**/**4a** production increased 26-fold upon introduction of the V753A mutation, as judged by the relative amount of **3a**/**4a** *vs*. the methyl product, **1**. However, Y755R resulted in an expressed but completely inactive enzyme in contrast to the flip in selectivity seen in Ery6TE (**Table 4.2**, R11).[63]

As in Chapter 2, molecular dynamics (MD) simulations were utilized to explore why the Tyr→Arg mutation is effective in one methyl-specific AT (EryAT6) but not in the Pik ATs. In EryAT6, Arg744 forms a salt bridge with the neighboring Asp743, in competition with Arg674 (**Figure 4.4**, Bottom Panel). Both arginines also interact with Asp613, helping to distribute the positive charge and also keep the active site from widening too far. In PikAT6 Y753R, Arg753 simply sits in the active site. Instead of competing with a salt bridge (Chapter 2, Section 2.2.2), Arg674 of EryAT6 has been replaced with Ala683—which in turn interacts hydrophobically with Val621, which is in the same position as Asp613 in EryAT6 (**Figure 4.4**, Bottom Panel). This nonpolar environment is not as welcoming of a large and charged residue like arginine, and the distance between the two catalytic residues widens commensurately (8.4 Å in PikAT6 Y753R *vs* 4.8 Å in EryAT6 Y744R). This distance is too far to maintain the interactions between Ser653 and His756 necessary for catalytic activity.

Given the MD results and that Y755R resulted in an inactive PikAIIITE, the corresponding tyrosine to valine mutation found in many ethylmalonyl-CoA-specific ATs was introduced into PikAIIITE (**Table 4.2**, R12).[170] This mutation, which has also recently been introduced into EryAT2 by Vögeli and coworkers,[152] resulted in a significantly more promiscuous module with nearly half of the products derived from **9a**. Next, both functional mutations were introduced into PikAIII and PikAIV and assayed as part of a bimodular system (**Table 4.2**, R14-R19). The valine to alanine mutation in PikAIII resulted in a more modest in propargyl product formation as compared to the wild-type bimodular system (**Table 4.2**, R15); however, the Tyr→Val mutation in either module flipped the selectivity of the bimodular system, preferring the non-natural substrate **9a** over the natural substrate **8**, as judged by the ratio of narbonolide products **2**, **5a**/**6a**, and **7a**, e.g., 21.4% **2**, 61.9% **5a**/**6a**, and 16.7% **7a** (**Table 4.2**, R16). With PikAIII Y755V, a 13-fold increase in preference for propargyl was seen in the single extension product as compared with the wild-type bimodular system, but unlike with the unstable chimera R3, 90% of the product was reduced (**Appendix C**, **Table C6**, R16 and R14, respectively). Subsequently, whereas the wild-type bimodular system only supports production of 0.3% di-propargyl narbonolides, nearly 17% of the narbonolides formed by PikAIII Y755V are derived from two propargyl extender units (**7a**). This successful selectivity shift demonstrates that even with other domains potentially impacting substrate selection (e.g., KS, KR, and/or TE), even the most stringent of systems can be engineered to take a desired new substrate.

Table 4.2. Engineered Product Distribution for One- and Two-Module Systems. The product distribution for the wild-type and mutant PikAIIITE and bimodular systems are shown in the chart. Error was ± 5% of indicated value. Percentages were calculated for 10-dML and narbonolide products separately. Relative activities were set to 100 for the two wild-type systems. N.D. = Not Detected. Total activities were normalized to enzyme concentration. **3a** and **4a** were summed to better highlight the effect of the AT.

				10-dML	Narbonolide		Relative Total Activity
	PikAIIITE	PikAIII	PikAIV	**3a + 4a**	**5a/6a**	**7a**	
R9	WT	--	--	0.8%	--	--	100.0
R10	V753A	--	--	21.1%	--	--	4.2
R11	Y755R	--	--	N.D.	--	--	0.0
R12	Y755V	--	--	42.2%	--	--	81.5
R13	WT (apo)	--	--	N.D.	--	--	0.0
R14	--	WT	WT	4.2%	23.6%	0.3%	100.0
R15	--	V753A	WT	7.7%	21.3%	0.8%	92.5
R16	--	Y755V	WT	55.9%	61.9%	16.7%	82.1
R17	--	WT	Y753V	4.4%	77.2%	1.3%	86.8
R18	--	Y755V	Y753V	44.6%	59.6%	14.7%	51.7
R19	--	WT (apo)	WT (apo)	N.D.	N.D.	N.D.	0.0

The PikAIII Y755V/PikAIV Y753V system was also probed for substrate tolerance beyond methyl and propargyl by including either ethyl-, allyl-, or butylmalonyl-CoA in equimolar concentrations with methylmalonyl-CoA (**Table 4.3**, **Table C5**). As desired, it showed improved activity with all three non-natural substrates tested. For ethyl, the non-natural narbonolide products were also more prevalent than the native product **2**. Gratifyingly, the allyl and butyl extenders that were not accepted to any degree by wild-type PikAIII (or even by the more promiscuous PikAIV for butyl)

were accepted by both mutant modules, even resulting in the double allyl product **7c**, albeit in reduced percentage yields as compared to the propargyl and ethyl products. This is particularly interesting given that in the previous study by Koryakina and coworkers, non-natural narbonolide products could not be produced with an engineered bimodular DEBS system.[63] Also of note, none of the single-extension butyl product was reduced by the KR, preventing production of any double-butyl narbonolide product and setting the stage for further engineering in the future.

Table 4.3. *In Vitro* Engineered Product Distribution with Non-Natural Extender Units. The product distribution for the PikAIII Y755V/PikAIV Y753V bimodular system are shown in the chart. Error was ± 5% of indicated value. Percentages were calculated for 10-dML and narbonolide products separately. Fold increase was calculated for the double mutant compared to wild-type and shown below percentages. # Fold increase based on limit of detection of 0.001%.

	PikAIII Y755V + PikAIV Y753V			
	10-dML		Narbonolide	
	3a-d	**4a-d**	**5a-d/6a-d**	**7a-d**
Propargyl (**a**)	41.0%	3.6%	59.6%	14.7%
	10.6x		2.5x	49.0x
Ethyl (**b**)	8.0%	1.8%	49.1%	17.4%
	6.1x		1.6x	24.9x
Allyl (**c**)	3.7%	4.3%	16.7%	3.3%
	>8000x #		3.9x	>3300x #
Butyl (**d**)	0.0%	11.3%	5.4%	0.0%
	>11300x #		>5400x #	--

4.3. Discussion

The promise of engineering PKSs for production of non-natural polyketides has been apparent ever since the modular nature of these assembly line pathways was first discovered nearly 30 years ago. PKSs provide the ability to template a series of reactions in order to achieve a desired product, with each module responsible for recruitment of a single extender unit into the polyketide scaffold. However, the envisioned goal of being able to stitch together any domain or module seamlessly

like Legos still cannot be attained without significant loss in enzyme function. In this study, even with two ATs sharing 88% identity, the AT swaps resulted in vastly different enzyme activity levels. Instead, to realize the potential of PKS engineering, specifically with substrate selection, site-specific mutations can be introduced to shift selectivity towards or away from a given substrate.

Herein, two very similar ATs from the pikromycin PKS were compared with each other and with the well-characterized EryAT6. Despite these three ATs sharing all residues known or predicted to influence substrate selectivity, they exhibit very different levels of substrate promiscuity. Additionally, mutations that cause a substrate selectivity flip in EryAT6 (Y744R) result in inactive enzymes in the Pik ATs. In other cases, such as with the Tyr→Val mutation explored here, the mutation results in a shift towards larger extenders, demonstrating that an AT with as little native promiscuity as PikAT5 can be engineered to prefer non-natural substrates (4.2% **3a**/**4a** to 55.9% **3a**/**4a**, **Table 4.2**, R14 and R16).

In its native setting, PikAIV accounted for the clear majority of non-natural extender unit incorporation by the bimodular system. Through AT swaps and mutagenesis, much of this discrepancy in promiscuity between PikAIII and PikAIV was shown to locate with the ATs themselves; however, as demonstrated by the 9-fold change in percent propargyl product for two PikAT5-containing modules (**Table 4.1**, R4 and R7) and the significant discrimination by PikKR5 against larger extenders (**Table 4.3**), the surrounding domains play a significant role in the distribution of natural *vs* non-natural products. The effect of the KR is especially notable regarding the incorporation of non-natural extender units—especially as the extenders become larger (e.g.,

butyl). As the selectivity of the AT is altered, the ketoreductase (and potentially other domains) will need to be exchanged or engineered to be compatible with the desired product. As engineering of these domains progresses, larger and more exotic extenders can be used to further push the limits of these modular assembly line-like enzymes.

Cumulatively, site-specific mutagenesis of PikAT5 and PikAT6 has led to robust yields of the first full-length polyketide with two non-natural extenders and to the first non-terminal methyl-specific module to take and process a non-natural extender unit. These new 10-dML and narbonolide analogues can be further derivatized through semi-synthetic chemistry, especially 'click chemistry.' In the future, PKS engineering for non-natural incorporation will depend on four main areas: (1) describing the rules of AT substrate selectivity to create orthogonal ATs, (2) the substrate throttling effect by downstream domains and modules, (3) improved understanding of protein-protein interactions for domain and module swaps, and (4) increasing the throughput of screening chimeric or mutant PKSs.

4.4. Materials and Methods

4.4.1. *General*

Materials and reagents were purchased from Sigma Aldrich (St. Louis, MO) unless otherwise noted. Isopropyl β-D-thioglactoside (IPTG) was purchased from Calbiochem (Gibbstown, NJ). The TP-pentaketide (**10**), the PikAIII-pET24b construct, the PikAIIITE-pET28a construct, and the PikAIV-pET24b construct were kindly provided by Dr. David Sherman at the University of Michigan. The *E. coli* BAP1 strain was provided by Dr. Blaine Pfeifer at the University at Buffalo.[171] Construction of the Ery6TE-pET28 plasmid was described previously.[63] All module

sequences are listed in **Appendix C**, **Table C1**. Primers were purchased from Integrated DNA Technologies (Coralville, IA). All *holo* proteins were expressed in BAP1 cells and all *apo* proteins were expressed in *E. coli* BL21 (DE3) cells. All primers and construct sequences are listed in **Appendix C**, **Table C2**.

4.4.2. *Mutagenesis of Modules*

Phusion polymerase (New England Biolabs, NEB, Ipswich, MA) was used for round-the-horn mutagenesis of all PikAIII and PikAIV mutants.[154] AT swaps of PikAIII AT and PikAIV AT were constructed using Gibson assembly (NEB) and boundaries defined by the Keasling group.[27] Correct mutants were transformed into BAP1 competent cells with the pRARE plasmid (Novagen) and plated onto LB agar plates (50 μg/mL kanamycin and 100 μg/ml spectinomycin).

4.4.3. *Lysate Preparation*

Modules were expressed overnight at 16 °C in 300 mL cultures of LB media with the appropriate antibiotics. Protein production was induced with 1 mM IPTG at OD_{600} of 0.6. After overnight expression, the culture was centrifuged at 4,700 rpm for 20 min, and the supernatant was discarded. The cells were resuspended in 1 mL module storage buffer (100 mM sodium phosphate, pH 7.4, 1 mM EDTA, 1 mM tris(2-carboxyethyl)phosphine (TCEP), 20% v/v glycerol, 0.1 μL Benzonase (NEB), and EDTA-free protease inhibitor cocktail (Roche, Basel, CH)) and sonicated using 51% amplitude, 10 sec on, 20 sec off for 10 min. After sonication, the lysed cells were centrifuged at 18,000 rpm for 1 h. The lysates were aliquoted and stored at -80 °C.

4.4.4. *MatB Reactions and Acyl-CoA Preparation*

MatB wild-type and MatB T207G/M306I were purified and 8mM malonyl-CoA (**8**, **9a-d**) stocks were set up as previously described.[156]

4.4.5. *Pentaketide Assay*

The pentaketide assay was set up with a total volume of 80 μL in 100 mM sodium phosphate, pH 7.0, and 20 mM $MgCl_2$. The reaction conditions included 1 mM TP-pentaketide (**10**), 1.75 mM of each competing malonyl-CoA from MatB reactions (**8**, **9a-d**), an NADPH regeneration system (5 mM glucose-6-phosphate, 500 μM $NADP^+$, and 0.008 U/mL glucose-6-phosphate dehydrogenase), and lysate containing the module or modules. For the bimodular reactions, the proteins were included as a total of 29.4 μL lysate. Module concentrations were calculated by Bradford assay and SDS-PAGE gel analysis. Reactions proceeded at room temperature for 16 h and were quenched with an equal volume of -20 °C methanol. After quenching, all reactions were centrifuged at 13,300 rpm three times for 3 h total, and the supernatant was filtered through a nylon 0.2 μm filter. Analysis was carried out on a high-resolution mass spectrometer – a ThermoFisher Scientific Exactive Plus MS, a benchtop full-scan Orbitrap™ mass spectrometer – using Heated Electrospray Ionization (HESI). The sample was analyzed via LC-MS injection into the mass spectrometer at a flow rate of 225 μL/min. The mobile phase B was acetonitrile with 0.1% formic acid and mobile phase A was water with 0.1% formic acid (see **Appendix C**, **Table C3** for gradient and scan parameters). The mass spectrometer was operated in positive ion mode. The LC column was a Thermo Hypersil Gold 50 x 2.1 mm, 1.9 μm particle size. This assay produces 10-dML and narbonolide products that can be seen as their $[M+H]^+$, $[M+H-H_2O]^+$, and $[M+Na]^+$ ions and keto-10-dML products that can be seen as their $[M+H]^+$ and $[M+Na]^+$ ions. Extracted ions for each

listed ion were summed for comparison purposes. For retention times, calculated masses, observed masses, representative extracted ion counts, and representative chromatograms, see **Appendix C**, **Tables C4**, **C5**, and **C6** and **Figure C2**.

4.4.6. *Homology Models and Molecular Dynamics*

Wild-type homology models for EryAT6, PikAT5, and PikAT6 were created using the I-TASSER online server.[124-126] Mutations were introduced into structurally-converged wild-type models with Discovery Studio 4.1 from Accelrys Software, Inc. (San Diego, CA). Molecular graphics and analyses of MD trajectories and PDB snapshots were performed with VMD 1.9.2 and UCSF Chimera 1.10.1.[134-136] Further analysis was performed with CPPTRAJ.[137] Images were rendered with POV-Ray.[138]

Using the AMBER14 software package, individual models' charges were neutralized with sodium ions in Xleap.[128] All models were then solvated with a 15 Å buffer of TIP3P water, also within Xleap. The enzymes and substrates were parameterized with ff12SB and GAFF force fields from the AMBER14 software package. Prior to production MD simulations, solvated systems were treated with four heating and seven minimization steps. Steps 2, 3, 5, and 11 heated the system to 300 K over times of 20-100 ps each. The first nine steps held the protein fixed, with the restraint constant being lowered each step. Steps 10 and 11 used no restraints. Minimization steps were completed when the change in the root mean square was below 0.01 kcal/mol•Å for the first two minimization steps and below 0.001 kcal/mol•Å for the remaining minimizations. Production simulations lasted between 10 ns and 30 ns for each model. Step times were 2 fs. The non-bonded interaction cut-off was imposed at 9.0 Å.

CHAPTER 5

Development of a Transcriptional Repressor Biosensor for Malonyl-CoA and Derivatives

5.1. Introduction

The ability to sense biomolecules—especially small molecules—*in situ* is one of the most important and useful technologies that can be developed for the areas of chemical and synthetic biology. For many of these small molecules, bacterial transcriptional regulators coupled to fluorescent readouts have been harnessed as biosensors as tools for metabolite detection, biosynthetic circuits, metabolic engineering, or directed evolution.[85, 172-174] Malonyl-CoA (mCoA) is one such small molecule, and it plays an integral part of a cell's metabolism in its unique role as a building block for primary and secondary metabolites. Because of this, mCoA has been the target of several genetically-encoded biosensors and associated metabolic engineering efforts.[175] While mCoA has been of particular interest for the production of fatty acids, flavonoids, and polyketides, its derivatives have received far less attention despite their many biological uses (**Figure 5.1D**).[153, 176-178] As these derivatives (e.g., methylmalonyl-CoA) are not required for primary metabolism but are instead often synthesized by individual enzymes found in secondary metabolite gene clusters (**Figure 5.1A-C**),[19] they should be prime targets for the high-throughput evolution enabled by a genetically-encoded biosensor.

FapR is a transcriptional regulator found in nearly all Gram-positive bacteria, and it acts as a global regulator for fatty acid biosynthesis.[179] Importantly, mCoA acts as the sole effector ligand known for FapR, and FapR has been characterized as the only member of its regulator family to go with a unique, highly-dynamic structure (**Figure 5.2A-B**).[179-181] Published crystal structures of FapR show a dimer, with each monomer comprised of a C-terminal ligand-binding domain and an N-

terminal domain that binds to its cognate DNA operator, *fapO*. Additionally, FapR has been shown to act as both a transcriptional activator and repressor, depending on the promoter.[182] Finally, FapR has been utilized as a mCoA biosensor in several cell types including *E. coli*,[182-184] yeast,[185-186] and mammalian.[187]

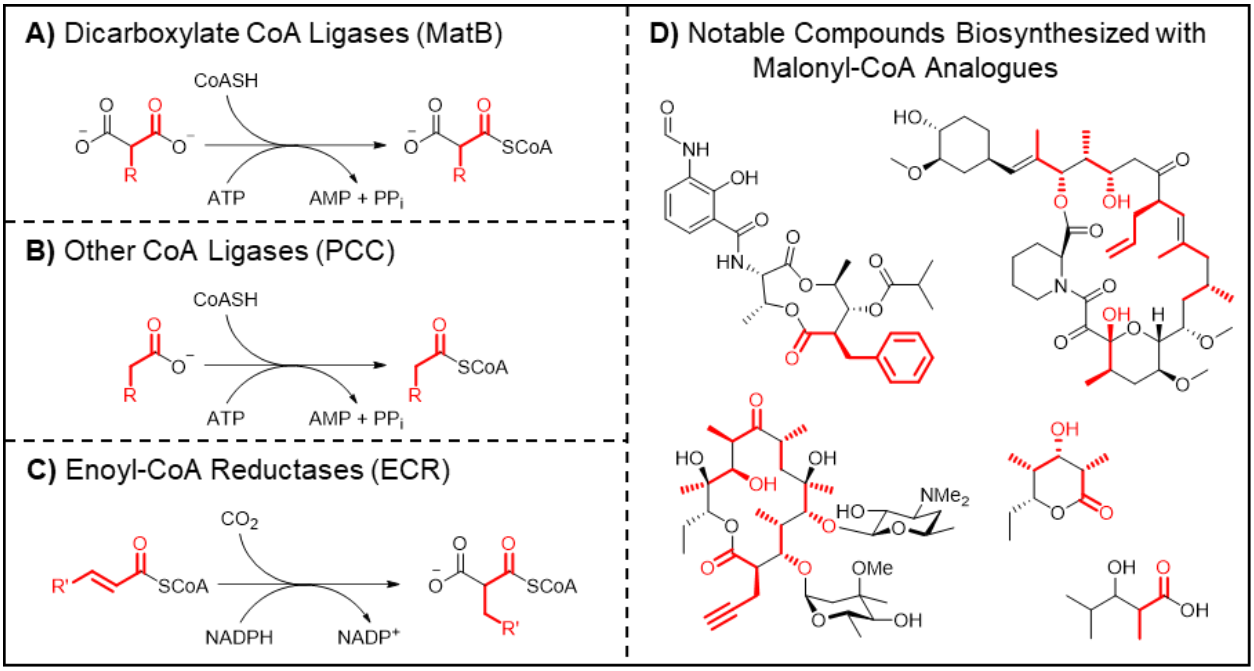


Figure 5.1. Origins and Uses of Malonyl-CoA Analogues. A/B/C) Three examples of enzymes classes responsible for production of malonyl-CoA derivatives are shown. Often these enzymes need to be engineered for broader substrate scope. **D)** Examples of biofuels and clinically-relevant compounds biosynthesized by polyketide synthases are shown with malonyl-CoA analogue-derived moieties highlighted in red.

To utilize FapR as a biosensor in *E. coli*, the *fapR* gene from *Bacillus subtilis* was first cloned into a plasmid under T7 control and LacI control alongside an *eGFP* gene also controlled by T7, LacI, and FapR.[183] This FapR biosensor, while not optimized or used as a tool, was an important first step by establishing that a FapR biosensor could be used to quantify mCoA over a linear range (0.1 to 1.1 nmol/mgDW). The next iterations of the FapR biosensor in *E. coli* were also out of the

Koffas lab, and by utilizing a T7 promoter (repressor) and a pGAP promoter (activator), fatty acid biosynthesis was balanced against mCoA biosynthesis.[182] With the same biosensors, the Koffas lab also explored the effect of the number of *fapO* sites placed downstream of the reporter gene promoter. In 2015, the Zhang lab created a fatty acid production negative feedback loop based on FapR and LacI to alleviate the toxicity from acetyl-CoA carboxylase (*acc*) overexpression, leading to a one-third improvement in fatty acid titer in *E. coli*.[184] In each of these examples, advancements in the FapR biosensor system were made, but each of the published biosensors also suffered from the flaws of its predecessor. Additionally, none of the biosensors were utilized for substrates beyond malonyl-CoA or for products beyond fatty acids. Herein, we present a FapR biosensor that requires no outside effectors (e.g., IPTG) and is optimized for a range of mCoA concentrations. FapR is also shown to recognize methylmalonyl-CoA (mmCoA) both *in vitro* and *in vivo*, and strides are made towards the creation of a biosensor specific for malonyl-CoA derivatives diversified at the C2-sidechain.

5.2. Results

5.2.1. *Refactoring the FapR Operon as an Efficient Malonyl-CoA Biosensor*

Of the previously published *E. coli* malonyl-CoA biosensors, one worked as a repressor dependent on T7 transcription controlled by the LacI/*lacO* system, another was dependent on the arabinose-dependent pBAD promoter, and the third acted as an activator with the constitutive pGAP promoter.[182-184] For the design of a new biosensor, a construct was desired that required no external inputs to function (e.g., IPTG) and that was robust enough to use for evolving FapR or other proteins of interest. Towards this goal, a new plasmid, pSENSE2FF, was designed and built based on the biosensor plasmid from an erythromycin biosensor.[85] Two constitutive *lac* promoters (no

lacO) control the transcription of a fluorescent reporter gene (*sfGFP*, Pro_{sfGFP}) and an operon containing *fapR* from *B. subtilis* and an additional gene (*mCherry* in the original construct, Pro_{fapR}). Downstream of Pro_{sfGFP}, a single copy of the 17 bp *fapO* operator was inserted to introduce FapR control over transcription of the reporter (**Figure 5.2C**/**D**). Super-folder GFP (sfGFP) was selected due to its higher brightness and low photobleaching.[188] The constitutive mCherry expression was included to confirm the presence of the biosensor in a cell even when no sfGFP was being produced. This would be especially useful during fluorescence-activated cell sorting (FACS) when trying to discriminate between small *E. coli* cells and surrounding noise. Later constructs would replace the *mCherry* gene with other genes of interest such as *matB* (**Figure 5.2D**).

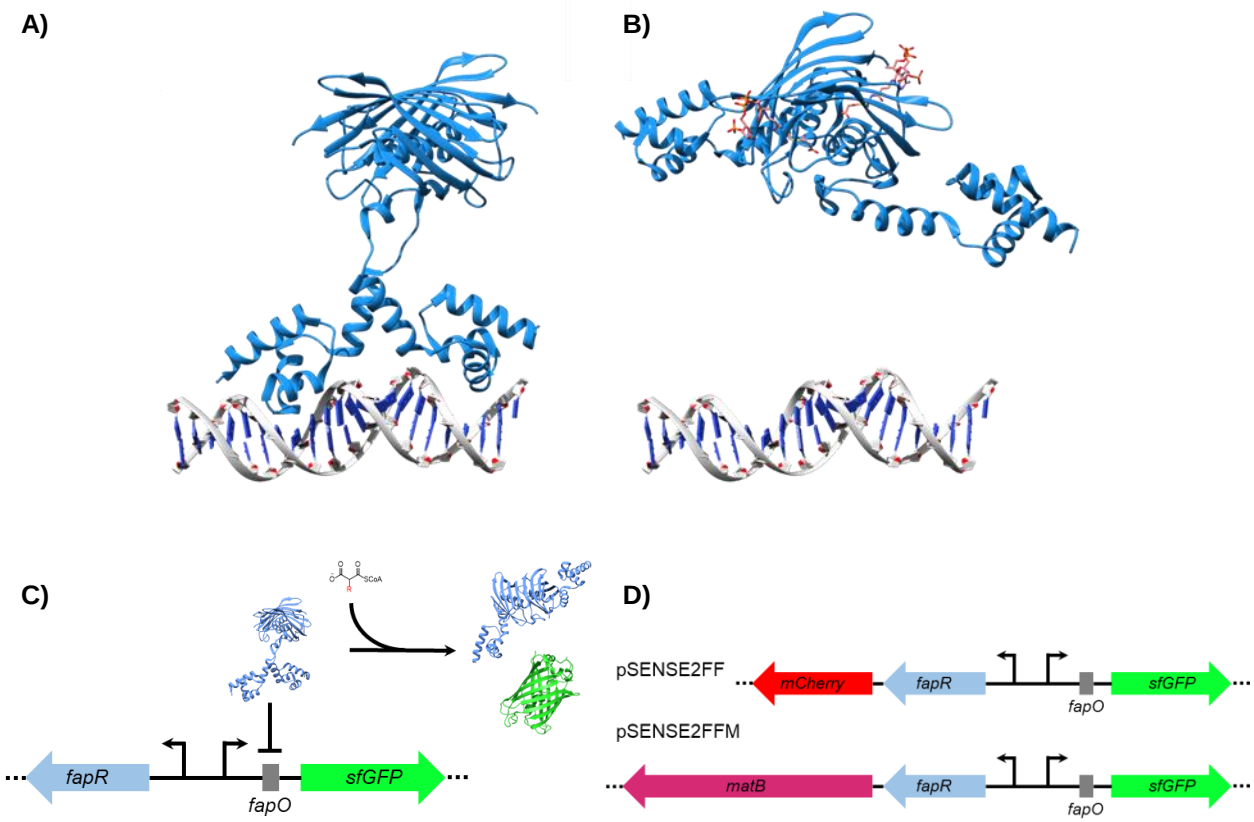


Figure 5.2. Redesigned Malonyl-CoA Biosensor. A) Crystal structure of FapR dimer bound to its operator sequence *fapO* (PDB: 4A12). **B)** Crystal structure of FapR dimer with two molecules of mCoA bound (PDB: 4A0Z). **C)** At low levels of malonyl-CoA, a FapR dimer binds to its cognate operator and represses transcription. At increased levels, malonyl-CoA binds and causes a conformational change, allowing transcription of a reporter gene. **D)** The two biosensor constructs built for this study.

5.2.2. *Optimization of Malonyl-CoA Biosensor* In Vivo

Once the initial biosensor had been constructed, it needed to be optimized for different uses and cellular environments (e.g., higher sensitivity or different strains). Previous work with FapR has focused primarily on varying the number of *fapO* binding sites, but even then, there was a finite limit before the effect leveled off.[182-183] Additionally, a single binding site should yield the most sensitive biosensor, so instead, an 18-member RBS library for *fapR* was constructed. As the original construct (RBS 1A1) had a low activation ratio (~5-fold, **Figure 5.3**) and less repressor

protein was hypothesized to result in higher activation, the RBS library was designed to have a maximum calculated transcription initiation rate (TIR) equal to that of RBS 1A1 (7594 au) based on the Salis RBS calculator.[189-191]

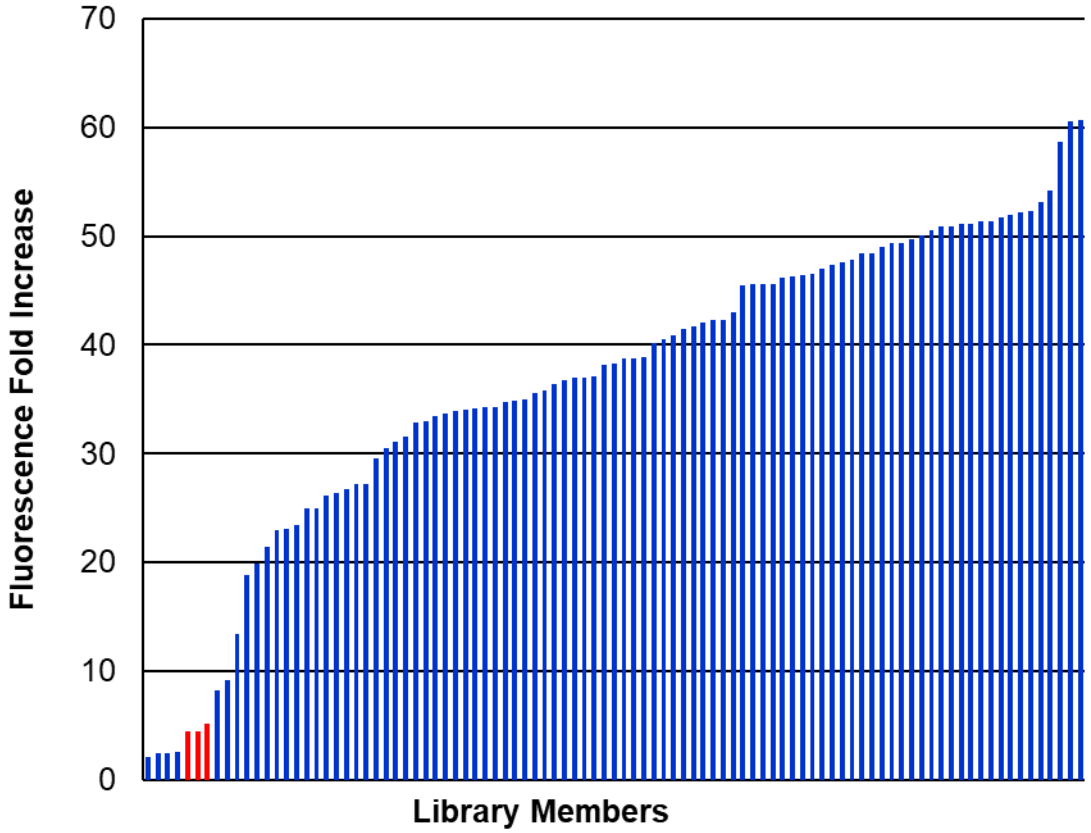


Figure 5.3. Representative RBS Library Screen. A FapR RBS library was screened with 0μM and 25μM cerulenin and compared with the original RBS (1A1, **red**). The larger the fold fluorescence increase, the better the response of the biosensor at the screened concentration of malonyl-CoA. 3 wild-type cultures and 93 RBS variants are shown.

Given that mCoA is not cell-permeable, to determine an activation ratio *in vivo*, the intracellular mCoA levels were artificially increased by using a small molecule inhibitor. Cerulenin is an antibiotic that acts as an equimolar inhibitor of the β-keto-acyl-ACP synthase, and as such, it causes a build-up in intracellular mCoA due to its inability to be processed as fatty acids.[192] The

linear response of [mCoA] to [cerulenin] has been previously established for the concentrations needed here.[183] To screen the FapR RBS library, cells containing the biosensors were grown in 96-well plates and after one hour of growth, treated with either 0 or 25 μM cerulenin, and allowed to grow overnight. A total of 288 cultures, with 3 1A1 cultures per plate, were screened to ensure full library coverage. As desired, most RBS variants resulted in significantly higher activation than the original 1A1 construct, with some reaching greater than 60-fold activation at 25 μM cerulenin (**Figure 5.3**). Gratifyingly, there was a mostly linear correlation between TIR and fluorescence activation.

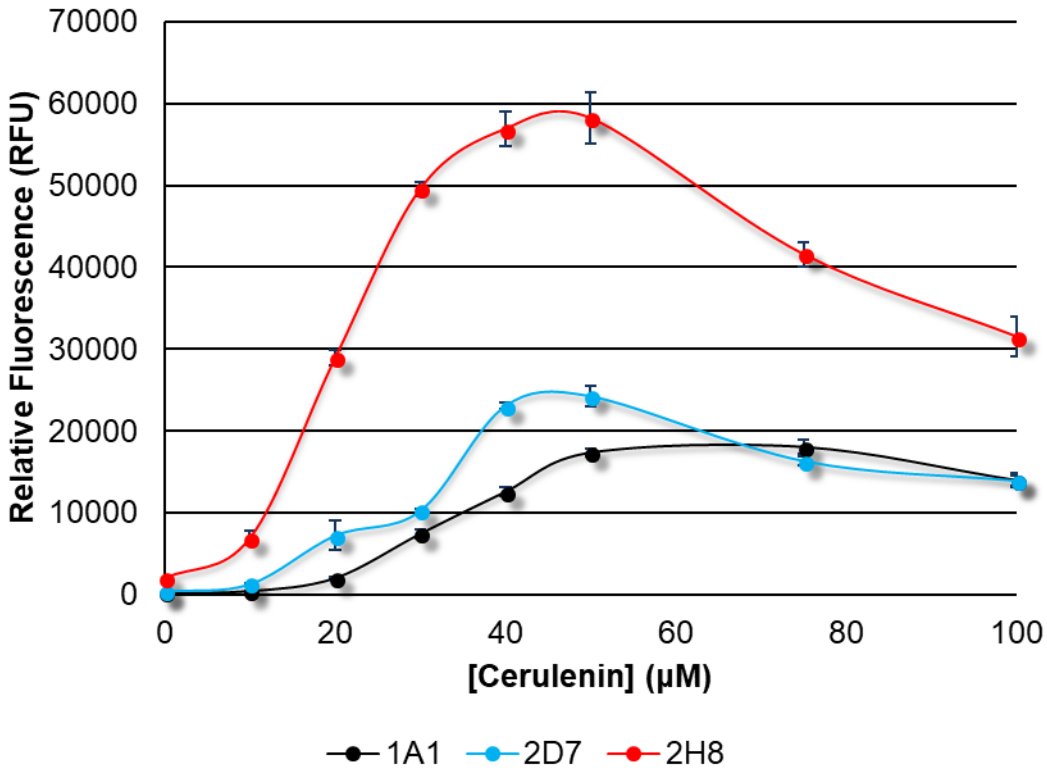


Figure 5.4. Dose-Response Curves from Select RBS Variants. Most selected RBS variants did not differ significantly from the original RBS except in fold fluorescence increase. Variant 2H8 had a significantly higher OFF (~2,000 vs ~200) and ON state (~58,000 vs ~18,000), in addition to a lower $K_{1/2}$ (19.81 ± 2.2 μM vs 30.89 ± 1.1 μM) and different peak response concentrations (50 μM vs 75 μM).

Several RBS variants were selected from this library for more thorough screening with cerulenin. Of these, two of the more interesting variants, 2D7 and 2H8, are shown alongside the original 1A1 construct (**Figure 5.4**). 2D7 did not vary much from 1A1 except in its fluorescence increase at concentrations of 50 μM and lower; however, like 1A1, its low maximum fluorescence at cerulenin concentrations in which sfGFP expression was not affected (< 75 μM) was not ideal for detection of a wide range of mCoA concentrations. Conversely, 2H8, with a very weak RBS, had a lower activation ratio than 2D7, but 2H8 had a maximal fluorescent response of 58,000 RFU, more than 3 times higher than 1A1. Unfortunately, this advantage came with the side effect of a leaky OFF state of ~2,000 RFU, nearly 10 times higher than 1A1's OFF state and almost 4 times higher than that of 2D7. Intriguingly, 2H8 was also calculated to have a $K_{1/2}$ value for cerulenin one-third lower and a lower peak fluorescent response (50 μM *vs* 75 μM) than 1A1, resulting in a more sensitive biosensor (**Figure 5.4**).

5.2.3. *Comparison of Biosensors in Different Strains*

While improvements in biosensor sensitivity such as those described in the previous section are not unprecedented after altering a repressor's RBS,[85] a mCoA biosensor has the additional caveat of having to account for basal levels of ligand. Many biosensors are used to detect compounds that are not native to the host organism, but mCoA has the complication of being required for cell growth. Additionally, the levels of mCoA fluctuate significantly during growth.[182] During the RBS library screen (**Figures 5.3** and **5.4**), the original biosensor, 1A1, was expressed in *E. coli* DH5α (hereon DH5α), but the other library members had been transformed into a commercial cloning strain termed *E. cloni* 10G (hereon 10G) similar to the more common DH5α and TOP10 strains. Often researchers use these cell lines interchangeably; however, due to the significant differences

in biosensing properties of 1A1 and 2H8 and to the indirect effect of using an inhibitor (cerulenin) to boost mCoA levels, 2H8 was transformed into DH5α, 10G, and TOP10 and compared to 1A1 in DH5α (**Figure 5.5**).

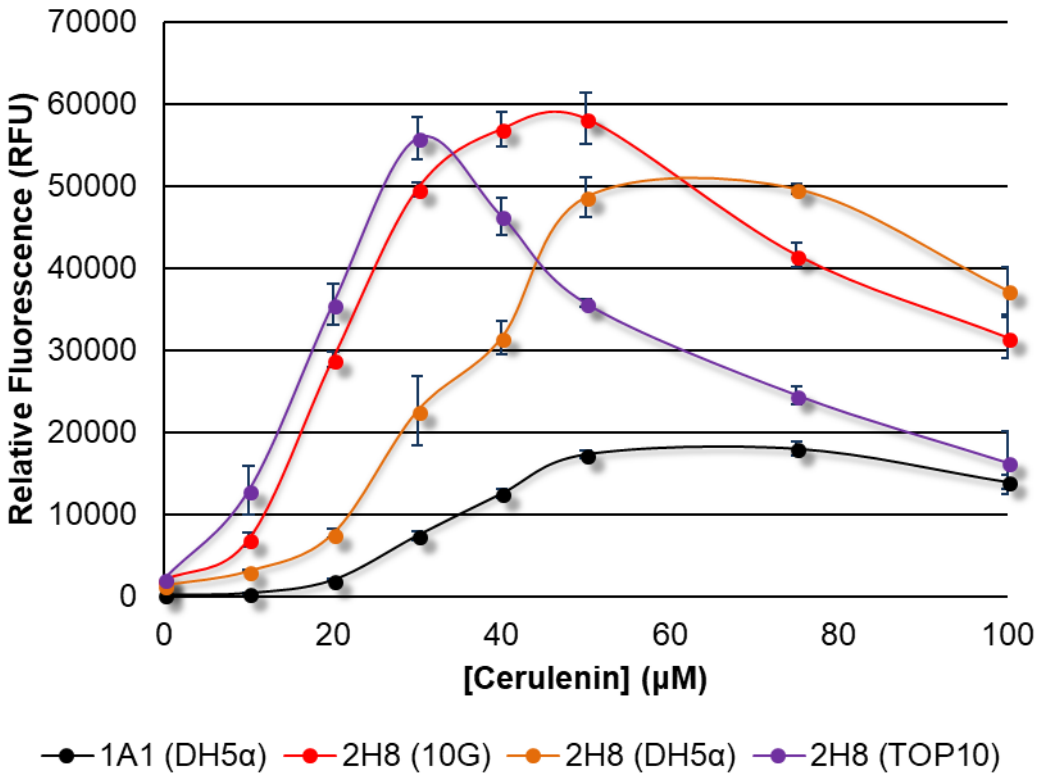


Figure 5.5. Biosensor Responses Vary Between Strains. Biosensor variant 2H8 was introduced into three different *E. coli* cloning strains: DH5α, 10G, and TOP10. Increasing cerulenin resulted in varied results among the three strains.

As **Figure 5.5** depicts, there are not only significant differences between biosensors 2H8 and 1A1, but there are also major differences in how these three *E. coli* cloning strains responded to cerulenin. Too much mCoA is toxic to the cell,[193-194] so a cell with higher endogenous levels of mCoA should result in both improved sensitivity to cerulenin and an earlier drop-off in fluorescence due to toxicity. This effect is seen most drastically in TOP10, where fluorescence peaks at 25 μM cerulenin, in contrast to 10G (50 μM) and DH5α (75 μM). While cloning strains

are not likely to be used for much beyond biosensor validation, these data highlight how even minor or even seemingly unrelated changes to the genome can affect either the FapR protein levels or the CoA and acyl-CoA pool (genotypes shown in **Appendix D**, **Table D1**).

5.2.4. *Detection of Methylmalonyl-CoA* In Vivo *and* In Vitro

Once the constructed biosensors had been optimized and characterized for mCoA detection, the next step was to work towards a functional biosensor for mCoA derivatives like mmCoA that are not natively produced in *E. coli*. A strain engineered for mmCoA-based production of polyketides, K207-3, was tested with the FapR biosensor for this purpose.[153] K207-3 has several genomic alterations compared with the previously-tested cloning strains, but importantly, it contains the propionyl-CoA carboxylase (PCC) genes *pccB* and *accA1* from *Steptomyces coelicolor* under T7 polymerase/LacI control.[195] This allows for mmCoA production from propionate when induced with IPTG.

The 2H8 biosensor was introduced into the K207-3 strain and screened by plate reader in the presence or absence of the inducer IPTG and increasing levels of the substrate sodium propionate. Upon induction of the PCC genes, fluorescence increased from 2,000 RFU to more than 15,000 RFU (**Figure 5.6**). A further nominal increase was seen with addition of sodium propionate; however, as mmCoA can be synthesized using endogenous propionate and propionyl-CoA,[176] exogenous propionate was likely not needed without an additional enzyme to consume mmCoA.

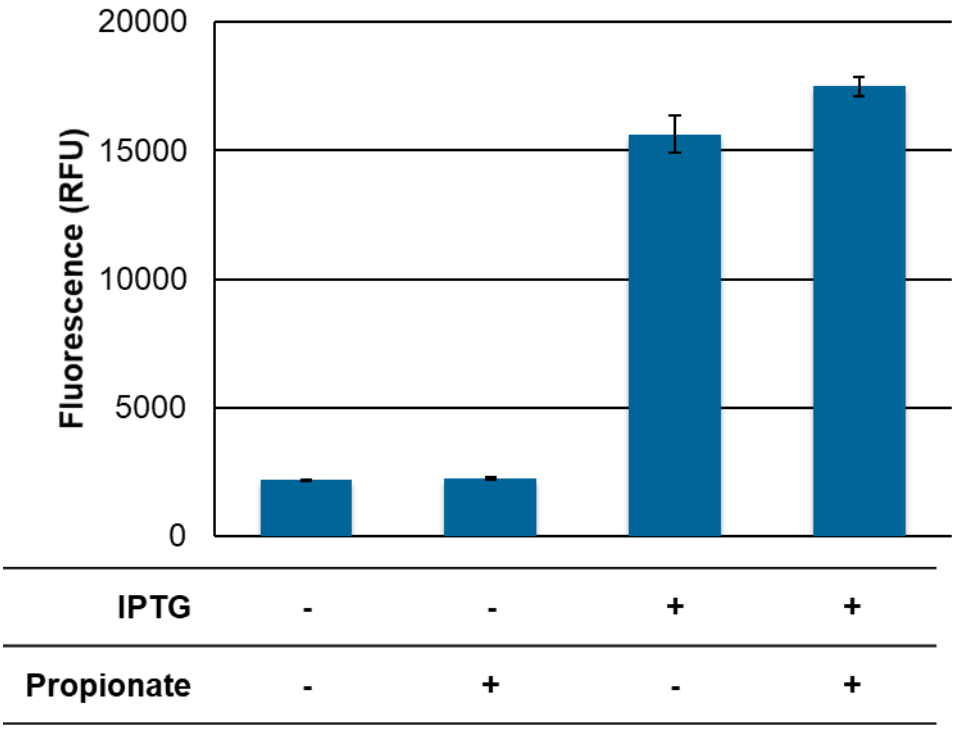


Figure 5.6. Use of Biosensor in Methylmalonyl-CoA-Producing Strain. *E. coli* K207-3 can biosynthesize methylmalonyl-CoA using propionate after induction with IPTG. Biosensor 2H8 was activated upon induction of the responsible genes by IPTG, where endogenous levels of propionate were sufficient. 30 mM propionate resulted in little to no increase in signal, likely due to no consumption of methylmalonyl-CoA in K207-3.

To confirm that FapR recognizes mmCoA, FapR was tested *in vitro* for DNA-binding and release with mCoA and mmCoA. Using purified wild-type FapR protein, an electrophoretic mobility shift assay (EMSA) was run with a 40 bp DNA fragment containing the *fapO* binding site (**Figure 5.7**). Incubation of the DNA with a two-fold excess of FapR resulted in near-complete binding and a corresponding gel shift. Varying concentrations of mCoA were titrated against FapR to determine the concentration at which half of the DNA would be bound and 200 μM was determined to be sufficient (data not shown). 200 μM mmCoA was then incubated with the FapR-*fapO* complex and a similar level of de-repression was observed (**Figure 5.7**). As previously reported, many other

acyl-CoAs did not result in de-repression of the FapR system,[180] making mmCoA only the second known ligand for FapR.

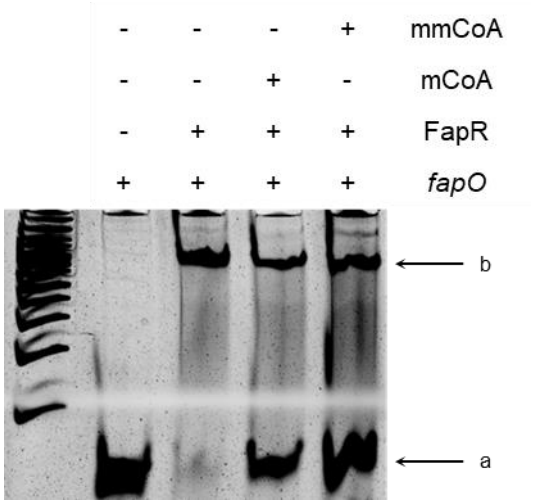


Figure 5.7. *In Vitro* Results Show FapR Recognizes Methylmalonyl-CoA. An EMSA gel demonstrating binding of a *fapO* oligo by purified FapR and then subsequent release upon addition of either malonyl-CoA or methylmalonyl-CoA. (**a**) unbound *fapO* and (**b**) bound *fapO*.

5.2.5. *Altering the Ligand Specificity of FapR*

While it is possible to utilize the wild-type FapR biosensor as a tool for mmCoA detection or even evolution, it will lack sensitivity due to the fluctuating levels of mCoA. To create a biosensor specific for non-native mCoA derivatives, recognition of endogenous mCoA needs to be eliminated. In an effort towards this goal, the crystal structure of FapR with mCoA bound (PDB: 2F3X) was inspected, and three residues near the 2-position of mCoA were identified: Phe99, Ile104, and Ala105. Phe99 was of particular interest due to its steric bulk and proximity to the mCoA (**Figure 5.8A**).

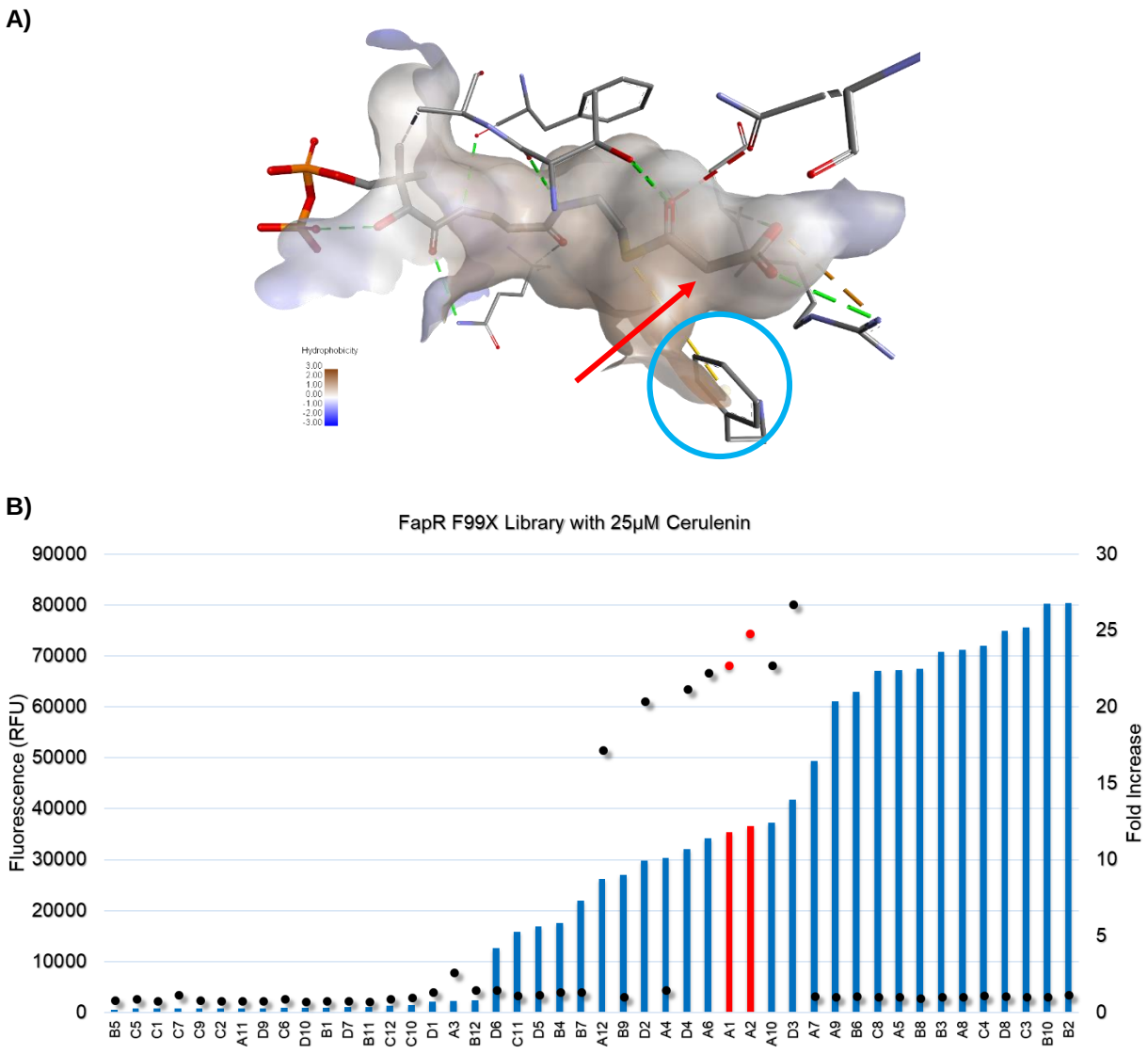


Figure 5.8. FapR Binding Pocket and F99X Saturation Library. A) The malonyl-CoA binding pocket of FapR wild-type with a hydrophobicity surface (brown = hydrophobic; blue = hydrophilic) from PDB 2F3X. The red arrow indicates the 2 position of malonyl-CoA and the blue circle indicates residue Phe99. **B)** Library was screened with 0μM and 25μM cerulenin to identify DNA-binding mutants that do not bind malonyl-CoA. Selected variants were sequenced and screened with higher cerulenin concentrations to ensure no effect from malonyl-CoA. The bars represent the fluorescence output with 25μM cerulenin and the dots represent the fold increase in fluorescence.

To see if mutagenesis at this residue would enable a functional repressor that does not recognize mCoA, a F99X saturation library was cloned into pSENSE2FF and screened with 0 μM and 25

μM cerulenin. With wild-type FapR (**Figure 5.8**, red bars), fluorescence reached nearly 40,000 RFUs in the presence of increased mCoA levels (~25-fold increase), but several members of the saturation library, including F99A, resulted in no increased fluorescence 25 μM cerulenin. The FapR mutant F99A (or another mutant from this library) could be used in conjunction with error-prone mutagenesis and FACS to sort for a biosensor that remains off in the absence of a larger mCoA derivative and turns on in its presence.

5.3. Discussion

The development of a robust, genetically-encoded biosensor for mCoA and its many derivatives, natural and non-natural, is one of the critical needs for our synthetic biology toolbox. The efforts described here provide a foundation towards this goal and a blueprint for future biosensor development. We have demonstrated the ability to optimize a constitutive biosensor with an endogenous metabolite, mCoA, and an exogenous metabolite, mmCoA, by focusing on RBS mutagenesis. We have also highlighted the effect different cellular backgrounds can have when working with sensitive transcription factor-based biosensors. Finally, we have taken the first step towards developing a biosensor specific for larger malonyl-CoA derivatives. Together, these refactored FapR biosensors can be utilized for metabolic engineering or perhaps more potently, for directed evolution of various enzymes responsible for the biosynthesis of natural or non-natural malonyl-CoA derivatives.

5.4. Materials and Methods

5.4.1. *General*

Materials and reagents were purchased from Sigma Aldrich (St. Louis, MO) unless otherwise noted. Isopropyl β-D-thioglactoside (IPTG) was purchased from Calbiochem (Gibbstown, NJ). Primers were purchased from Integrated DNA Technologies (Coralville, IA). *E. cloni* 10G electrocompetent cells purchased from Lucigen Corporation (Middleton, WI). Cerulenin was purchased from Cayman Chemical (Ann Arbor, MI). All cultures were grown in LB media with 100 μg/mL ampicillin unless otherwise stated. Absorbance and fluorescence readings were taken in clear flat-bottom and black flat-bottom 96-well plates (Greiner Bio-One), respectively, in a BioTek Hybrid Synergy 4 plate reader.

5.4.2. *Construction of Plasmids and Error-Prone Libraries*

The pSENSE2 plasmid was synthesized by Twist Bioscience (San Francisco, CA) in two fragments (1 and 2). The two pieces were assembled using standard restriction digestion and ligation procedures. The FapR_FapO containing a codon-optimized *fapR* gene and a *fapO* operator was synthesized by Genewiz, Inc. (South Plainfield, NJ) and cloned into pSENSE2 between the *KpnI* and *SpeI* sites to give pSENSE2FF. MatB was cloned into pET28a as previously described.[76] *matB* was amplified with the pET28a RBS, a 5' *KpnI* site, and a 3' *NotI* site. MatB wild-type and MatB T207G/M306I were cloned into pSENSE2FF to give pSENSE2FFM. *FapR* was amplified with a 5' *NcoI* site and a 3' *XhoI* site and cloned into pET28a to give fapR_pET28a (with a C-terminal His_6 tag). Error-prone mutagenesis was performed according to manufacturer's instructions for Mutazyme II (Agilent Technologies, Santa Clara, CA). Sequences are listed in Appendix D.

5.4.3. *RBS Library Construction and Screening*

The original RBS in front of *fapR* was calculated to have a transcription initiation rate (TIR) of 7594 using the Salis online RBS calculator.[189-191] The calculator was used to design an 18-member RBS library with a maximum TIR of 7594. Site-directed mutagenesis was used to produce the RBS library (VRGAGGH). Library was grown in 96-deepwell plates with 500 μL media. Cultures were grown overnight and 10μL from each well were used to inoculate 440 μL media. Plates were grown for one hour at 37°C and 350 rpm. Plates were then treated with either 5 μM, 12.5 μM, or 25 μM cerulenin or the corresponding volume of DMSO. Plates were then grown an additional 15 hours. Plates were centrifuged at 3,500 rpm for 7 min, and cell pellets were resuspended in 500 μL PBS. 279 individual colonies from the RBS library were screened at different cerulenin concentrations. OD_{600} and GFP fluorescence readings were taken and normalized.

5.4.4. *Measuring Methylmalonyl-CoA Production in* E. coli *K207-3*

pSENSE2FF-2H8 was transformed into *E. coli* K207-3 and grown overnight in a 96-deepwell plate at 37 °C. A new 96-deepwell plate with 440 μL LB and 100 μg/mL ampicillin was inoculated with 10μL overnight culture and grown at 37 °C and 350 rpm for 2.5 h. The cultures were induced with combinations of 1mM IPTG and different concentrations of sodium propionate and allowed to grow for 16 h at 37 °C and 350 rpm. The 96-deepwell plate was centrifuged at 3500 rpm for 7 min, and the supernatant was discarded. Cell pellets were resuspended with 1 mL PBS. OD_{600} and GFP fluorescence readings were taken and normalized.

5.4.5. *Electrophoretic Mobility Shift Assays*

FapR_pET28a was transformed into BL21 (DE3) and grown in a 300 mL culture of LB with 100 μg/mL ampicillin. The culture was grown at 37 °C and 250 rpm until OD_{600} reached 0.6, and then the culture was induced with 1 mM IPTG and allowed to express at 22 °C and 250 rpm for 20 h. FapR was purified on a nickel column with a Bio-Rad BioLogic LP FPLC (Wash buffer: 50 mM Tris-HCl, pH 8, 300 mM NaCl, 20 mM imidazole, and 10% v/v glycerol, Elution buffer: 0 mM Tris-HCl, pH 8, 300 mM NaCl, 200 mM imidazole, and 10% v/v glycerol), concentrated with a 10 kDa MWCO filter, and stored in storage buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, and 10% v/v glycerol). The *fapO* oligos were annealed per standard procedure: 5'-CCTAGGACTATTAGTACCTAGTCTTAATTGTCCGGCATCC-3' and 5'-GGATGCCGG ACAATTAAGACTAGGTACTAATAGTCCTAGG-3'.

EMSA reactions were run in 20 μL volumes at 30 °C for 40 min with 2 pmol *fapO* and varying concentrations of FapR in reaction buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 5% v/v glycerol, 100 mM KCl, 0.01 mg/mL BSA). Reactions were run out on a 6% PAGE gel at 220 V. The gel was stained with SYBR Green I and imaged on a Typhoon FLA 7000. 6 pmol of FapR was determined to be the optimal amount (data not shown) and was used in subsequent reactions with 200 μM of either mCoA or mmCoA.

CHAPTER 6

Development of a Transcription Factor-Based Biosensor for Acyclic Terpenes

6.1. Introduction

6.1.1. *Directed Evolution*

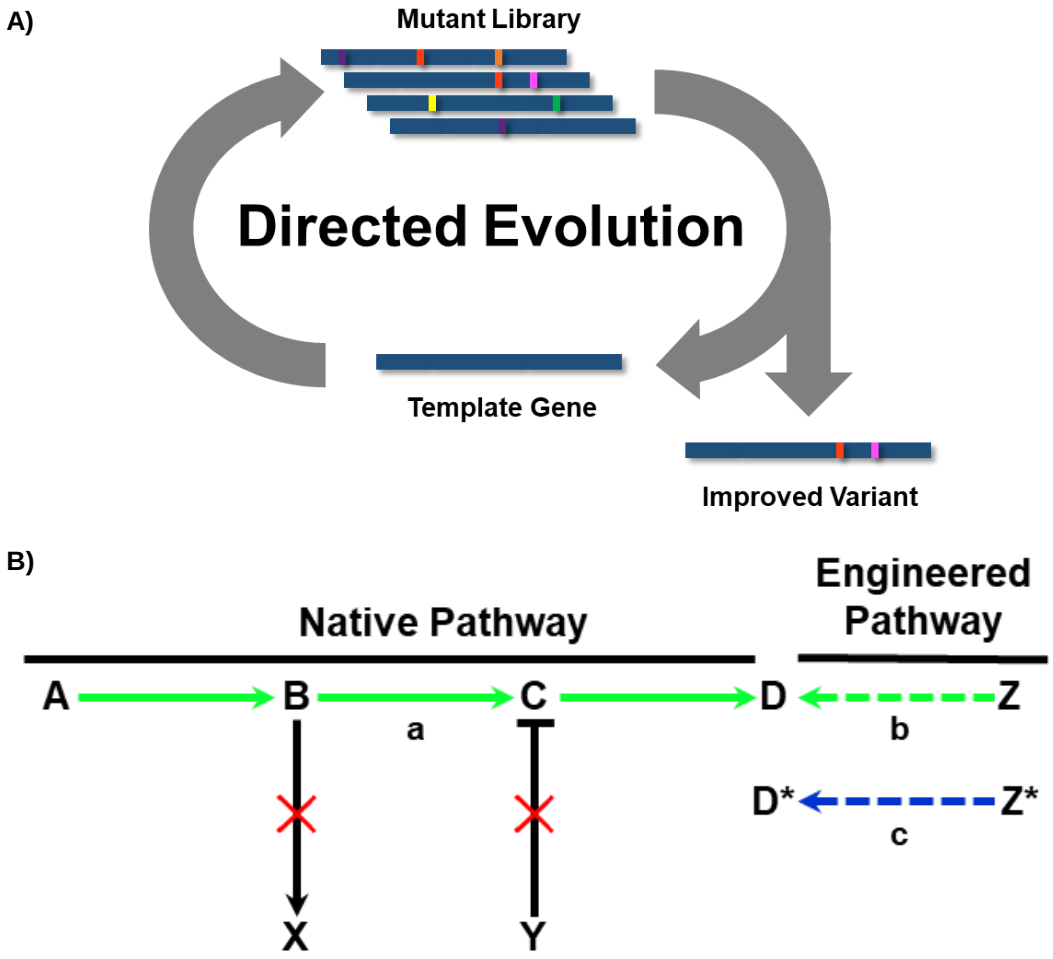


Figure 6.1. Directed Evolution and Uses for an Acyclic Terpene Biosensor. A) Directed evolution is an approach to protein engineering wherein multiple rounds of mutagenesis and screening can lead to mutants that would be impossible to predict. **B)** A genetically-encoded biosensor can be used for directed evolution of a native pathway (**a**) or an engineered pathway (**b**) for greater production of a desired compound. A promiscuous biosensor could also be used to optimize a novel pathway with non-natural analogues yielding products similar to the native ligands (**c**).

Directed evolution is the laboratory version of natural evolution (**Figure 6.1A**).[196] Instead of natural mutations arising randomly and selected by natural fitness, mutations in directed evolution

are introduced through *in vitro* mutagenic and selected for a desired phenotype through a carefully chosen screen or selection.[197] While a well-designed screen is able to sort through millions of possibilities for mutations that could not be predicted rationally, a poor screen will give false positives or limit the size of a mutant library that can be assayed. Thus, it is critical to spend the time perfecting a screen before subjecting a gene or pathway to multiple rounds of random mutagenesis and screening. Depending on the throughput of the screen, higher rates of mutation can be introduced at once. Currently, for engineering terpene biosynthesis, most work is done using low-throughput screens or selections based on mass spectrometry, a colorimetric reporter, or product toxicity.[105, 109, 198-199] The work described herein is the directed evolution of a repressor protein, AtuR, to act as a high-throughput *in vivo* biosensor for acyclic terpene production. Once evolved, this biosensor could be used as a tool for evolving several other enzymes or pathways (**Figure 6.1B**). A robust terpene biosensor therefore could have consequences in any field dependent on the production of terpenes, from biofuels to pharmaceuticals.

6.1.2. *Repressor Proteins and AtuR*

Repressor proteins, a subset of transcriptional regulators, act to prevent the expression of certain genes through binding to specific DNA sequences and blocking RNA polymerase from binding to the promoter. Bacteria often use these regulatory proteins to respond quickly and efficiently to stimuli. For example, in several *Pseudomonas* species, the (de)activation of the acyclic terpene utilization (Atu) pathway is modulated by AtuR.[200-201] This pathway is responsible for the catabolism of various monoterpenes such as citronellol (**1**, **Figure 6.3**) to units of acetyl-CoA.[202]

Förster-Fromme and coworkers established the role of AtuR through a variety of experiments

including electrophoretic shift mobility assays (EMSAs) to determine its DNA binding sequence (*atuO*, a 13bp sequence) and the creation of an AtuR-knockout *P. aeruginosa* strain. In the knock-out strain, proteins from the Atu pathway could be detected in cell lysates with or without monoterpenes present. In the wild-type strain, these proteins could only be detected in the presence of at least one of the six molecules shown in **Figure 6.3**. No expression was observed with non-branched molecules such as octanol or octanate, confirming the ligand profile of AtuR.

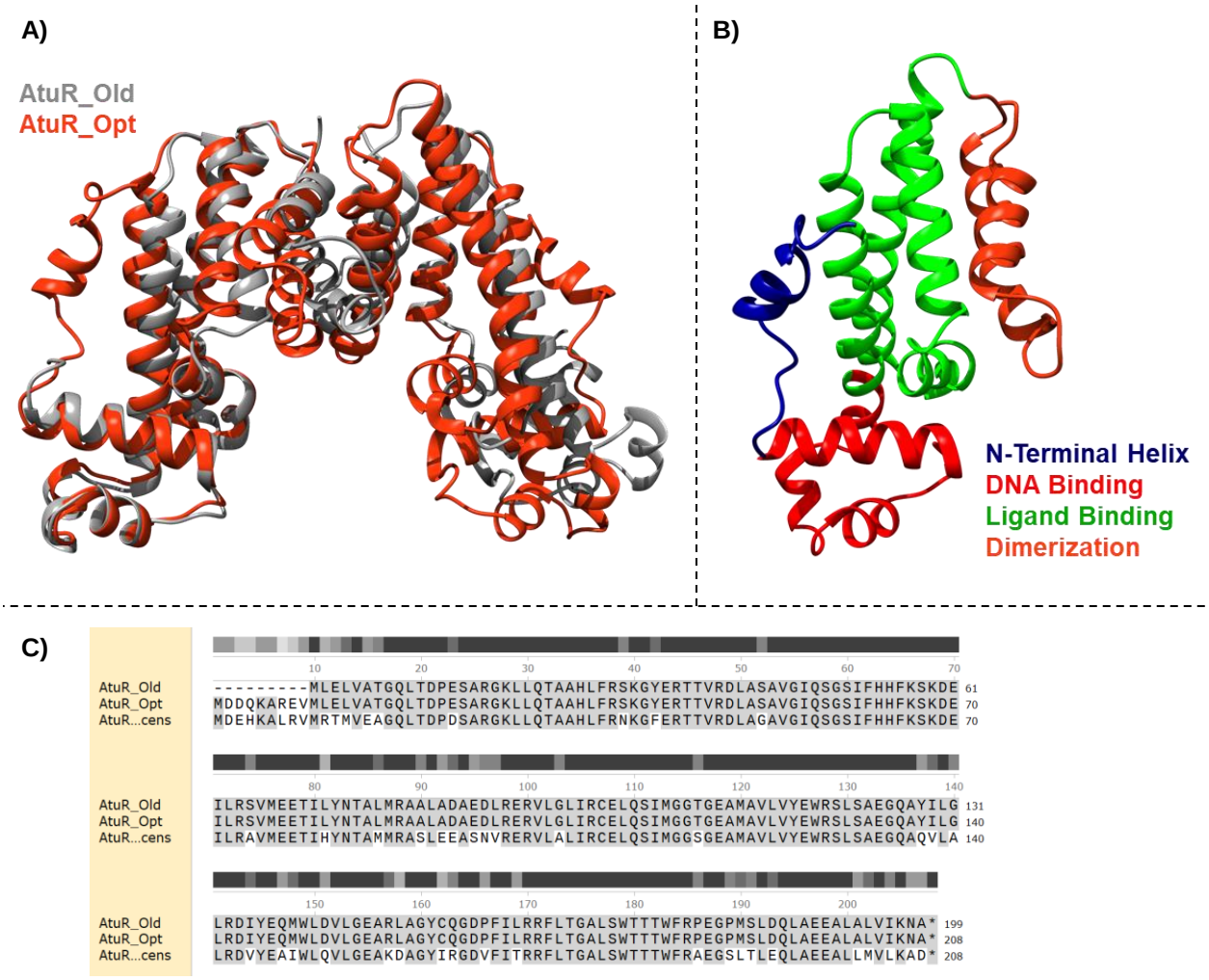


Figure 6.2. Homology Models of AtuR Using Published and Corrected Sequences. A) AtuR, the repressor protein responsible for regulation of the acyclic terpene utilization (atu) pathway in *Pseudomonas aeruginosa*, was originally characterized as a dimer (198 amino acids per monomer), but sequence alignments and RBS/promoter analysis led to the redefined AtuR with 207 amino acids per monomer. The new sequence has an additional N-terminal helix. **B)** An AtuR monomer has four regions: an N-terminal helix of unknown function, a DNA-binding domain, a ligand-binding domain made of 4 helices, and a dimerization interface. **C)** Sequence alignment of the old and new AtuR sequences and an AtuR homologue from *Pseudomonas fluorescens*.

Through sequence similarity, AtuR was categorized as a TetR-type repressor protein. This family is known for recognizing small molecules, often secondary metabolites including the polyketide tetracycline (recognized by TetR).[203] While none of the >100 solved unique TetR structures have more than 36% sequence identity to each other, all of the structures have nine or more helices,

including three found in the DNA-binding domain (DBD), with the remainder making up the ligand-binding and dimerization domains (**Figure 6.2B**).[204] There is no published crystal structure of AtuR; however, a 1.7 Å structure of PfiT, the *P. fluorescens* homologue of AtuR (78% sequence identity) was published in 2013.[205] A homology model was constructed for AtuR based on this crystal structure (**Figure 6.2A**), but little headway can been made with rational redesign of its ligand specificity due to the complicated interactions between the three protein regions of a repressor protein. In this chapter, AtuR is described for two purposes: the evolution of AtuR for improved recognition of acyclic terpenes and the use of AtuR as a model system for biosensor construction.

6.2. Results and Discussion

6.2.1. *Construction of an Acyclic Terpene Biosensor in pMLGFP: Initial Considerations*

Repressor proteins have been used as biosensors previously, but to the best of our knowledge, never to directly detect a terpene substrate.[173, 182, 199, 206-207] The goal was to develop a whole-cell fluorescent reporter for the production of acyclic terpenes, so ideally, GFP would only be produced in the presence of these small molecules, which are not naturally produced in *E. coli*. In addition to any interest in the valuable commodity chemicals some of these terpenes are, and in keeping with the scope of this dissertation, it was expected that non-natural acyclic compounds might be more readily prepared and detected than their cyclic counterparts. To create a terpene biosensor, *atuR* was subcloned along with two copies of its putative binding site, as found in its native operon, into a modified version of the plasmid host for an erythromycin biosensor, pMLGatuR (pMLG).[85, 208] pMLG houses the gene for the repressor protein and green fluorescent protein (*gfp*)

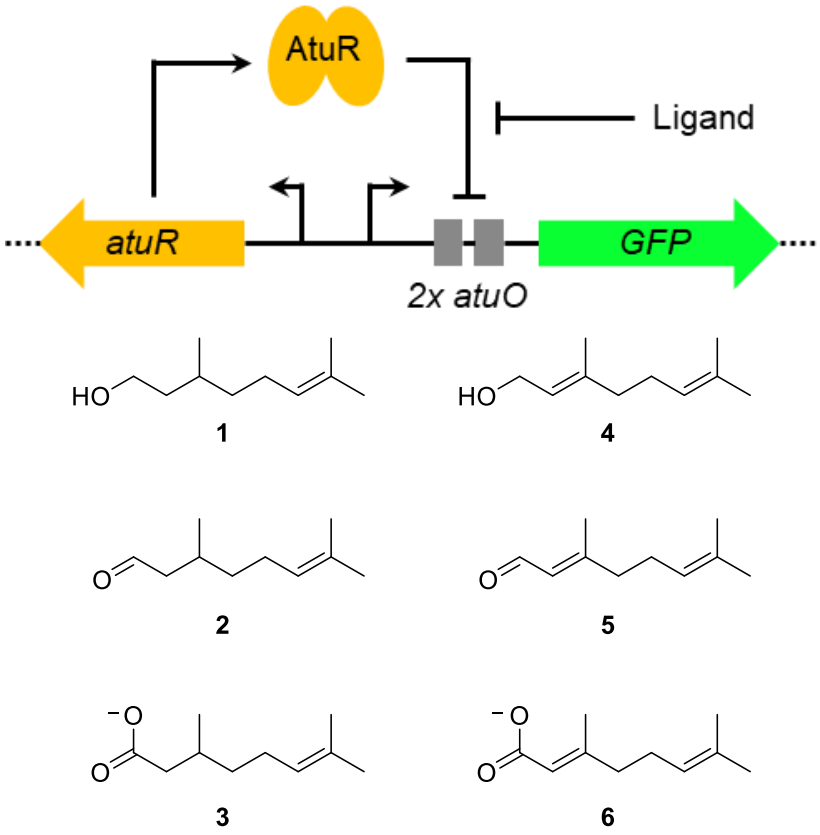


Figure 6.3. Original pMLG AtuR Biosensor and Natural Substrates. Both the repressor protein gene *atuR* and the reporter gene *GFP* are under constitutive control. Once expressed, the AtuR dimer binds to *atuO*, a 13bp recognition site, preventing transcription of the reporter gene. Compounds **1-6** are proposed to act as ligands for AtuR, resulting in a conformational shift preventing binding to *atuO*.

under constitutive expression (**Figure 6.3**). However, in this prototype system, pMLG displays leaky expression of GFP—likely due to poor DNA-binding by AtuR—and only shows a 10-30% increase in fluorescence upon induction with terpenes (**Figure 6.4B**). This result matches well with previously-published electrophoretic mobility shift assay (EMSA) data which showed AtuR was required in 8-fold excess to saturate both DNA binding sites.[200] While an exact binding affinity of AtuR with its operator has not been determined, it is likely quite low. Additionally, the crystal structure of the AtuR homolog, PfiT, indicates a distance of nearly 50 Å between the DBDs of the dimer, significantly further than the 35 Å of the typical TetR-type repressor (**Figure 6.2A**).[205, 209] Thus, before utilizing AtuR as a biosensor for metabolic engineering, it needs to be engineered to

improve not only its ligand-recognition, but also its DNA-binding ability.

The directed evolution of AtuR requires screening for several desirable traits simultaneously: higher expression, protein thermostability, tighter DNA binding, and higher ligand affinity. The higher the expression levels, the better the DNA binding, as described earlier with EMSA. As *atuR* is a *Pseudomonas* gene being transcribed in *E. coli*, it will naturally have some codons that are less common in the heterologous host. As such, any screen could pull out expression mutants—those that do not have an amino acid change, only silent mutations favorable for transcription or translation. Increased thermostability could lead to a protein that can tolerate additional unstable mutations in future rounds of mutagenesis—ones that may improve the other desired activities. This could be achieved by screening AtuR and its variants at 37 °C (wild-type expression was at 30 °C for Förster-Fromme and Jendrossek).[200] While the main goal of sorting remains to increase the fluorescent signal-to-noise ratio of background expression compared to the induced (ON) state, improved expression and thermostability are highly desirable traits even if not directly screened for. However, due to its high OFF state and relatively low ON state (*vide infra*), AtuR poses a significant challenge in this respect, as it is inherently difficult to screen for a good OFF state (no ligand present) and ON state (ligand present) simultaneously.

6.2.2. *Screening an AtuR Error-Prone Library Using pMLG as Template*

Without a good rational approach to engineering AtuR for improved DNA-binding, a random mutagenesis approach was taken towards evolving AtuR for the desired characteristics. For the first round of evolution, an initial error-prone library with 2-3 amino acid mutations per kilobase was generated (**Figure 6.4A**). Taking advantage of the fluorescent output of the biosensor,

fluorescence-activated cell sorting (FACS) was used to screen millions of random mutants from this library. Less than 1% of the lowest-fluorescent cells were collected in the absence of acyclic terpenes and were grown up as an enriched collection. They were then grown up as colonies and tested individually in a 96-well plate format. Cells were resuspended in PBS, measured for GFP fluorescence in the absence or presence of ligand (mixture of **1**-**6**), and corrected for cell density with OD_{600} by plate reader.

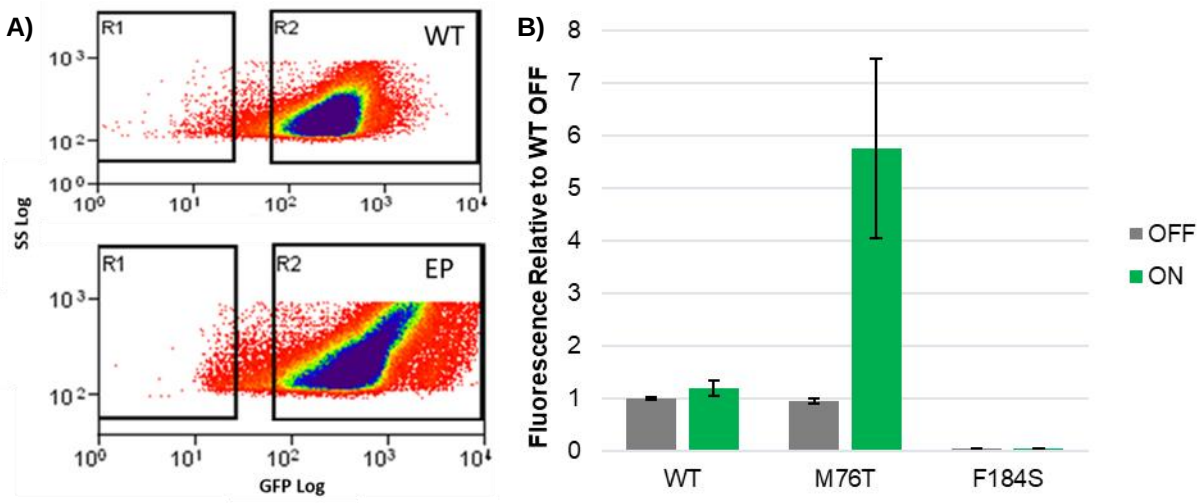


Figure 6.4. Initial Screening of Error-Prone AtuR Library. A) FACS plots showing GFP fluorescence for wild-type (WT) and error-prone (EP) populations of AtuR_Old in pMLG. **B)** Fluorescence from AtuR variants shows improved biosensor activity by the M76T mutant and an always-OFF F184S mutation. OFF = DMSO only. ON = 100 µM mixture of **1**-**6**.

In the first round of plate reader screening, two notable AtuR variants were identified. The first supported a 6-fold increase in fluorescence activation in the presence of 100 µM mixture of the natural ligands. This mutant had a single silent mutation and M76T, which is located at the interface of the DBD and the ligand-binding domain (LBD). Thus, the single mutation M67T provides the first evidence that the ligand specificity of AtuR can be manipulated by directed evolution. The second variant displayed 30-fold lower fluorescence in the OFF state compared to

the wild-type AtuR, but it showed no activation in the presence of terpenes. This variant also had a single silent mutation, but two amino acid substitutions as well: M148L near the top of the LBD, and F184S located in the dimerization region. Site-directed mutagenesis was used to isolate each of the three mutations to confirm their roles. M76T retained the mutant activity in the absence of the silent mutation (**Figure 6.4B**). M148L did not appear to have any effect on AtuR when separated from F184S, but F184S alone resulted in an always-OFF state (**Figure 6.4B**). The latter mutation was noteworthy for eliminating any reporter leakiness—especially for later being recognized as creating a functional repressor from a truncated protein.

AtuR M76T and F184S were each used as templates for a second round of error-prone mutagenesis. Mutations were introduced at a rate of 2-3 additional amino acid mutations per gene. One negative FACS sort per library was performed to select for correct OFF state variants. The enriched populations were initially screened for ON state variants by 96-well plate; however, mostly always-OFF mutants appeared, so an improved construct was needed. This prompted a re-analysis of the AtuR biosensor and subsequently resulted in a complete rebuild.

6.2.3. *Refinement and Recharacterization of the AtuR Biosensor*

After the initial evolution of AtuR was performed with limited success, as described in Sections 6.2.1 and 6.2.2, the model and sequence for AtuR was examined more closely, and the published sequence of AtuR (AtuR_Old) was determined to have used an incorrect start codon. The corrected sequence (AtuR_Opt, **Figure 6.2**) was utilized from thereon. To create a biosensor construct that could be used in a more robust screening process, a plasmid based on pMLG—with a few key improvements—was designed and built: pSENSE. To improve FACS screening, the brighter and

more photostable super-folder GFP (sfGFP) replaced GFPuv from pMLG as the fluorescent reporter.[188] Additionally, to better distinguish between living cells and instrument noise, mCherry was constitutively expressed using the same promoter as the repressor gene. This second

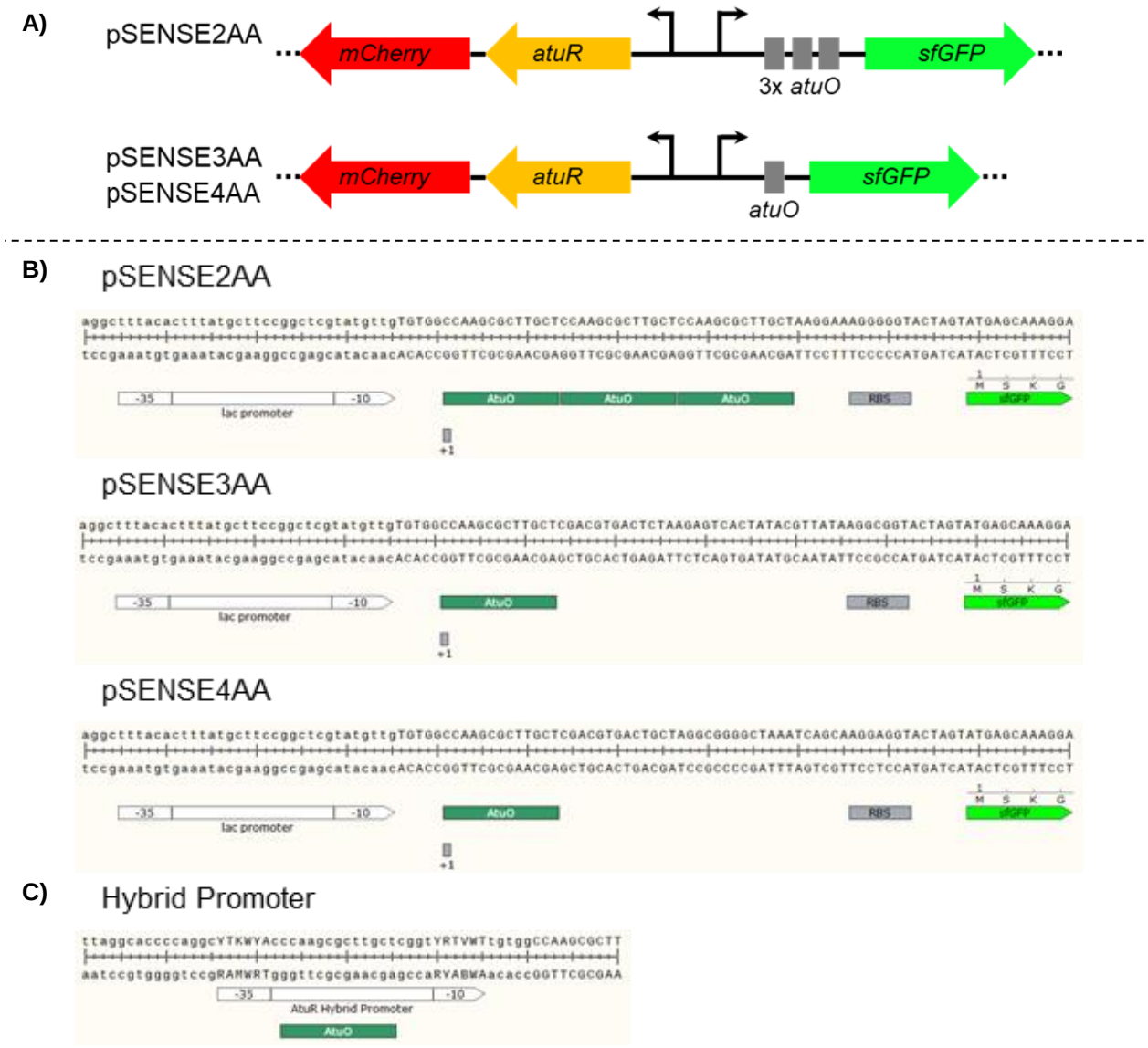


Figure 6.5. New AtuR Biosensors and AtuO Placement. A) pMLG was replaced with a new plasmid, pSENSE2, containing improved cloning sites, a brighter fluorescent reporter (*sfGFP*), and a constitutive *mCherry* reporter. All new biosensors contained the *atuR_Opt* sequence and varied only in the region directly upstream of *sfGFP*. **B)** pSENSE2AA contained 3 consecutive *atuO* sites. pSENSE3AA and 4AA only retained the upstream copy of *atuO*; however, 3AA was never fluorescent. **C)** A hybrid promoter containing *atuO* was designed with 300 variations.

fluorescent output should allow for more accurate positive and negative sorting. Finally, for ease of biosensor construction, common restriction sites were introduced in all three multiple cloning sites.

In the first iteration of the AtuR pSENSE biosensor, pSENSE2AA (**Figure 6.5**), three consecutive *atuO* binding sites were present between the *lac* promoter and *sfGFP*. Three sites were used due to the previously-characterized poor binding of AtuR to its operator; however, when the codon-optimized and full-length *atuR_Opt* gene was cloned into pSENSE2AA, fluorescence was low at any concentration of acyclic terpenes (up to the solubility limit). In the absence of the repressor, the sfGFP fluorescence was significantly higher than that from pMLG, as expected (data not shown). To weaken the AtuR repression, pSENSE3AA and pSENSE4AA (**Figure 6.5**) were built from pSENSE2AA, removing the two downstream *atuO* sites. In pSENSE3AA however, no sfGFP expression was observed even in the absence of repressor—despite a calculated *sfGFP* RBS strength equal to that in pSENSE2AA and pSENSE4AA (data not shown).[189-190] Perplexingly, pSENSE4AA shared the always OFF characteristic with pSENSE2AA in the presence of AtuR.

In addition to exploring the cis-acting *atuO* sites, the effects of the trans-acting repressor protein were also analyzed using *in vitro* studies with purified protein and a DNA ligand. Purified AtuR_Opt was tested with a series of DNA probes using EMSA. Usually EMSA is utilized with larger fragments of DNA—such as the several hundred base pair fragments used to originally identify the *atuO* sites;[200] however, to confirm the binding site for the new sequence, only the 13bp of *atuO* was used for AtuR binding. EMSA showed AtuR_Opt binding to *atuO* with a higher affinity than that published for AtuR_Old—albeit without exact parameters calculated in either

case—potentially due to the extra N-terminal helix present in the new structure (**Appendix E**, **Figure E1**). Furthermore, an oligonucleotide made up of the three consecutive *atuO* sites from pSENSE2AA also showed three shift patterns, indicating binding of three AtuR dimers simultaneously. These two data show that AtuR_Opt can bind both *atuO* and geraniol, but it does not explain the lack of biosensor activation in the presence of acyclic terpenes with the pSENSE constructs.

In addition to the differences in repressor protein sequence, the DNA sequences surrounding the various *atuO* sites may have played a role in the binding or release of AtuR. In pMLG, the two *atuO* sites were surrounded by the same sequences found in *P. aeruginosa*, but in the pSENSE vectors, only the 13bp required for binding were used. In the native setting, the *atuO* sites are located in between the -10 and -35 boxes of the putative promoter regions. To test if *atuO* might work better in the middle of the promoter rather than downstream, *atuO* was cloned into the middle of the *lac* promoter. Simultaneously, one of the native *atuO*-containing *Pseudomonas* promoters was swapped with the *lac* promoter, but unsurprisingly, neither promoter resulted in any sfGFP expression (data not shown). A final experiment was designed to create a functional *atuO* promoter in *E. coli*: a hybrid promoter with the consensus sequence of both promoters from *Pseudomonas aeruginosa* and the *lac* promoter from *E. coli* (**Figure 6.5C**). Once constructed, this ~300-member promoter library could be sorted with FACS or by plate reader in the absence (and subsequent presence) of AtuR_Opt.

In studying the AtuR repressor system in a biosensor format, the many challenges have led to the exploration of potential bottlenecks and optimizations of the system. Poor binding by AtuR_Old

led to the discovery of the correct AtuR_Opt sequence. Difficulties with accurate FACS sorting resulted in the construction of a biosensor built to address these issues. Finally, the several iterations of pMLG and pSENSE biosensors highlighted the importance of the location, number, and surroundings of repressor protein binding sites, and future studies with this system should result in even further advancements in both biosensor design and progress towards a fully functional acyclic terpene biosensor.

6.3. Materials and Methods

6.3.1. *General*

Materials and reagents were purchased from Sigma Aldrich (St. Louis, MO) unless otherwise noted. Isopropyl β-D-thioglactoside (IPTG) was purchased from Calbiochem (Gibbstown, NJ). Primers were purchased from Integrated DNA Technologies (Coralville, IA). Cells were grown in LB media (Fisher Scientific) and 100 μg mL^{-1} ampicillin (Sigma Aldrich) at 37 °C and 270 rpm for 16 h, unless noted otherwise. Negative controls had DMSO (Fisher Scientific) added in an equal volume to any terpenes added to positive controls. All terpenes were purchased from Santa Cruz Biotechnology, Inc. except for citral (Sigma Aldrich) and geraniol (Acros Organics). 1-octanol (Sigma Aldrich) was used as a second negative control. Terpenes and octanol were diluted in DMSO prior to addition to cell cultures. pET28-MHL was a gift from Cheryl Arrowsmith (Addgene plasmid # 26096). The sequences for *atuR_Old* and the codon-optimized *atuR_Opt* are listed in Appendix E. The sequence for pSENSE2AA is also listed in Appendix E.

6.3.2. *Construction and Mutagenesis of pMLG Plasmids*

The *atuR* gene was amplified from *Pseudomonas aeruginosa* PAO1 (ATCC 15692) genomic DNA using the primers 5'-GGAATTCGCATGCCTGGAGCTG GTGGCTACCG-3' and 5'-CCCCCTGCCGCAGGGATCAACACCCTGCACTTCCTCCTG-3' and cloned into the vector pMLG. The AtuR intergenic region was cloned with the primers 5'-CCCGAATTCTACCAAGCAAGCGCTTGGTTG-3' and 5'-CCCCCTAGGCGACCGAGCAA GCGCTTG-3'. The pMLG vector was constructed from the vector pMLGFP, which was kindly provided by Dr. Ashton Cropp (Virginia Commonwealth University), by using site-directed mutagenesis to introduce appropriate restriction sites.[210] *atuR* was cloned between *Nde*I and *Kpn*I sites. The intergenic region was cloned between the *Eco*RI and *Avr*II sites. Error-prone mutagenesis was carried out using the standard protocol for the GeneMorph II Random Mutagenesis Kit (Agilent Technologies) and the primers 5'-ATACATATGCTGGAGCTGGTGGC-3' and 5'-ATAGGTACCTCAGGCGTTC TTGATC-3'. The individual mutant a*tuR* genes were made using standard procedures for site-directed mutagenesis and the following primers: M76T, 5'-CTGCGCTCGGTGAC GGAAGAGACCATC-3' and 5'-GATGGTCTTCTTCCGTCACCGAGCGCAG-3'; M148L, 5'-GACATCTACGAGCAGCTGTGGCTCGACGTG-3' and 5'-CACGTCGAGCCACAGCTGCT CGTAGATGTC-3'; and F184S, 5'-CACCACCTGGTCCCGTCCGGAAG-3' and 5'-CTTCCG GACGGGACCAGGTGGTG-3'. Plasmids from site-directed mutagenesis were transformed into chemically-competent *E. coli* DH5α cells by standard methods. Error-prone *atuR* libraries were transformed electrocompetently into *E. cloni* 10G ELITE cells (Lucigen).

6.3.3. *FACS Screening*

Samples were analyzed and sorted on a MoFlo XDP cell sorter (Beckman Coulter, CA). Cultures were diluted 10-fold with LB media and passed through a 30 µm disposable Celtric filter (Partec, Germany) prior to FACS. GFP was excited using the 100 mW 488 nm laser. Cells were collected into 1 mL LB media and allowed to recover for 4 h at 37 °C before adding antibiotic and growing for 16 h.

6.3.4. *Plate Screening*

Individual colonies were selected from LB agar/ampicillin plates with sterile toothpicks and added to 500 µL wells for growth at 37 °C and 350 rpm for 3 h. 10 µL were taken from each well and used to inoculate 500 µL cultures with at least one of DMSO, 1-octanol, or one of the six natural terpene ligands (**1**-**6**) at concentrations ranging from 10 µM to 1 mM. At least one negative and one positive plate were grown overnight for 16 h at 37 °C and 350 rpm for each colony. Cells were pelleted and resuspended in 1 mL phosphate buffered saline (PBS) per well. 100 µL from each well was transferred to clear flat-bottom 96-well plates and another 100 µL to black flat-bottom 96-well plates (Greiner Bio-One). Cell density (600 nm) and GFP fluorescence (488 nm excitation, 515 nm emission) was measured on a BioTek Hybrid Synergy 4 plate reader. Fluorescence output was divided by OD_{600} for comparison purposes. Controls were run to ensure linear fluorescence response relative to OD_{600} at associated levels. At least one wild-type AtuR sample was run per 96-well plate for consistency.

6.3.5. *Construction of pSENSE Plasmids*

Two halves of the pSENSE vector were synthesized by Twist Bioscience (San Francisco, CA) and circularized together via Gibson Assembly (standard procedure).[155] The 3AA *atuO* binding site was prepared by phosphorylating and annealing complementary oligos and cloning between the *Afe*I and *Spe*I sites. The 4AA *atuO* binding site was introduced from 3AA using 'round-the-horn mutagenesis and the following 5' phosphorylated primers: 5'- AAATCAGCAAGGAG GTACTAGTATGAGCAAAGGAGAAG-3' and 5'- AGCCCCGCCTAGCAGTCACGTCGAGC AAGC-3'.[154] The *atuR_Opt* gene was cloned in between *Nde*I and *Kpn*I. The *sfGFP* gene was cloned in between *Spe*I and *Hin*dIII.

6.3.6. *Purification of AtuR*

The *atuR_Opt* gene was cloned into pET28-MHL using the restriction sites *Nde*I and *Hin*dIII and transformed into *E. coli* BL21 (DE3). The cells were grown in 300 mL of LB media with 50 μg/mL kanamycin at 37 °C and 250 rpm until OD_{600} reached 0.6, and then the culture was induced with 1mM IPTG and allowed to express at 22 °C for 20 h. AtuR_Opt was purified on a nickel column with a Bio-Rad BioLogic LP FPLC (Wash buffer: 50 mM Tris-HCl, pH 8, 300 mM NaCl, 20 mM imidazole, and 10% v/v glycerol; Elution buffer: 50 mM Tris-HCl, pH 8, 300 mM NaCl, 200 mM imidazole, and 10% v/v glycerol), concentrated with a 10 kDa MWCO filter (EMD Millipore, Burlington, MA), and stored as single-use aliquots in storage buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, and 10% v/v glycerol) at -20 °C.

6.3.7. *Electrophoretic Mobility Shift Assays*

The *atuO3x* oligos were annealed per standard procedure: 5'- GCCAAGCGCTTGCTCCAAGC GCTTGCTCCAAGCGCTTGCT-3' and 5'- AGCAAGCGCTTGGAGCAAGCGCTTGGAGCA AGCGCTTGGC-3'.

EMSA reactions were run in 20 μL volumes at 30 °C for 30 minutes with 1 pmol *atuO3x* and purified AtuR_Opt ranging from 0-10 pmol in reaction buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 5% v/v glycerol, 100 mM KCl, and 0.01 mg/mL BSA). Reactions were run out on a 6% PAGE gel at 220 V. The gel was stained with SYBR Green I and imaged on a Typhoon FLA 7000.

6.3.8. *Construction of AtuR Homology Models*

Both AtuR homology models were constructed using the I-TASSER server.[124-126] The AtuR_Old model was based on the PfiT crystal structure PDB 4MO7, and the AtuR_Opt model was based on the PfiT crystal structure PDB 4MXM.[205]

CHAPTER 7

Future Directions

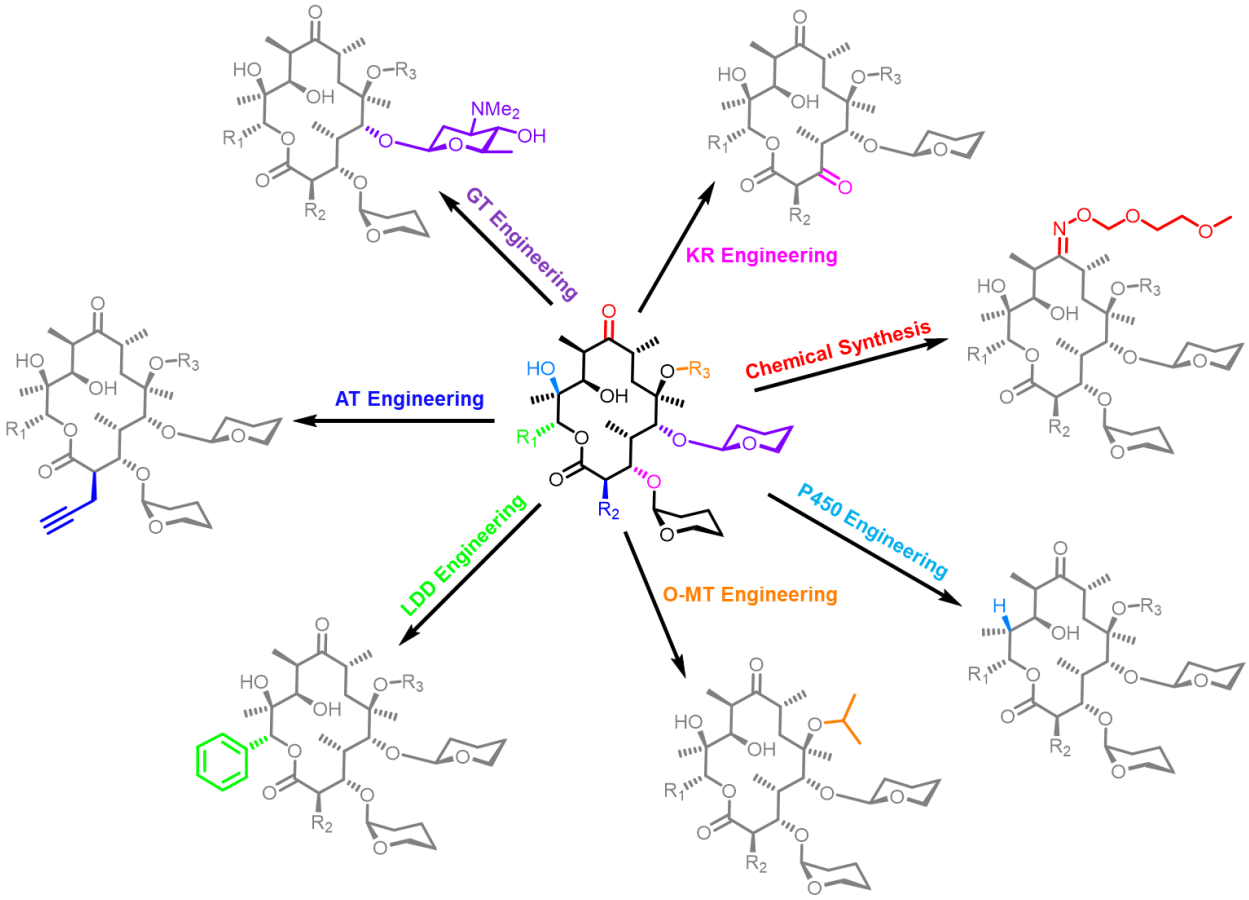


Figure 7.1. Polyketide Diversification. The described work focuses mainly on AT engineering in polyketide synthases; however, there are many other aspects of polyketide products that can be manipulated. A few of these approaches are shown. GT = glycosyltransferase; KR = ketoreductase; O-MT = O-methyltransferase; LDD = loading didomain; AT = acyltransferase.

7.1. Future Directions of Polyketide Synthase Acyltransferase Engineering

As discussed in Chapters 2-4, engineering PKS AT domains holds the potential to alter the substituent at every other carbon in the backbone of each polyketide natural product. The need for regioselectively controlling the biosynthesis of polyketides is highlighted particularly strongly by the example of switching a methyl to a hydrogen in the rifamycin backbone, yielding an antibiotic

potent against rifamycin-resistant tuberculosis—one of the great public health threats of our time.[211] To be able to make these modifications at will is the 'holy grail' of PKS engineering, but the field has not reached that point to date. However, the insight into substrate selectivity described in this dissertation has laid a foundation for further AT engineering. The work from Chapters 2-4 cover mutations in three AT domains: EryAT6, PikAT5, and PikAT6. These ATs are important for various reasons, but they do not begin to encompass all of the types of ATs or substrates found in nature. Just as mutations from EryAT6 were introduced into the Pik ATs, so too should the new mutations and selectivity-conferring motifs be moved into new and more diverse ATs to develop comprehensive rules. The field also has now accrued enough knowledge to begin working with clinically-relevant polyketides such as rifamycin, rapamycin, and FK506.

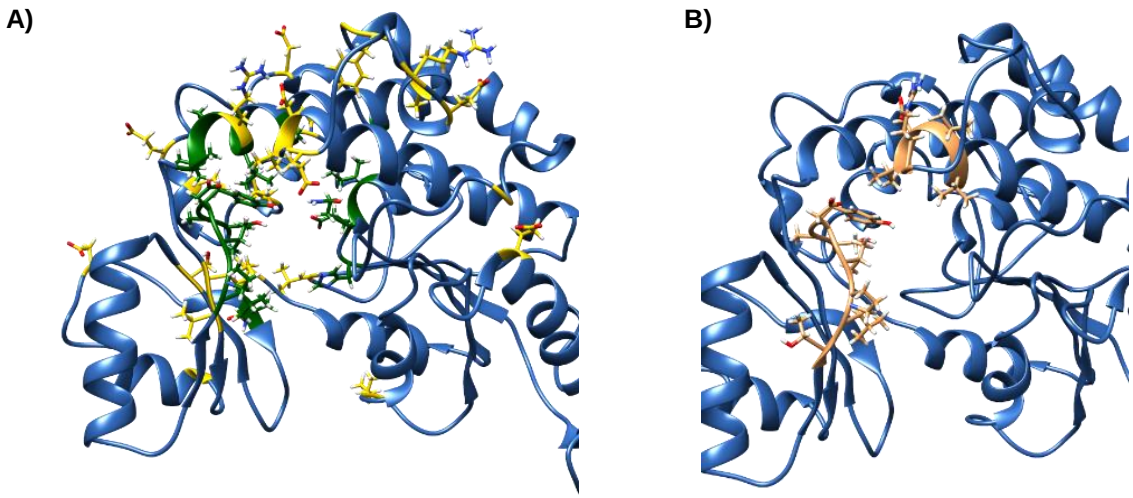


Figure 7.2. Current *vs* Proposed Mutations for EryAT6. A) Individual mutations have been made at each of the positions highlighted in green. More than two dozen additional mutations were designed for altering substrate selectivity at other residues highlighted in yellow. **B)** The AT substrate selectivity motifs (highlighted) need to be further refined regarding length, and additional motifs can be introduced from newly-discovered ATs for novel substrates.

While testing known mutations in other ATs is one important avenue, more can still be done with the model ATs probed in this dissertation: EryAT6, PikAT5, and PikAT6. Using MD simulations and sequence alignments as a guide, dozens of residues are potentially responsible for substrate selectivity (**Figure 7.2**). It is unlikely that all or even most of these residues would need to be changed to achieve a fully specific AT, but these are residues that should and need to be tested in the same context as the mutations described in Chapters 3 and 4. Furthermore, with the advent of genome mining, ATs with new substrates are being discovered at higher rates than at any other time before now.[212] These ATs should be exploited for their unique substrate scopes as well as utilizing their specificity motifs in more traditional ATs.

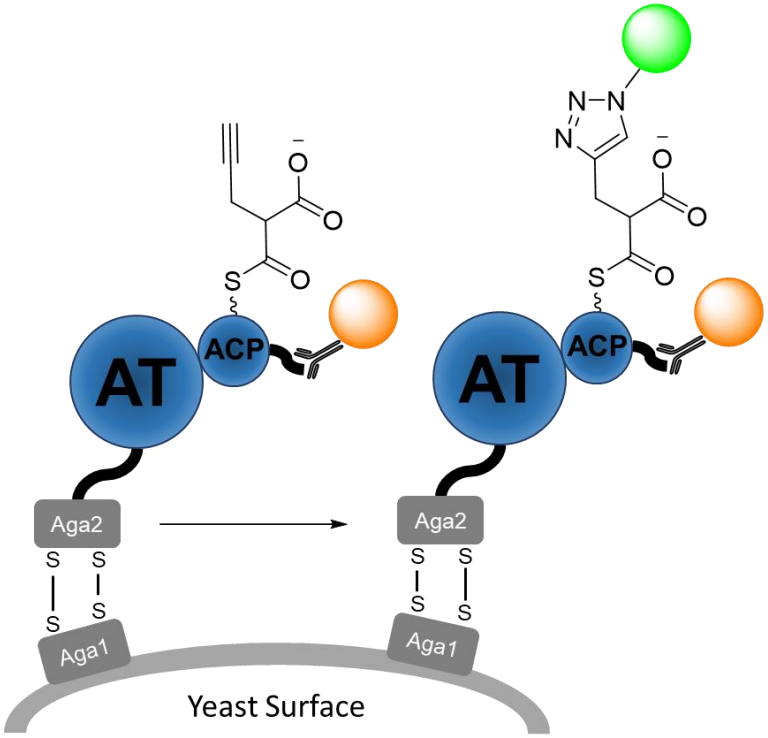


Figure 7.3. Yeast Cell Surface Display for PKSs. An AT-ACP didomain can be displayed on a yeast surface with a C-terminal *c-myc* tag. A fluorescently-tagged anti-*c-myc* antibody will be incubated to confirm expression of full construct. When the yeast is incubated with a propargyl- or azide-containing extender unit, the ACP linked to the accepting AT will be ligated to a fluorophore with a complementary click chemistry handle. The resulting cells can be sorted by flow cytometry. Competing substrates such as methylmalonyl-CoA can also be included.

Working with these ATs and their mutants is time-consuming and expensive work. To make the next leap in PKS engineering, tools for directed evolution are likely needed. One such method, the use of a genetically-encoded macrolide biosensor, is highlighted in **Figure 1.5**.[85] This method, while providing high throughput, requires the use of a full (and specific) PKS pathway and would also require additional engineering of the transcription factor to recognize each desired product. This is a worthwhile route, and much information can be gained rapidly through its use, but it should not be depended on as a sole strategy for would-be PKS engineers. An alternative method, inspired by the work of Dr. Chris Ladner (*unpublished data*) and a similar scheme for non-ribosomal peptide synthetases by the Hilvert group,[213] is to take advantage of yeast cell surface display technology (**Figure 7.3**). To utilize this technology to its highest potential, an AT-ACP didomain would be expressed and displayed on the yeast cell surface, flanked by a C-terminal *c-myc* tag. The intact didomain would be preferred over separate domains for several reasons: (1) a didomain would be encoded in a single gene, (2) an anti-*c-myc* antibody with a fluorescent tag could be used to detect cells containing fully-expressed and displayed didomains, (3) instead of having only a small percentage of ATs and ACPs in close proximity and in the correct orientation, every didomain can ostensibly catalyze a single reaction, and (4) working with a didomain (or even a tridomain with a KS) enables evolving active chimeras and prevents mutations that would disrupt the native covalent nature of the domains. The yeast cells are incubated with a desired concentration of a malonyl-CoA derivative containing a click handle, i.e. an alkyne or an azide. A small molecule fluorophore with the corresponding click handle is added after sufficient time for the AT to react is provided. Any ATs that successfully transferred the malonyl group from CoA to their cognate ACPs can then be identified by screening with fluorescence-activated cell sorting (FACS). Because this reaction occurs outside the cells, the only malonyl-CoAs present will be

those fed into the reaction—an advantage if trying to evolve a malonyl-CoA-specific AT, for example. Alternatively, competing substrates that lack the "click" handle can be fed in to compete with the desired substrate if specificity as opposed to promiscuity is desired. While this screen is limited initially to engineering ATs for substrates modified with a click handle, once an AT has been evolved to specifically recognize only a single substrate, this assay can be utilized to screen AT-ACP chimeras or even non-CoA- or non-ACP-linked extender units (as described in Section 7.2).

7.2. Synthetic Extender Units for Non-Natural Polyketide Production

AT engineering (Chapters 2-4) is an integral part of the work needed to achieve maximum polyketide diversification (**Figure 7.1**), but often, the non-native substrates are not easily accessed. The development of a directed evolution tool for engineering malonyl-CoA derivative production (Chapter 5) can help with this barrier; however, there may be cases in which there is no obvious starting point for biosynthesis, e.g., aminomalonyl-CoA (Samantha Carpenter, *unpublished results*). Additionally, there may be instances when enzyme evolution or genome engineering is cost- or time-prohibitive. In these cases, it may be necessary to chemically synthesize extender units. The most common synthetic extender units, derived from malonyl-*N*-acetylcysteamine (malonyl-SNAC, **Figure 7.4**, top right), have been used to great effect both *in vitro* and *in vivo*.[61-62, 143, 151, 214] In many studies published to date using a malonyl-SNAC, these synthetic substrates have either been considered to be taken up well by the cell and incorporated efficiently by a PKS,[214] taken up by the cell to an unknown degree,[62] or simply outcompeted by an intra-cellular malonyl-CoA derivative.[151] These observations highlight two potential obstacles to efficient *in vivo* use of

malonyl-SNACs: (1) their bioavailability may be limited due to the charged carboxylate and (2) the synthetic thioesters may not be recognized by AT domains as well as a CoA thioester.

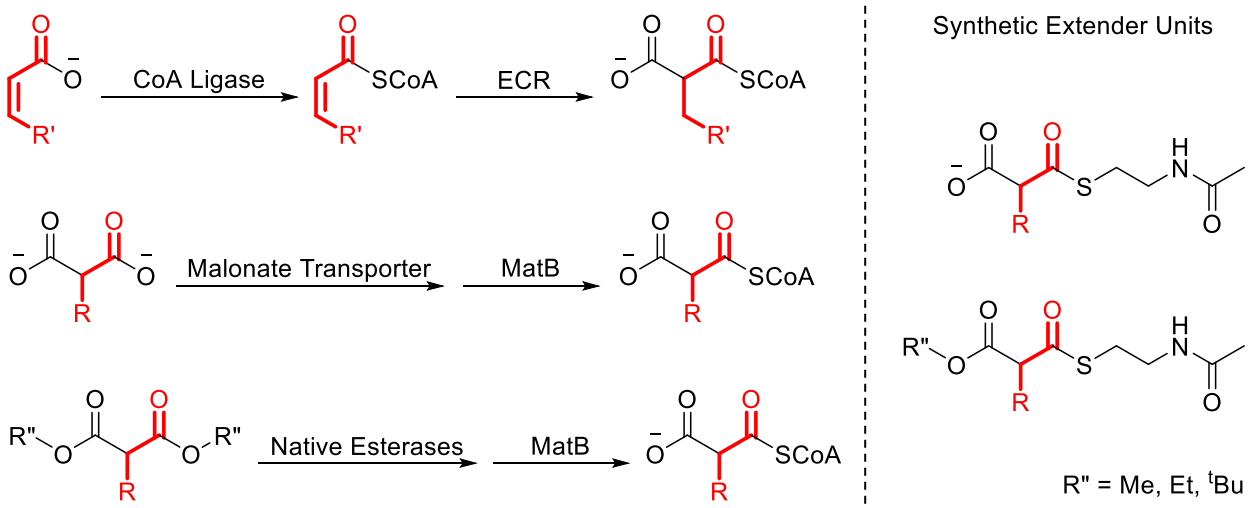


Figure 7.4. Current and Proposed Methods of Non-Native Polyketide Extender Unit Delivery. Left. Various methods for *in vivo* production of non-native extender units are shown. These include the use of enoyl-CoA reductases (ECRs) and MatB-like enzymes. In some cases, additional proteins are required for the substrates to enter the cell or be converted by a CoA ligase. **Right.** Malonyl-SNAC analogues can be fed in, but the charged carboxylate can limit their bioavailability. Malonyl-SNAC esters may be better compounds to feed into cultures.

To alter the malonyl-SNACs for passing through the membrane more readily, the malonyl carboxylate can be converted into an ester. These uncharged precursors should pass through the membrane more easily and with less need for media buffering. Once inside the cell, native esterases should yield the desired malonyl-SNAC substrate, much like with the malonyl diesters used in conjunction with MatB-like enzymes (**Figure 7.4**). Notably, the synthetic route to malonyl-SNAC derivatives goes through a *tert*-butyl ester intermediate, but the Schulz group has demonstrated that a *tert*-butyl methylmalonyl-SNAC is not easily cleaved enzymatically.[214] Instead, a methyl or ethyl ester is likely a better choice to increase intracellular concentrations of

synthetic substrate. This increase could help overcome any K_m differences between the NAC and CoA thioesters as well.

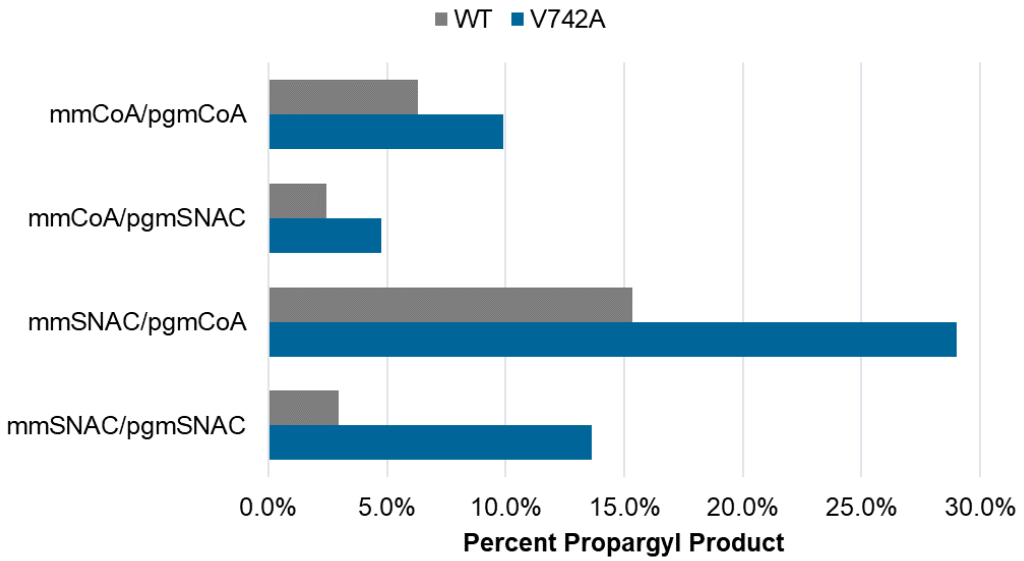


Figure 7.5. *In Vitro* Substrate Competition Assay for EryAT6 Variants. Malonyl-CoA and malonyl-SNAC derivatives (mm = methylmalonyl; pgm = propargylmalonyl) were fed in at equal concentrations with Ery6TE wild-type and V742A. The CoA thioesters were incorporated at higher rates than their SNAC counterparts. The presence of the V742A mutation in EryAT6 also contributed to a difference in discrimination between CoA and SNAC thioesters.

Just as increasing the intracellular concentration of synthetic extender units could lead to higher levels of incorporation, so too could leveraging any potential K_m differences between CoA and NAC thioesters. The SNAC moiety is only a small fraction of the entire CoA structure, and it lacks many of the functional groups responsible for substrate binding, e.g., the phosphates. Because it is missing these interactions, the substrate is likely more flexible in its enzyme-bound conformation. This added flexibility may enable new substrate conformations that are better recognized by certain natural or mutant ATs. Additionally, the tighter binding of CoA compared to SNAC can lead to improvement of non-natural extender unit incorporation if leveraged appropriately. To test both of

these hypotheses, equal concentrations of two of the following substrates were incubated *in vitro* with either wild-type or the V742A variants of Ery6TE: methylmalonyl-CoA (mmCoA), methylmalonyl-SNAC (mmSNAC), propargylmalonyl-CoA (pgmCoA), or propargylmalonyl-SNAC (pgmSNAC) (**Figure 7.5**). The malonyl-CoA derivatives were produced by *in vitro* MatB reactions, and the malonyl-SNAC derivatives were synthesized by Anuran Kumar Gayen. With the wild-type enzyme and both CoA derivatives, the V742A variant incorporated ~1.6x more propargyl product than did wild-type. With mmCoA and pgmSNAC, the percent propargyl predictably dropped for both modules. When providing the non-natural pgmCoA against mmSNAC, propargyl incorporation rates increased ~3-fold for both enzymes. In both cases with one CoA and one SNAC, the V742A mutant incorporated ~1.9x more propargyl than did wild-type, a small increase over both substrates provided as CoA thioesters. However, when both substrates were provided as the smaller SNACs, the V742A mutant incorporated 4.6x more propargyl than did the wild-type construct. This effect could be explained due to the added flexibility of the YASH loop in the V742A mutant and the increased reliance on any binding interactions by the propargyl moiety when missing the CoA interactions. These preliminary results set a foundation for co-utilization of CoA- and SNAC-linked extender units both *in vitro* and *in vivo* for the derivatization of polyketide products.

7.3. Construction of a Ketolide Precursor *In Vivo*

While AT engineering is the focus of much of this dissertation, it is certainly not the only means of altering the structure of polyketides (**Figure 7.1**). For example, the newest clinically-relevant analogues of erythromycin A are monoglycosylated macrolides called ketolides (**Figure 7.6**). However, each of these semi-synthetic analogues requires 9 or more synthetic steps, increasing

cost and limiting the ease of chemical space exploration during drug discovery.[215] Ideally, access to a biosynthetically-derived intermediate for the ketolide class would be available.

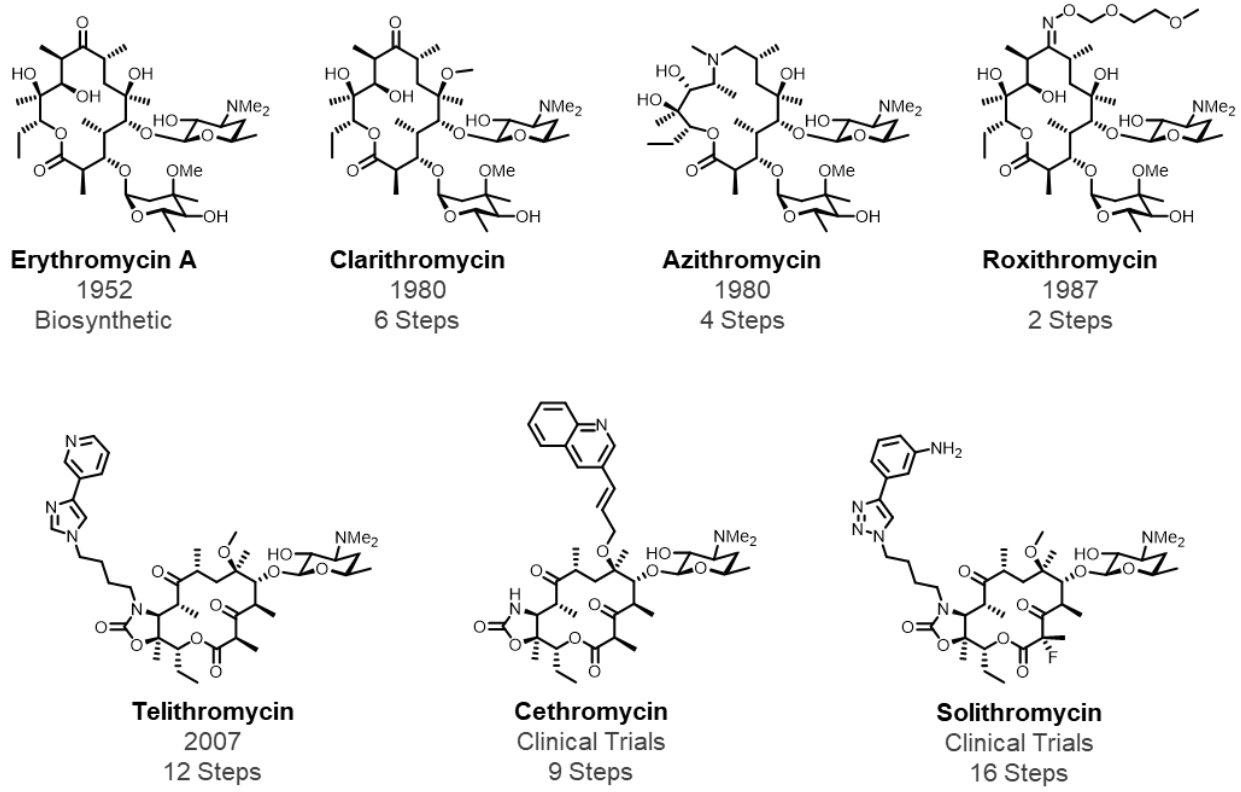


Figure 7.6. Clinically-Relevant Macrolide and Ketolide Antibiotics. Erythromycin A, a potent polyketide antibiotic, is biosynthesized by a PKS found in *Saccharopolyspora erythraea*. Since its discovery in 1952, three semisynthetic macrolides with improved pharmacological properties (**Top Row**) have been primarily utilized in the clinic; however, twenty years passed before the approval of the first ketolide antibiotic, telithromycin. The ketolides (**Bottom Row**) are a promising new class of ribosome-targeting antibiotics, but they are among the most expensive drugs to produce due to the large number of synthetic steps from erythromycin A.

Taking solithromycin, the first clinical fluoroketolide, as a proof-of-principle case, a retrosynthetic analysis provided a biologically-feasible intermediate requiring a number of changes to the erythromycin biosynthetic gene cluster (**Figure 7.7**). Within the DEBS PKS, a dehydratase (DH) domain or DH-containing module would need to be introduced in place of module 2, thus yielding an alkenyl handle for introduction of the carbamate moiety. Currently, the proposed module is

module 4 of the rapamycin pathway due to the conservation of the intermodular linkers from that pathway and its place as an early-stage module at the C-terminus of a polypeptide. Additionally, for introduction of the gem-fluoro-methyl group, the first module of the hectochlorin PKS (HctD) is proposed to be able to catalyze this chemistry. In this case, the malonyl-CoA-specific AT would accept fluoromalonyl-CoA before adding the methyl stereospecifically using the C-methyltransferase (CM) domain. The functional ketoreductase (KR) domain can be inactivated through a single Tyr→Phe mutation.[216] Challenges remaining with this approach involve the

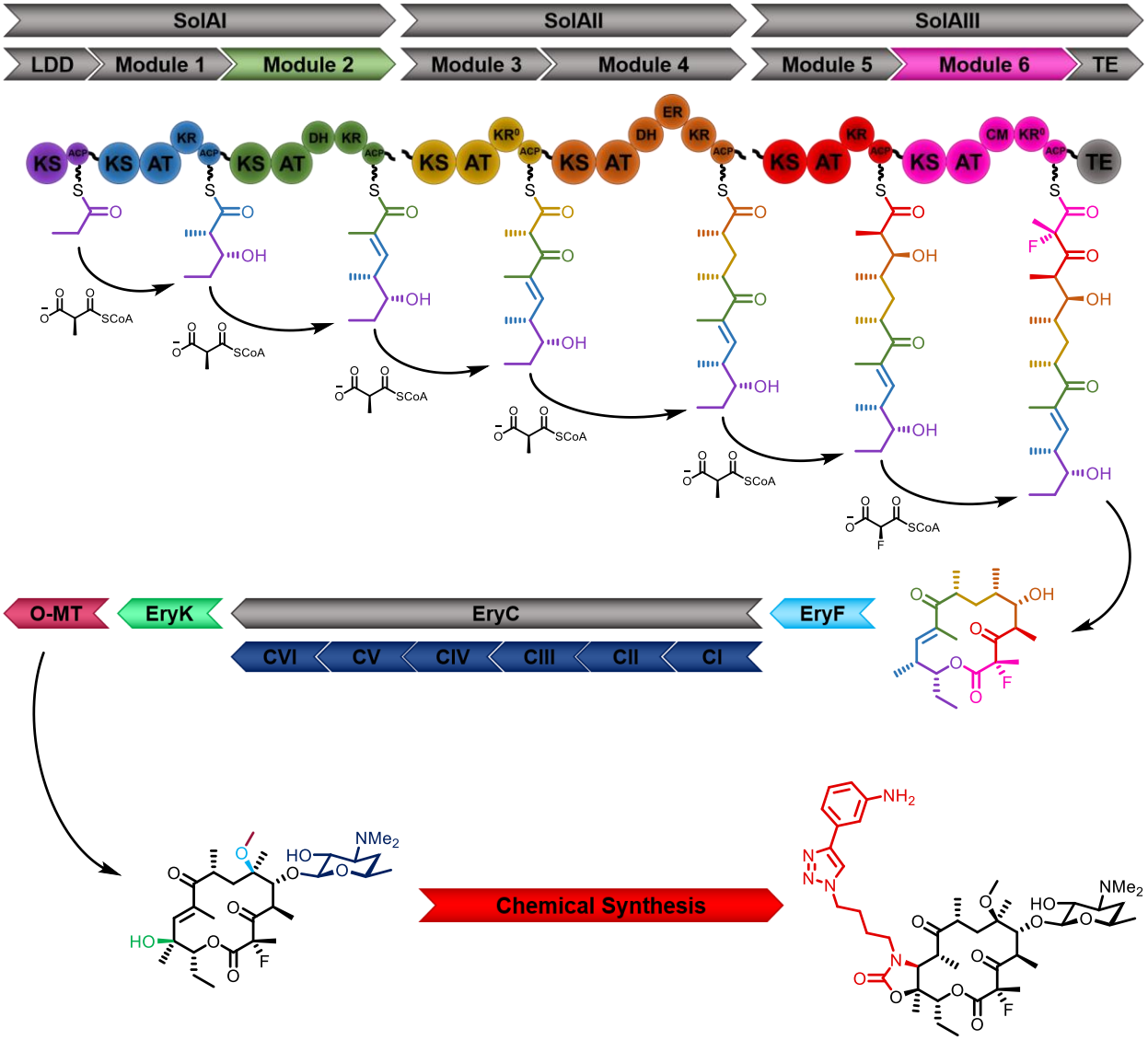


Figure 7.7. Proposed Novel Pathway for Production of Solithromycin. In *S. erythraea*, the core of erythromycin A is produced by six PKS assembly line-like modules. We propose to replace the second and sixth modules with modified modules from rapamycin and hectochlorin biosynthesis. Post-PKS enzymes, including an engineered O-methyltransferase (O-MT), provide a biosynthetic route to a ketolide precursor for semisynthesis.

questionable preference of fluoromalonyl-CoA over endogenous malonyl-CoA and the lack of shared homology between DEBS ketosynthase (KS) domains and HctD's KS. Downstream of the PKS, several genes could be eliminated; however, a single *O*-methyltransferase would need to be engineered to perform the non-natural methylation of the C6-hydroxyl (Yiwei Li, *in progress*). A potential starting scaffold for this enzyme would be StfMIII from the steffimycin biosynthetic gene

cluster due to the similarity of the native and proposed substrates.[217] Once a functional pathway has been optimized for production, the biosynthetically-derived fluoroketolide could be used as a building block for further enzymatic or chemical derivatization for the production of known or novel antibiotics.

7.4. Directed Evolution of MatB, Other CoA Ligases, and FapR

Following the characterization and development of a malonyl-CoA FapR biosensor from Chapter 5, the biosensor can be utilized as a directed evolution tool for CoA ligases. Of the various CoA ligases responsible for production of malonyl-CoA derivatives (**Figure 5.1**), MatB and MatB-like enzymes are likely the best candidate for *in vivo* use to make non-natural malonyl-CoAs. In natural product gene clusters, most non-(methyl)malonyl-CoAs are synthesized by enoyl-CoA reductases (ECRs), often requiring two or more enzymes to turn over a non-natural substrate.[19] While the throughput of a genetically-encoded fluorescent biosensor can account for a multi-enzyme pathway, MatB still remains the top choice when feasible due to a lower metabolic burden, a smaller library size, and plentiful commercial substrates. With the current biosensor plasmid, fluorescence-activated cell sorting (FACS) will be used to sort cells containing individual members of an error-prone library of MatB (or a DNA-shuffled library of MatB-like enzymes). Negative sorting can be done in the absence of a fed-in malonyl diester, and subsequent positive sorts can include decreasing concentrations of the corresponding malonyl diesters. By screening for high production initially, high k_{cat} mutants can be pulled out before lowering the substrate concentrations to find improved K_m mutants.

To increase the throughput and further reduce the cost of screening libraries, the *sfGFP* reporter can be coupled with a second fusion gene, *cat-upp*. The *cat-upp* gene is a fusion gene between a chloramphenicol acetyltransferase (*cat*) and a uracil phosphoribosyltransferase (*upp*) gene, and it has been previously used as part of a growth selection for tRNA synthetase engineering.[218] CAT provides resistance to chloramphenicol when expressed, and UPP is lethal to the cell in the presence of 5-fluorouracil (5-FU). By including both reporter genes, the newly-constructed biosensor will have both a semi-quantitative readout (sfGFP fluorescence) and a dual positive/negative growth selection (CAT-UPP expression in the presence of a given compound). This plasmid could be particularly useful when evolving FapR to more specifically recognize a desired malonyl-CoA over others. One could even imagine using this system to evolve FapR to recognize just CoA rather than a malonyl-CoA. The higher throughput of a growth selection also allows for more thorough coverage of large or high-error rate libraries, especially for multi-enzyme systems like ECRs or more unique pathways like the allylmalonyl-CoA cluster from *Streptomyces hygroscopicus*.[159]

7.5. Coupling of AtuR Biosensor with IPK, IspA, and GES

Given the high potential of a genetically-encoded biosensor for acyclic terpenes (**Figure 6.1**), AtuR should be used as more than a study of a biosensor construct. Rather, the AtuR biosensor should be coupled with multiple enzymes or pathways for directed evolution. As AtuR recognizes several acyclic monoterpenes (**Figure 6.3**), it can be utilized for the production of terpene precursors IPP and DMAPP, the monoterpene precursor geranyl pyrophosphate (GPP), or the dephosphorylated acyclic monoterpenes. For this proposed directed evolution system, pSENSE4AA (or a similar construct) would replace *mCherry* with isopentenyl phosphate kinase

(*IPK*) for constitutive expression. IPK has been reported to catalyze the conversion of dimethylallyl alcohol (DMAA) to DMAPP (and subsequent isomerization by *E. coli* Idi provides IPP).[219] A second plasmid such as pCDFDuet-1 or pACYCDuet-1 could house a GPP synthase gene, *IspA S81F* (a mutant selective for C_{10} products),[220] and a geraniol synthase, tVoGES (a truncated version of GES from *Valeriana officinalis*) (**Figure 7.8**).[221] Libraries of IPK could be screened in the presence of the substrate DMAA and an inhibitor of the natural pathway, fosmidomycin, for optimization or engineering of a novel terpene precursor pathway.[222]

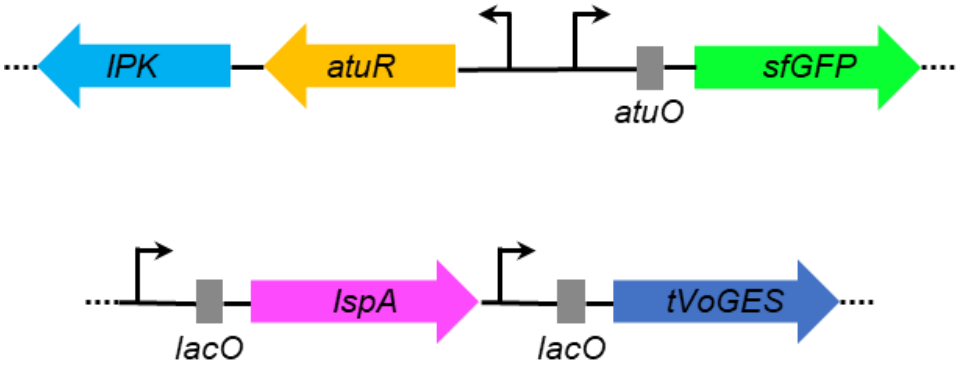


Figure 7.8. Use of AtuR as a Directed Evolution Tool. A modular two-plasmid system for evolving natural or non-natural terpene production in *E. coli*.

Additionally, as described by Dr. Sean Lund and Rachael Hall (*manuscript in prep.*), IspA is capable of linking DMAPP with various short-chain alcohols to yield GPP analogues. To date, the kinetics for many of these substrates is sub-optimal, and none of the tested extended analogues have been turned over by a terpene synthase. As GES is the simplest transformation for GPP, it is a good first step for evolving an enzyme that can utilize these non-natural analogues. Provided AtuR's promiscuity towards recognizing at least six acyclic terpenes, it is possible that these new analogues would also activate the AtuR biosensor, enabling directed evolution. To avoid direct

competition with any natural geraniol produced by the endogenous DMAPP and IPP, an *in vitro* screen could be used with microfluidics and purified transcription-translation components.

REFERENCES

1. Wang, H.; Fewer, D. P.; Holm, L.; Rouhiainen, L.; Sivonen, K., Atlas of Nonribosomal Peptide and Polyketide Biosynthetic Pathways Reveals Common Occurrence of Nonmodular Enzymes. *Proc. Natl. Acad. Sci. USA* **2014,** *111* (25), 9259-9264.
2. Weissman, K. J., Chapter 1 Introduction to Polyketide Biosynthesis. In *Methods Enzymol.*, Academic Press: 2009; Vol. 459, pp 3-16.
3. Kennedy, J., Mutasynthesis, Chemobiosynthesis, and Back to Semi-Synthesis: Combining Synthetic Chemistry and Biosynthetic Engineering for Diversifying Natural Products. *Nat Prod Rep* **2008,** *25* (1), 25-34.
4. Klein, A. S.; Domröse, A.; Bongen, P.; Brass, H. U. C.; Classen, T.; Loeschcke, A.; Drepper, T.; Laraia, L.; Sievers, S.; Jaeger, K.-E.; Pietruszka, J., New Prodigiosin Derivatives Obtained by Mutasynthesis in Pseudomonas Putida. *ACS Synth Biol* **2017,** *6* (9), 1757-1765.
5. Chen, L.; Li, Y.; Yue, Q.; Loksztejn, A.; Yokoyama, K.; Felix, E. A.; Liu, X.; Zhang, N.; An, Z.; Bills, G. F., Engineering of New Pneumocandin Side-Chain Analogues from Glarea Lozoyensis by Mutasynthesis and Evaluation of Their Antifungal Activity. *ACS Chem. Biol.* **2016,** *11* (10), 2724-2733.
6. Jacobsen, J. R.; Hutchinson, C. R.; Cane, D. E.; Khosla, C., Precursor-Directed Biosynthesis of Erythromycin Analogs by an Engineered Polyketide Synthase. *Science* **1997,** *277* (5324), 367-369.
7. Kinoshita, K.; Pfeifer, B. A.; Khosla, C.; Cane, D. E., Precursor-Directed Polyketide Biosynthesis in Escherichia Coli. *Bioorganic & medicinal chemistry letters* **2003,** *13* (21), 3701-3704.

8. Ward, S. L.; Desai, R. P.; Hu, Z.; Gramajo, H.; Katz, L., Precursor-Directed Biosynthesis of 6-Deoxyerythronolide B Analogues Is Improved by Removal of the Initial Catalytic Sites of the Polyketide Synthase. *J Ind Microbiol Biotechnol* **2007,** *34* (1), 9-15.

9. Murli, S.; MacMillan, K. S.; Hu, Z.; Ashley, G. W.; Dong, S. D.; Kealey, J. T.; Reeves, C. D.; Kennedy, J., Chemobiosynthesis of Novel 6-Deoxyerythronolide B Analogues by Mutation of the Loading Module of 6-Deoxyerythronolide B Synthase 1. *Appl Environ Microbiol* **2005,** *71* (8), 4503-4509.

10. Wright, P. M.; Seiple, I. B.; Myers, A. G., The Evolving Role of Chemical Synthesis in Antibacterial Drug Discovery. *Angewandte Chemie (International ed. in English)* **2014,** *53* (34), 8840-8869.

11. Cane, D. E.; Walsh, C. T., The Parallel and Convergent Universes of Polyketide Synthases and Nonribosomal Peptide Synthetases. *Chem Biol* **1999,** *6* (12), R319-325.

12. Cane, D. E.; Walsh, C. T.; Khosla, C., Harnessing the Biosynthetic Code: Combinations, Permutations, and Mutations. *Science* **1998,** *282* (5386), 63-68.

13. Newman, D. J.; M., C. G., Natural Products as Sources of New Drugs from 1981 to 2014. *Journal of Natural Products* **2016,** *79*, 629-661.

14. Cragg, G. M.; Newman, D. J., Natural Products: A Continuing Source of Novel Drug Leads. *Biochim. Biophys. Acta* **2013,** *1830* (6), 3670-3695.

15. Maier, M. E., Design and Synthesis of Analogues of Natural Products. *Org. Biomol. Chem.* **2015,** *13* (19), 5302-5343.

16. Li, G.; Wang, J.; Reetz, M. T., Biocatalysts for the Pharmaceutical Industry Created by Structure-Guided Directed Evolution of Stereoselective Enzymes. *Bioorganic & Medicinal Chemistry* **2017**.

17. Cortes, J.; Haydock, S. F.; Roberts, G. A.; Bevitt, D. J.; Leadlay, P. F., An Unusually Large Multifunctional Polypeptide in the Erythromycin-Producing Polyketide Synthase of Saccharopolyspora Erythraea. *Nature* **1990,** *348* (6297), 176-178.

18. Weissman, K. J.; Leadlay, P. F., Combinatorial Biosynthesis of Reduced Polyketides. *Nature Reviews Microbiology* **2005,** *3* (12), 925-936.

19. Chan, Y. A.; Podevels, A. M.; Kevany, B. M.; Thomas, M. G., Biosynthesis of Polyketide Synthase Extender Units. *Nat. Prod. Rep.* **2009,** *26* (1), 90-114.

20. Hopwood, D. A.; Malpartida, F.; Kieser, H. M.; Ikeda, H.; Duncan, J.; Fujii, I.; Rudd, B. A.; Floss, H. G.; Ōmura, S., Production of 'Hybrid' Antibiotics by Genetic Engineering. *Nature* **1985,** *314* (6012), 642--644.

21. Chen, T. S.; Inamine, E. S.; Hensens, O. D.; Zink, D.; Ostlind, D. A., Directed Biosynthesis of Avermectins. *Arch. Biochem. Biophys.* **1989,** *269* (2), 544-547.

22. Dutta, S.; Whicher, J. R.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Congdon, G. R.; Narayan, A. R.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G., Structure of a Modular Polyketide Synthase. *Nature* **2014,** *510* (7506), 512-517.

23. Whicher, J. R.; Dutta, S.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Dosey, A. M.; Narayan, A. R. H.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G., Structural Rearrangements of a Polyketide Synthase Module During Its Catalytic Cycle. *Nature* **2014,** *510* (7506), 560-564.

24. Mathews, I. I.; Allison, K.; Robbins, T.; Lyubimov, A. Y.; Uervirojnangkoorn, M.; Brunger, A. T.; Khosla, C.; DeMirci, H.; McPhillips, S. E.; Hollenbeck, M.; Soltis, M.; Cohen, A. E., The Conformational Flexibility of the Acyltransferase from the Disorazole Polyketide Synthase Is Revealed by an X-Ray Free-Electron Laser Using a Room-

Temperature Sample Delivery Method for Serial Crystallography. *Biochemistry* **2017,** *56* (36), 4751-4756.

25. Menzella, H. G.; Carney, J. R.; Santi, D. V., Rational Design and Assembly of Synthetic Trimodular Polyketide Synthases. *Chem. Biol.* **2007,** *14* (2), 143-151.

26. Menzella, H. G.; Reid, R.; Carney, J. R.; Chandran, S. S.; Reisinger, S. J.; Patel, K. G.; Hopwood, D. A.; Santi, D. V., Combinatorial Polyketide Biosynthesis by De Novo Design and Rearrangement of Modular Polyketide Synthase Genes. *Nat Biotech* **2005,** *23* (9), 1171-1176.

27. Hansen, D. A.; Koch, A. A.; Sherman, D. H., Identification of a Thioesterase Bottleneck in the Pikromycin Pathway through Full-Module Processing of Unnatural Pentaketides. *J Am Chem Soc* **2017,** *139* (38), 13450-13455.

28. Kakule, T. B.; Lin, Z.; Schmidt, E. W., Combinatorialization of Fungal Polyketide Synthase–Peptide Synthetase Hybrid Proteins. *J. Am. Chem. Soc.* **2014,** *136* (51), 17882-17890.

29. Murphy, A. C.; Hong, H.; Vance, S.; Broadhurst, R. W.; Leadlay, P. F., Broadening Substrate Specificity of a Chain-Extending Ketosynthase through a Single Active-Site Mutation. *Chemical Communications* **2016,** *52* (54), 8373-8376.

30. Donadio, S.; Staver, M. J.; McAlpine, J. B.; Swanson, S. J.; Katz, L., Modular Organization of Genes Required for Complex Polyketide Biosynthesis. *Science* **1991,** *252* (5006), 675-679.

31. Donadio, S.; Katz, L., Organization of the Enzymatic Domains in the Multifunctional Polyketide Synthase Involved in Erythromycin Formation in Saccharopolyspora Erythraea. **1992,** *111* (1), 51–60.

32. Donadio, S.; McAlpine, J. B.; Sheldon, P. J.; Jackson, M.; Katz, L., An Erythromycin Analog Produced by Reprogramming of Polyketide Synthesis. *Proc Natl Acad Sci U S A* **1993,** *90* (15), 7119-7123.

33. Cortes, J.; Wiesmann, K.; Roberts, G.; Brown, M.; Staunton, J.; Leadlay, P., Repositioning of a Domain in a Modular Polyketide Synthase to Promote Specific Chain Cleavage. *Science* **1995,** *268* (5216), 1487-1489.

34. Kao, C. M.; Katz, L.; Khosla, C., Engineered Biosynthesis of a Complete Macrolactone in a Heterologous Host. *Science* **1994,** *265* (5171), 509-512.

35. Wiesmann, K. E.; Cortes, J.; Brown, M. J.; Cutter, A. L.; Staunton, J.; Leadlay, P. F., Polyketide Synthesis in Vitro on a Modular Polyketide Synthase. *Chem Biol* **1995,** *2* (9), 583-589.

36. Pieper, R.; Luo, G.; Cane, D. E.; Khosla, C., Cell-Free Synthesis of Polyketides by Recombinant Erythromycin Polyketide Synthases. *Nature* **1995,** *378* (6554), 263-266.

37. Kao, C. M.; Luo, G.; Katz, L.; Cane, D. E.; Khosla, C., Manipulation of Macrolide Ring Size by Directed Mutagenesis of a Modular Polyketide Synthase. *J. Am. Chem. Soc.* **1995,** *117* (35), 9105-9106.

38. Kao, C. M.; Luo, G.; Katz, L.; Cane, D. E.; Khosla, C., Engineered Biosynthesis of Structurally Diverse Tetraketides by a Trimodular Polyketide Synthase. *J. Am. Chem. Soc.* **1996,** *118* (38), 9184-9185.

39. Marsden, A. F.; Barrie, W.; Cortés, J.; Dunster, N. J.; Staunton, J.; Leadlay, P. F., Engineering Broader Specificity into an Antibiotic-Producing Polyketide Synthase. *Science* **1998,** *279* (5348), 199-202.

40. Rowe, C. J.; Böhm, I. U.; Thomas, I. P.; Wilkinson, B.; Rudd, B. A. M.; Foster, G.; Blackaby, A. P.; Sidebottom, P. J.; Roddis, Y.; Buss, A. D.; Staunton, J.; Leadlay, P. F., Engineering a Polyketide with a Longer Chain by Insertion of an Extra Module into the Erythromycin-Producing Polyketide Synthase. *Chem. Biol.* **2001,** *8* (5), 475-485.

41. Ranganathan, A.; Timoney, M.; Bycroft, M.; Jesús, C.; Thomas, I. P.; Wilkinson, B.; Kellenberger, L.; Hanefeld, U.; Galloway, I. S.; Staunton, J.; Leadlay, P. F., Knowledge-Based Design of Bimodular and Trimodular Polyketide Synthases Based on Domain and Module Swaps: A Route to Simple Statin Analogues. *Chem. Biol.* **1999,** *6* (10), 731-741.

42. Menzella, H. G.; Reid, R.; Carney, J. R.; Chandran, S. S.; Reisinger, S. J.; Patel, K. G.; Hopwood, D. a.; Santi, D. V., Combinatorial Polyketide Biosynthesis by De Novo Design and Rearrangement of Modular Polyketide Synthase Genes. *Nature Biotechnology* **2005,** *23* (9), 1171-1176.

43. Chandran, S. S.; Menzella, H. G.; Carney, J. R.; Santi, D. V., Activating Hybrid Modular Interfaces in Synthetic Polyketide Synthases by Cassette Replacement of Ketosynthase Domains. *Chem. Biol.* **2006,** *13* (5), 469-474.

44. Oliynyk, M.; Brown, M. J.; Cortes, J.; Staunton, J.; Leadlay, P. F., A Hybrid Modular Polyketide Synthase Obtained by Domain Swapping. *Chem Biol* **1996,** *3* (10), 833-839.

45. Ruan, X.; Pereda, A.; Stassi, D. L.; Zeidner, D.; Summers, R. G.; Jackson, M.; Shivakumar, A.; Kakavas, S.; Staver, M. J.; Donadio, S.; Katz, L., Acyltransferase Domain Substitutions in Erythromycin Polyketide Synthase Yield Novel Erythromycin Derivatives. *J. Bacteriol.* **1997,** *179* (20), 6416-6425.

46. Stassi, D. L.; Kakavas, S. J.; Reynolds, K. A.; Gunawardana, G.; Swanson, S.; Zeidner, D.; Jackson, M.; Liu, H.; Buko, A.; Katz, L., Ethyl-Substituted Erythromycin Derivatives

Produced by Directed Metabolic Engineering. *Proc. Natl. Acad. Sci. USA* **1998,** *95* (13), 7305-7309.

47. Zhang, J.; Yan, Y. J.; An, J.; Huang, S. X.; Wang, X. J.; Xiang, W. S., Designed Biosynthesis of 25-Methyl and 25-Ethyl Ivermectin with Enhanced Insecticidal Activity by Domain Swap of Avermectin Polyketide Synthase. *Microbial cell factories* **2015,** *14* (152).

48. Bedford, D.; Jacobsen, J. R.; Luo, G.; Cane, D. E.; Khosla, C., A Functional Chimeric Modular Polyketide Synthase Generated Via Domain Replacement. *Chem Biol* **1996,** *3* (10), 827-831.

49. McDaniel, R.; Kao, C. M.; Fu, H.; Hevezi, P.; Claes Gustafsson; Betlach, M.; Ashley, G.; Cane, D. E.; Khosla, C., Gain-of-Function Mutagenesis of a Modular Polyketide Synthase. *J. Am. Chem. Soc.* **1997,** *119* (18), 4309-4310.

50. Donadio, S.; Staver, M. J.; McAlpine, J. B.; Swanson, S. J.; Katz, L., Biosynthesis of the Erythromycin Macrolactone and a Rational Approach for Producing Hybrid Macrolides. *Gene* **1992,** *115* (1-2), 97-103.

51. McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Betlach, M.; Ashley, G., Multiple Genetic Modifications of the Erythromycin Polyketide Synthase to Produce a Library of Novel "Unnatural" Natural Products. *Proc. Natl. Acad. Sci. USA* **1999,** *96* (5), 1846-1851.

52. Hans, M.; Hornung, A.; Dziarnowski, A.; Cane, D. E.; Khosla, C., Mechanistic Analysis of Acyl Transferase Domain Exchange in Polyketide Synthase Modules. *Journal of the American Chemical Society* **2003,** *125*, 5366-5374.

53. Chemler, J. A.; Tripathi, A.; Hansen, D. A.; O'Neil-Johnson, M.; Williams, R. B.; Starks, C.; Park, S. R.; Sherman, D. H., Evolution of Efficient Modular Polyketide Synthases by Homologous Recombination. *Journal of the American Chemical Society* **2015,** *137* (33), 10603--10609.

54. Klaus, M.; Ostrowski, M. P.; Austerjost, J.; Robbins, T.; Lowry, B.; Cane, D. E.; Khosla, C., Protein-Protein Interactions, Not Substrate Recognition, Dominate the Turnover of Chimeric Assembly Line Polyketide Synthases. *Journal of Biological Chemistry* **2016,** *291* (31), 16404-16415.

55. Klaus, M.; Ostrowski, M. P.; Austerjost, J.; Robbins, T.; Lowry, B.; Cane, D. E.; Khosla, C., Protein Protein Interactions, Not Substrate Recognition, Dominates the Turnover of Chimeric Assembly Line Polyketide Synthases. *The Journal of Biological Chemistry* **2016,** *291* (31), 16404-16415.

56. Yuzawa, S.; Deng, K.; Wang, G.; Baidoo, E. E.; Northen, T. R.; Adams, P. D.; Katz, L.; Keasling, J. D., Comprehensive in Vitro Analysis of Acyltransferase Domain Exchanges in Modular Polyketide Synthases and Its Application for Short-Chain Ketone Production. *ACS Synth. Biol.* **2017,** *6* (1), 139-147.

57. Reeves, C. D.; Sumati, M.; Ashley, G. W.; Piagentini, M.; Hutchinson, C. R.; McDaniel, R., Alteration of the Substrate Specificity of a Modular Polyketide Synthase Acyltransferase Domain through Site-Specific Mutations. *Biochemistry* **2001,** *40* (51), 15464-15470.

58. Koch, A. A.; Hansen, D. A.; Shende, V. V.; Furan, L. R.; Houk, K. N.; Jimenez-Oses, G.; Sherman, D. H., A Single Active Site Mutation in the Pikromycin Thioesterase Generates a More Effective Macrocyclization Catalyst. *J Am Chem Soc* **2017,** *139* (38), 13456-13465.

59. Ad, O.; Thuronyi, B. W.; Chang, M. C. Y., Elucidating the Mechanism of Fluorinated Extender Unit Loading for Improved Production of Fluorine-Containing Polyketides. *Proc. Natl. Acad. Sci. USA* **2017,** *114* (5), E660-E668.

60. Musiol-Kroll, E. M.; Zubeil, F.; Schafhauser, T.; Hartner, T.; Kulik, A.; McArthur, J.; Koryakina, I.; Wohlleben, W.; Grond, S.; Williams, G. J.; Lee, S. Y.; Weber, T., Polyketide Bioderivatization Using the Promiscuous Acyltransferase Kircii. *ACS Synth. Biol.* **2017,** *6* (3), 421-427.

61. Bravo-Rodriguez, K.; Klopries, S.; Koopmans, K. R. M.; Sundermann, U.; Yahiaoui, S.; Arens, J.; Kushnir, S.; Schulz, F.; Sanchez-Garcia, E., Substrate Flexibility of a Mutated Acyltransferase Domain and Implications for Polyketide Biosynthesis. *Chem. Biol.* **2015,** *22* (11), 1425-1430.

62. Sundermann, U.; Bravo-Rodriguez, K.; Klopries, S.; Kushnir, S.; Gomez, H.; Sanchez-Garcia, E.; Schulz, F., Enzyme-Directed Mutasynthesis: A Combined Experimental and Theoretical Approach to Substrate Recognition of a Polyketide Synthase. *ACS Chem. Biol.* **2013,** *8* (2), 443-450.

63. Koryakina, I.; Kasey, C.; McArthur, J. B.; Lowell, A. N.; Chemler, J. A.; Li, S.; Hansen, D. A.; Sherman, D. H.; Williams, G. J., Inversion of Extender Unit Selectivity in the Erythromycin Polyketide Synthase by Acyltransferase Domain Engineering. *ACS Chem. Biol.* **2017,** *12* (1), 114-123.

64. Koryakina, I.; McArthur, J. B.; Draelos, M. M.; Williams, G. J., Promiscuity of a Modular Polyketide Synthase Towards Natural and Non-Natural Extender Units. *Org. Biomol. Chem.* **2013,** *11* (27), 4449-4458.

65. Keatinge-Clay, A. T., Polyketide Synthase Modules Redefined. *Angew. Chem. Int. Ed.* **2017,** *56* (17), 4658-4660.

66. Zhang, L.; Hashimoto, T.; Qin, B.; Hashimoto, J.; Kozone, 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., Characterization of Giant Modular Pkss Provides Insight into Genetic Mechanism for Structural Diversification of Aminopolyol Polyketides. *Angew. Chem. Int. Ed.* **2017,** *56* (7), 1740-1745.

67. Lowry, B.; Li, X.; Robbins, T.; Cane, D. E.; Khosla, C., A Turnstile Mechanism for the Controlled Growth of Biosynthetic Intermediates on Assembly Line Polyketide Synthases. *ACS central science* **2016,** *2* (1), 14-20.

68. Eng, C. H.; Backman, T. W. H.; Bailey, C. B.; Magnan, C.; García Martín, H.; Katz, L.; Baldi, P.; Keasling, J. D., Clustercad: A Computational Platform for Type I Modular Polyketide Synthase Design. *Nucleic Acids Research* **2017**, gkx893-gkx893.

69. Hagen, A.; Poust, S.; Rond, T. d.; Fortman, J. L.; Katz, L.; Petzold, C. J.; Keasling, J. D., Engineering a Polyketide Synthase for in Vitro Production of Adipic Acid. *ACS Synth. Biol.* **2016,** *5* (1), 21-27.

70. Zheng, J.; Piasecki, S. K.; Keatinge-Clay, A. T., Structural Studies of an A2-Type Modular Polyketide Synthase Ketoreductase Reveal Features Controlling A-Substituent Stereochemistry. *ACS Chem. Biol.* **2013,** *8* (9), 1964-1971.

71. Nielsen, M. L.; Isbrandt, T.; Petersen, L. M.; Mortensen, U. H.; Andersen, M. R.; Hoof, J. B.; Larsen, T. O., Linker Flexibility Facilitates Module Exchange in Fungal Hybrid Pks-Nrps Engineering. *PLOS ONE* **2016,** *11* (8), e0161199.

72. Chemler, J. A.; Tripathi, A.; Hansen, D. A.; O'Neil-Johnson, M.; Williams, R. B.; Starks, C.; Park, S. R.; Sherman, D. H., Evolution of Efficient Modular Polyketide Synthases by Homologous Recombination. *J. Am. Chem. Soc.* **2015,** *137* (33), 10603-10609.

73. Wlodek, A.; Kendrew, S. G.; Coates, N. J.; Hold, A.; Pogwizd, J.; Rudder, S.; Sheehan, L. S.; Higginbotham, S. J.; Stanley-smith, A. E.; Warneck, T.; Nur-e-alam, M.; Radzom, M.; Martin, C. J.; Overvoorde, L.; Samborskyy, M.; Alt, S.; Heine, D.; Carter, G. T.; Graziani, E. I.; Koehn, F. E.; McDonald, L.; Alanine, A.; Rodríguez Sarmiento, R. M.; Chao, S. K.; Ratni, H.; Steward, L.; Norville, I. H.; Sarkar-tyson, M.; Moss, S. J.; Leadlay, P. F.; Wilkinson, B.; Gregory, M. A., Diversity Oriented Biosynthesis Via Accelerated Evolution of Modular Gene Clusters. *Nature Communications* **2017,** *8*, 1-10.

74. Lu, C.; Zhang, X.; Jiang, M.; Bai, L., Enhanced Salinomycin Production by Adjusting the Supply of Polyketide Extender Units in Streptomyces Albus. *Metabolic engineering* **2016,** *35*, 129-137.

75. Huang, J.; Yu, Z.; Li, M.-H.; Wang, J.-D.; Bai, H.; Zhou, J.; Zheng, Y.-G., High Level of Spinosad Production in the Heterologous Host Saccharopolyspora Erythraea. *Applied and Environmental Microbiology* **2016,** *82* (18), 5603-5611.

76. Koryakina, I.; Williams, G. J., Mutant Malonyl-Coa Synthetases with Altered Specificity for Polyketide Synthase Extender Unit Generation. *Chembiochem* **2011,** *12* (15), 2289-2293.

77. Thuronyi, B. W.; Privalsky, T. M.; Chang, M. C. Y., Engineered Fluorine Metabolism and Fluoropolymer Production in Living Cells. *Angew. Chem. Int. Ed.* **2017,** *56* (44), 13637-13640.

78. Basitta, P.; Westrich, L.; Rösch, M.; Kulik, A.; Gust, B.; Apel, A. K., Agos: A Plug-and-Play Method for the Assembly of Artificial Gene Operons into Functional Biosynthetic Gene Clusters. *ACS Synth. Biol.* **2017,** *6* (5), 817-825.

79. Fang, L.; Guell, M.; Church, G. M.; Pfeifer, B. A., Heterologous Erythromycin Production across Strain and Plasmid Construction. *Biotechnol Prog* **2017**.

80. Jiang, M.; Fang, L.; Pfeifer, B. A., Improved Heterologous Erythromycin a Production through Expression Plasmid Re-Design. *Biotechnol Prog* **2013,** *29* (4), 862-869.

81. Clevenger, K. D.; Bok, J. W.; Ye, R.; Miley, G. P.; Verdan, M. H.; Velk, T.; Chen, C.; Yang, K.; Robey, M. T.; Gao, P.; Lamprecht, M.; Thomas, P. M.; Islam, M. N.; Palmer, J. M.; Wu, C. C.; Keller, N. P.; Kelleher, N. L., A Scalable Platform to Identify Fungal Secondary Metabolites and Their Gene Clusters. *Nature Chemical Biology* **2017,** *13*, 895.

82. Zhang, M. M.; Wong, F. T.; Wang, Y.; Luo, S.; Lim, Y. H.; Heng, E.; Yeo, W. L.; Cobb, R. E.; Enghiad, B.; Ang, E. L.; Zhao, H., Crispr–Cas9 Strategy for Activation of Silent Streptomyces Biosynthetic Gene Clusters. *Nature Chemical Biology* **2017,** *13*, 607.

83. Liu, Y.; Wei, W. P.; Ye, B. C., High Gc Content Cas9-Mediated Genome-Editing and Biosynthetic Gene Cluster Activation in Saccharopolyspora Erythraea. *ACS Synth Biol* **2018**.

84. Rebets, Y.; Schmelz, S.; Gromyko, O.; Tistechok, S.; Petzke, L.; Scrima, A.; Luzhetskyy, A., Design, Development and Application of Whole-Cell Based Antibiotic-Specific Biosensor. *Metabolic engineering* **2018,** *47*, 263-270.

85. Kasey, C. M.; Zerrad, M.; Li, Y.; Cropp, T. A.; Williams, G. J., Development of Transcription Factor-Based Designer Macrolide Biosensors for Metabolic Engineering and Synthetic Biology. *ACS Synth Biol* **2017**.

86. Maimone, T. J.; Baran, P. S., Modern Synthetic Efforts toward Biologically Active Terpenes. *Nature Chemical Biology* **2007,** *3* (7), 396-407.

87. Monteiro, J. L. F.; Veloso, C. O., Catalytic Conversion of Terpenes into Fine Chemicals. *Topics in Catalysis* **2004,** *27* (1), 169-180.

88. Steele, C. L.; Crock, J.; Bohlmann, J.; Croteau, R., Sesquiterpene Synthases from Grand Fir (Abies Grandis). Comparison of Constitutive and Wound-Induced Activities, and Cdna Isolation, Characterization, and Bacterial Expression of Delta-Selinene Synthase and Gamma-Humulene Synthase. *J Biol Chem* **1998,** *273* (4), 2078-2089.

89. Brill, Z. G.; Condakes, M. L.; Ting, C. P.; Maimone, T. J., Navigating the Chiral Pool in the Total Synthesis of Complex Terpene Natural Products. *Chemical Reviews* **2017**.

90. Ro, D.-K.; Paradise, E. M.; Ouellet, M.; Fisher, K. J.; Newman, K. L.; Ndungu, J. M.; Ho, K. A.; Eachus, R. A.; Ham, T. S.; Kirby, J.; Chang, M. C. Y.; Withers, S. T.; Shiba, Y.; Sarpong, R.; Keasling, J. D., Production of the Antimalarial Drug Precursor Artemisinic Acid in Engineered Yeast. *Nature* **2006,** *440* (7086), 940-943.

91. Westfall, P. J.; Pitera, D. J.; Lenihan, J. R.; Eng, D.; Woolard, F. X.; Regentin, R.; Horning, T.; Tsuruta, H.; Melis, D. J.; Owens, A.; Fickes, S.; Diola, D.; Benjamin, K. R.; Keasling, J. D.; Leavell, M. D.; McPhee, D. J.; Renninger, N. S.; Newman, J. D.; Paddon, C. J., Production of Amorphadiene in Yeast, and Its Conversion to Dihydroartemisinic Acid, Precursor to the Antimalarial Agent Artemisinin. *Proceedings of the National Academy of Sciences* **2012,** *109* (3), E111-E118.

92. Paddon, C. J.; Westfall, P. J.; Pitera, D. J.; Benjamin, K.; Fisher, K.; McPhee, D.; Leavell, M. D.; Tai, A.; Main, A.; Eng, D.; Polichuk, D. R.; Teoh, K. H.; Reed, D. W.; Treynor, T.; Lenihan, J.; Jiang, H.; Fleck, M.; Bajad, S.; Dang, G.; Dengrove, D.; Diola, D.; Dorin, G.;

Ellens, K. W.; Fickes, S.; Galazzo, J.; Gaucher, S. P.; Geistlinger, T.; Henry, R.; Hepp, M.; Horning, T.; Iqbal, T.; Kizer, L.; Lieu, B.; Melis, D.; Moss, N.; Regentin, R.; Secrest, S.; Tsuruta, H.; Vazquez, R.; Westblade, L. F.; Xu, L.; Yu, M.; Zhang, Y.; Zhao, L.; Lievense, J.; Covello, P. S.; Keasling, J. D.; Reiling, K. K.; Renninger, N. S.; Newman, J. D., High-Level Semi-Synthetic Production of the Potent Antimalarial Artemisinin. *Nature* **2013,** *496* (7446), 528-532.

93. Kong, J.; Yang, Y.; Wang, W.; Cheng, K.; Zhu, P., Artemisinic Acid: A Promising Molecule Potentially Suitable for the Semi-Synthesis of Artemisinin. *RSC Advances* **2013,** *3* (21), 7622-7641.

94. Turconi, J.; Griolet, F.; Guevel, R.; Oddon, G.; Villa, R.; Geatti, A.; Hvala, M.; Rossen, K.; Göller, R.; Burgard, A., Semisynthetic Artemisinin, the Chemical Path to Industrial Production. *Organic Process Research & Development* **2014,** *18* (3), 417-422.

95. Peplow, M., Synthetic Biology's First Malaria Drug Meets Market Resistance. In *Nature*, England, 2016; Vol. 530, pp 389-390.

96. Demiray, M.; Tang, X.; Wirth, T.; Faraldos, J. A.; Allemann, R. K., An Efficient Chemoenzymatic Synthesis of Dihydroartemisinic Aldehyde. *Angew. Chem. Int. Ed.* **2017,** *56* (15), 4347-4350.

97. Tang, X.; Demiray, M.; Wirth, T.; Allemann, R. K., Concise Synthesis of Artemisinin from a Farnesyl Diphosphate Analogue. *Bioorganic & Medicinal Chemistry* **2017**.

98. Cascon, O.; Touchet, S.; Miller, D. J.; Gonzalez, V.; Faraldos, J. A.; Allemann, R. K., Chemoenzymatic Preparation of Germacrene Analogues. *Chemical Communications* **2012,** *48* (78), 9702-9704.

99. Faraldos, J. A.; Miller, D. J.; González, V.; Yoosuf-Aly, Z.; Cascón, O.; Li, A.; Allemann, R. K., A 1,6-Ring Closure Mechanism for (+)-Δ-Cadinene Synthase? *J. Am. Chem. Soc.* **2012,** *134* (13), 5900-5908.

100. Cane, D. E.; Tsantrizos, Y. S., Aristolochene Synthase. Elucidation of the Cryptic Germacrene a Synthase Activity Using the Anomalous Substrate Dihydrofarnesyl Diphosphate. *J. Am. Chem. Soc.* **1996,** *118* (42), 10037-10040.

101. Jin, Q.; Williams, D. C.; Hezari, M.; Croteau, R.; Coates, R. M., Stereochemistry of the Macrocyclization and Elimination Steps in Taxadiene Biosynthesis through Deuterium Labeling. *The Journal of organic chemistry* **2005,** *70* (12), 4667-4675.

102. Jin, Y.; Williams, D. C.; Croteau, R.; Coates, R. M., Taxadiene Synthase-Catalyzed Cyclization of 6-Fluorogeranylgeranyl Diphosphate to 7-Fluoroverticillenes. *J Am Chem Soc* **2005,** *127* (21), 7834-7842.

103. Vattekkatte, A.; Gatto, N.; Schulze, E.; Brandt, W.; Boland, W., Inhibition of a Multiproduct Terpene Synthase from Medicago Truncatula by 3-Bromoprenyl Diphosphates. *Org Biomol Chem* **2015,** *13* (16), 4776-4784.

104. Miller, D. J.; Yu, F.; Knight, D. W.; Allemann, R. K., 6- and 14-Fluoro Farnesyl Diphosphate: Mechanistic Probes for the Reaction Catalysed by Aristolochene Synthase. *Org Biomol Chem* **2009,** *7* (5), 962-975.

105. Lauchli, R.; Rabe, K. S.; Kalbarczyk, K. Z.; Tata, A.; Heel, T.; Kitto, R. Z.; Arnold, F. H., High-Throughput Screening and Directed Evolution of Terpene Synthase-Catalyzed Cylization(). *Angewandte Chemie (International ed. in English)* **2013,** *52* (21), 10.1002/anie.201301362.

106. Bowers, W.; Nishino, C.; Montgomery, M.; Nault, L.; Nielson, M., Sesquiterpene Progenitor, Germacrene A: An Alarm Pheromone in Aphids. *Science* **1977,** *196* (4290), 680-681.

107. Andreas, K.; Benjamin, S.; Vanessa, H.; Clara, O.; Katja, S.; Kimia, E.; Sascha, B.; Jeroen, D.; Lukas, L.; Sven, W., Exploiting the Synthetic Potential of Sesquiterpene Cyclases for Generating Unnatural Terpenoids. *Angew. Chem. Int. Ed. 0* (ja).

108. Little, D. B.; Croteau, R. B., Alteration of Product Formation by Directed Mutagenesis and Truncation of the Multiple-Product Sesquiterpene Synthases Δ-Selinene Synthase and Γ-Humulene Synthase. *Archives of Biochemistry and Biophysics* **2002,** *402* (1), 120-135.

109. Yoshikuni, Y.; Ferrin, T. E.; Keasling, J. D., Designed Divergent Evolution of Enzyme Function. *Nature* **2006,** *440* (7087), 1078-1082.

110. Alberti, F.; Khairudin, K.; Venegas, E. R.; Davies, J. A.; Hayes, P. M.; Willis, C. L.; Bailey, A. M.; Foster, G. D., Heterologous Expression Reveals the Biosynthesis of the Antibiotic Pleuromutilin and Generates Bioactive Semi-Synthetic Derivatives. *Nat Commun* **2017,** *8* (1), 1831.

111. Kuzuyama, T.; Noel, J. P.; Richard, S. B., Structural Basis for the Promiscuous Biosynthetic Prenylation of Aromatic Natural Products. *Nature* **2005,** *435* (7044), 983-987.

112. Itoh, T.; Tokunaga, K.; Matsuda, Y.; Fujii, I.; Abe, I.; Ebizuka, Y.; Kushiro, T., Reconstitution of a Fungal Meroterpenoid Biosynthesis Reveals the Involvement of a Novel Family of Terpene Cyclases. *Nature Chemistry* **2010,** *2* (10), 858-864.

113. Leipoldt, F.; Zeyhle, P.; Kulik, A.; Kalinowski, J.; Heide, L.; Kaysser, L., Diversity of Abba Prenyltransferases in Marine Streptomyces Sp. Cnq-509: Promiscuous Enzymes for the Biosynthesis of Mixed Terpenoid Compounds. *PLoS One* **2015,** *10* (12), e0143237.

114. Botta, B.; Monache, G. D.; Menendez, P.; Boffi, A., Novel Prenyltransferase Enzymes as a Tool for Flavonoid Prenylation. *Trends in Pharmacological Sciences* **2005,** *26* (12), 606-608.

115. Kim, B. S.; Sherman, D. H.; Reynolds, K. A., An Efficient Method for Creation and Functional Analysis of Libraries of Hybrid Type I Polyketide Synthases. *Protein Eng Des Sel* **2004,** *17* (3), 277-284.

116. Kalkreuter, E.; Williams, G. J., Engineering Enzymatic Assembly Lines for the Production of New Antimicrobials. *Curr. Opin. Microbiol.* **2018,** *45*, 140-148.

117. Wong, F. T.; Jin, X.; Mathews, I. I.; Cane, D. E.; Khosla, C., Structure and Mechanism of the Trans-Acting Acyltransferase from the Disorazole Synthase. *Biochemistry* **2011,** *50* (30), 6539-6548.

118. Oefner, C.; Schulz, H.; D'Arcy, A.; Dale, G. E., Mapping the Active Site of Escherichia Coli Malonyl-Coa-Acyl Carrier Protein Transacylase (Fabd) by Protein Crystallography. *Acta Crystallogr D Biol Crystallogr* **2006,** *62* (Pt 6), 613-618.

119. Keatinge-Clay, A. T.; Shelat, A. A.; Savage, D. F.; Tsai, S.-C.; Miercke, L. J. W.; O'Connell Iii, J. D.; Khosla, C.; Stroud, R. M., Catalysis, Specificity, and Acp Docking Site of Streptomyces Coelicolor Malonyl-Coa:Acp Transacylase. *Structure (London, England : 1993)* **2003,** *11* (2), 147-154.

120. Park, H.; Kevany, B. M.; Dyer, D. H.; Thomas, M. G.; Forest, K. T., A Polyketide Synthase Acyltransferase Domain Structure Suggests a Recognition Mechanism for Its Hydroxymalonyl-Acyl Carrier Protein Substrate. *PLoS One* **2014,** *9* (10), e110965.

121. Del Vecchio, F.; Petkovic, H.; Kendrew, S. G.; Low, L.; Wilkinson, B.; Lill, R.; Cortes, J.; Rudd, B. A.; Staunton, J.; Leadlay, P. F., Active-Site Residue, Domain and Module Swaps in Modular Polyketide Synthases. *J. Ind. Microbiol. Biotechnol.* **2003,** *30* (8), 489-494.

122. Tang, Y.; Chen, A. Y.; Kim, C. Y.; Cane, D. E.; Khosla, C., Structural and Mechanistic Analysis of Protein Interactions in Module 3 of the 6-Deoxyerythronolide B Synthase. *Chem Biol* **2007,** *14* (8), 931-943.

123. Tang, Y.; Kim, C. Y.; Mathews, II; Cane, D. E.; Khosla, C., The 2.7-Angstrom Crystal Structure of a 194-Kda Homodimeric Fragment of the 6-Deoxyerythronolide B Synthase. *Proc Natl Acad Sci U S A* **2006,** *103* (30), 11124-11129.

124. Zhang, Y., I-Tasser Server for Protein 3d Structure Prediction. *BMC Bioinfom.* **2008,** *9*, 40.

125. Roy, A.; Kucukural, A.; Zhang, Y., I-Tasser: A Unified Platform for Automated Protein Structure and Function Prediction. *Nat. Protoc.* **2010,** *5* (4), 725-738.

126. Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y., The I-Tasser Suite: Protein Structure and Function Prediction. *Nat. Methods* **2015,** *12* (1), 7-8.

127. Dunn, B. J.; Cane, D. E.; Khosla, C., Mechanism and Specificity of an Acyltransferase Domain from a Modular Polyketide Synthase. *Biochemistry* **2013,** *52* (11), 1839-1841.

128. D.A. Case, J. T. B., R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K.M. Merz, G. Monard, P. Needham, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, R. Salomon-Ferrer, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, D.M. York, and P.A. Kollman *Amber 2015*, University of California, San Francisco, 2015.

129. Counterman, A. E.; Clemmer, D. E., Volumes of Individual Amino Acid Residues in Gas-Phase Peptide Ions. *J. Am. Chem. Soc.* **1999,** *121* (16), 4031-4039.

130. Petkovic, H.; Kendrew, S.; Leadlay, P. Polyketides and Their Synthesis. 2004.

131. Poust, S.; Yoon, I.; Adams, P. D.; Katz, L.; Petzold, C. J.; Keasling, J. D., Understanding the Role of Histidine in the Ghsxg Acyltransferase Active Site Motif: Evidence for Histidine Stabilization of the Malonyl-Enzyme Intermediate. *PLoS ONE* **2014,** *9* (10), e109421.

132. E. Vanquelef, S. S., G. Marquant, E. Garcia, G. Klimerak, J.C. Delepine, P. Cieplak, and F.Y. Dupradeau, R.E.D. Server: A Web Service for Deriving Resp and Esp Charges and Building Force Field Libraries for New Molecules and Molecular Fragments. *Nucl. Acids Res.* **2011,** *39*, W511.

133. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.;

Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Gaussian, Inc.: Wallingford, CT, USA, 2009.

134. W. Humphrey, A. D., and K. Schulten, Vmd - Visual Molecular Dynamics. *J. Molec. Graphics* **1996,** *14*, 33-38.

135. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., Ucsf Chimera--a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004,** *25* (13), 1605-1612.

136. Sanner, M. F.; Olson, A. J.; Spehner, J. C., Reduced Surface: An Efficient Way to Compute Molecular Surfaces. *Biopolymers* **1996,** *38* (3), 305-320.

137. Roe, D. R.; Cheatham, T. E., Ptraj and Cpptraj: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013,** *9* (7), 3084-3095.

138. Persistence of Vision Pty Ltd *Persistence of Vision Raytracer (Version 3.6)*, 2004.

139. Donadio, S.; Katz, L., Organization of the Enzymatic Domains in the Multifunctional Polyketide Synthase Involved in Erythromycin Formation in Saccharopolyspora Erythraea. *Gene* **1992,** *111* (1), 51-60.

140. Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A., The Specificity-Conferring Code of Adenylation Domains in Nonribosomal Peptide Synthetases. *Chem Biol* **1999,** *6* (8), 493-505.

141. Khayatt, B. I.; Overmars, L.; Siezen, R. J.; Francke, C., Classification of the Adenylation and Acyl-Transferase Activity of Nrps and Pks Systems Using Ensembles of Substrate Specific Hidden Markov Models. *PLoS One* **2013,** *8* (4), e62136.

142. Petkovic, H.; Sandmann, A.; Challis, I. R.; Hecht, H.-J.; Silakowski, B.; Low, L.; Beeston, N.; Kuscer, E.; Garcia-Bernardo, J.; Leadlay, P. F.; Kendrew, S. G.; Wilkinson, B.; Muller,

R., Substrate Specificity of the Acyl Transferase Domains of Epoc from the Epothilone Polyketide Synthase. *Org. Biomol. Chem.* **2008,** *6* (3), 500-506.

143. 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., Structural Basis of a Broadly Selective Acyltransferase from the Polyketide Synthase of Splenocin. *Angew. Chem. Int. Ed.* **2018,** *57* (20), 5823-5827.

144. Mortison, J. D.; Kittendorf, J. D.; Sherman, D. H., Synthesis and Biochemical Analysis of Complex Chain-Elongation Intermediates for Interrogation of Molecular Specificity in the Erythromycin and Pikromycin Polyketide Synthases. *J. Am. Chem. Soc.* **2009,** *131* (43), 15784-15793.

145. Rachid, S.; Huo, L.; Herrmann, J.; Stadler, M.; Kopcke, B.; Bitzer, J.; Muller, R., Mining the Cinnabaramide Biosynthetic Pathway to Generate Novel Proteasome Inhibitors. *Chembiochem* **2011,** *12* (6), 922-931.

146. Greule, A.; Intra, B.; Flemming, S.; Rommel, M. G.; Panbangred, W.; Bechthold, A., The Draft Genome Sequence of Actinokineospora Bangkokensis 44ehw(T) Reveals the Biosynthetic Pathway of the Antifungal Thailandin Compounds with Unusual Butylmalonyl-Coa Extender Units. *Molecules (Basel, Switzerland)* **2016,** *21* (11).

147. Walker, M. C.; Thuronyi, B. W.; Charkoudian, L. K.; Lowry, B.; Khosla, C.; Chang, M. C. Y., Expanding the Fluorine Chemistry of Living Systems Using Engineered Polyketide Synthase Pathways. *Science (New York, N.Y.)* **2013,** *341* (6150), 1089-1094.

148. Khosla, C.; Tang, Y.; Chen, A. Y.; Schnarr, N. A.; Cane, D. E., Structure and Mechanism of the 6-Deoxyerythronolide B Synthase. *Annual review of biochemistry* **2007,** *76*, 195-221.

149. Chen, A. Y.; Schnarr, N. A.; Kim, C. Y.; Cane, D. E.; Khosla, C., Extender Unit and Acyl Carrier Protein Specificity of Ketosynthase Domains of the 6-Deoxyerythronolide B Synthase. *J Am Chem Soc* **2006,** *128* (9), 3067-3074.

150. Wong, F. T.; Chen, A. Y.; Cane, D. E.; Khosla, C., Protein-Protein Recognition between Acyltransferases and Acyl Carrier Proteins in Multimodular Polyketide Synthases. *Biochemistry* **2010,** *49* (1), 95-102.

151. Moller, D.; Kushnir, S.; Grote, M.; Ismail-Ali, A.; Koopmans, K. R. M.; Calo, F.; Heinrich, S.; Diehl, B.; Schulz, F., Flexible Enzymatic Activation of Artificial Polyketide Extender Units by Streptomyces Cinnamonensis into the Monensin Biosynthetic Pathway. *Lett. Appl. Microbiol.* **2018,** *67*, 226-234.

152. Vögeli, B.; Geyer, K.; Gerlinger, P. D.; Benkstein, S.; Cortina, N. S.; Erb, T. J., Combining Promiscuous Acyl-Coa Oxidase and Enoyl-Coa Carboxylase/Reductases for Atypical Polyketide Extender Unit Biosynthesis. *Cell Chem. Biol.* **2018,** *25* (7), 833-839.e834.

153. Murli, S.; Kennedy, J.; Dayem, L. C.; Carney, J. R.; Kealey, J. T., Metabolic Engineering of Escherichia Coli for Improved 6-Deoxyerythronolide B Production. *J Ind Microbiol Biotechnol* **2003,** *30* (8), 500-509.

154. Moore, S. 'Round-the-Horn Site-Directed Mutagenesis. https://openwetware.org/wiki/'Round-the-horn_site-directed_mutagenesis.

155. Gibson, D. G.; Young, L.; Chuang, R. Y.; Venter, J. C.; Hutchison, C. A., 3rd; Smith, H. O., Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat. Methods* **2009,** *6* (5), 343-345.

156. Koryakina, I.; McArthur, J.; Randall, S.; Draelos, M. M.; Musiol, E. M.; Muddiman, D. C.; Weber, T.; Williams, G. J., Poly Specific Trans-Acyltransferase Machinery Revealed Via Engineered Acyl-Coa Synthetases. *ACS Chem. Biol.* **2013,** *8* (1), 200-208.

157. Zhang, J.; Liang, Y.; Zhang, Y., Atomic-Level Protein Structure Refinement Using Fragment-Guided Molecular Dynamics Conformation Sampling. *Structure (London, England : 1993)* **2011,** *19* (12), 1784-1795.

158. Chan, Y. A.; Boyne, M. T., 2nd; Podevels, A. M.; Klimowicz, A. K.; Handelsman, J.; Kelleher, N. L.; Thomas, M. G., Hydroxymalonyl-Acyl Carrier Protein (Acp) and Aminomalonyl-Acp Are Two Additional Type I Polyketide Synthase Extender Units. *Proc. Natl. Acad. Sci. USA* **2006,** *103* (39), 14349-14354.

159. Mo, S.; Kim, D. H.; Lee, J. H.; Park, J. W.; Basnet, D. B.; Ban, Y. H.; Yoo, Y. J.; Chen, S. W.; Park, S. R.; Choi, E. A.; Kim, E.; Jin, Y. Y.; Lee, S. K.; Park, J. Y.; Liu, Y.; Lee, M. O.; Lee, K. S.; Kim, S. J.; Kim, D.; Park, B. C.; Lee, S. G.; Kwon, H. J.; Suh, J. W.; Moore, B. S.; Lim, S. K.; Yoon, Y. J., Biosynthesis of the Allylmalonyl-Coa Extender Unit for the Fk506 Polyketide Synthase Proceeds through a Dedicated Polyketide Synthase and Facilitates the Mutasynthesis of Analogues. *J. Am. Chem. Soc.* **2011,** *133* (4), 976-985.

160. Chang, C.; Huang, R.; Yan, Y.; Ma, H.; Dai, Z.; Zhang, B.; Deng, Z.; Liu, W.; Qu, X., Uncovering the Formation and Selection of Benzylmalonyl-Coa from the Biosynthesis of Splenocin and Enterocin Reveals a Versatile Way to Introduce Amino Acids into Polyketide Carbon Scaffolds. *J. Am. Chem. Soc.* **2015,** *137* (12), 4183-4190.

161. Yoo, H. G.; Kwon, S. Y.; Kim, S.; Karki, S.; Park, Z. Y.; Kwon, H. J., Characterization of 2-Octenoyl-Coa Carboxylase/Reductase Utilizing Pteb from Streptomyce Avermitilis. *Biosci., Biotechnol., Biochem.* **2011,** *75* (6), 1191-1193.

162. 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., Biosynthesis of the Salinosporamide a Polyketide Synthase Substrate Chloroethylmalonyl-Coenzyme a from *S*-Adenosyl-L-Methionine. *Proc. Natl. Acad. Sci. USA* **2009,** *106* (30), 12295-12300.

163. McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Ashley, G., Multiple Genetic Modifications of the Erythromycin Polyketide Synthase to Produce a Library of Novel "Unnatural" Natural Products. *Proc Natl Acad Sci U S A* **1999,** *96* (5), 1846-1851.

164. Xue, Y.; Zhao, L.; Liu, H. W.; Sherman, D. H., A Gene Cluster for Macrolide Antibiotic Biosynthesis in Streptomyces Venezuelae: Architecture of Metabolic Diversity. *Proc. Natl. Acad. Sci. USA* **1998,** *95* (21), 12111-12116.

165. Jung, W. S.; Kim, E.; Yoo, Y. J.; Ban, Y. H.; Kim, E. J.; Yoon, Y. J., Characterization and Engineering of the Ethylmalonyl-Coa Pathway Towards the Improved Heterologous Production of Polyketides in Streptomyces Venezuelae. *Appl. Microbiol. Biotechnol.* **2014,** *98* (8), 3701-3713.

166. Bonnett, Shilah A.; Rath, Christopher M.; Shareef, A.-R.; Joels, Joanna R.; Chemler, Joseph A.; Håkansson, K.; Reynolds, K.; Sherman, David H., Acyl-Coa Subunit Selectivity in the Pikromycin Polyketide Synthase Pikaiv: Steady-State Kinetics and Active-Site Occupancy Analysis by Fticr-Ms. *Chem. Biol.* **2011,** *18* (9), 1075-1081.

167. Pieper, R.; Gokhale, R. S.; Luo, G.; Cane, D. E.; Khosla, C., Purification and Characterization of Bimodular and Trimodular Derivatives of the Erythromycin Polyketide Synthase. *Biochemistry* **1997,** *36* (7), 1846-1851.

168. Bycroft, M.; Weissman, K. J.; Staunton, J.; Leadlay, P. F., Efficient Purification and Kinetic Characterization of a Bimodular Derivative of the Erythromycin Polyketide Synthase. *Eur. J. Biochem.* **2000,** *267* (2), 520-526.

169. Holzbaur, I. E.; Ranganathan, A.; Thomas, I. P.; Kearney, D. J.; Reather, J. A.; Rudd, B. A.; Staunton, J.; Leadlay, P. F., Molecular Basis of Celmer's Rules: Role of the Ketosynthase Domain in Epimerisation and Demonstration That Ketoreductase Domains Can Have Altered Product Specificity with Unnatural Substrates. *Chem. Biol.* **2001,** *8* (4), 329-340.

170. Smith, L.; Hong, H.; Spencer, J. B.; Leadlay, P. F., Analysis of Specific Mutants in the Lasalocid Gene Cluster: Evidence for Enzymatic Catalysis of a Disfavoured Polyether Ring Closure. *ChemBioChem* **2008,** *9* (18), 2967-2975.

171. Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C., Biosynthesis of Complex Polyketides in a Metabolically Engineered Strain of E. Coli. *Science* **2001,** *291* (5509), 1790-1792.

172. Zhang, F.; Carothers, J. M.; Keasling, J. D., Design of a Dynamic Sensor-Regulator System for Production of Chemicals and Fuels Derived from Fatty Acids. *Nat. Biotechnol.* **2012,** *30* (4), 354-359.

173. Siedler, S.; Schendzielorz, G.; Binder, S.; Eggeling, L.; Bringer, S.; Bott, M., Soxr as a Single-Cell Biosensor for Nadph-Consuming Enzymes in Escherichia Coli. *ACS Synth Biol* **2014,** *3* (1), 41-47.

174. Dahl, R. H.; Zhang, F.; Alonso-Gutierrez, J.; Baidoo, E.; Batth, T. S.; Redding-Johanson, A. M.; Petzold, C. J.; Mukhopadhyay, A.; Lee, T. S.; Adams, P. D.; Keasling, J. D.,

Engineering Dynamic Pathway Regulation Using Stress-Response Promoters. *Nat. Biotechnol.* **2013,** *31* (11), 1039-1046.

175. Johnson, A. O.; Gonzalez-Villanueva, M.; Wong, L.; Steinbüchel, A.; Tee, K. L.; Xu, P.; Wong, T. S., Design and Application of Genetically-Encoded Malonyl-Coa Biosensors for Metabolic Engineering of Microbial Cell Factories. *Metab. Eng.* **2017,** *44*, 253-264.

176. Zhang, H.; Boghigian, B. A.; Pfeifer, B. A., Investigating the Role of Native Propionyl-Coa and Methylmalonyl-Coa Metabolism on Heterologous Polyketide Production in Escherichia Coli. *Biotechnol Bioeng* **2010,** *105* (3), 567-573.

177. Dayem, L. C.; Carney, J. R.; Santi, D. V.; Pfeifer, B. A.; Khosla, C.; Kealey, J. T., Metabolic Engineering of a Methylmalonyl-Coa Mutase-Epimerase Pathway for Complex Polyketide Biosynthesis in Escherichia Coli. *Biochemistry* **2002,** *41* (16), 5193-5201.

178. Cai, W.; Zhang, W., Engineering Modular Polyketide Synthases for Production of Biofuels and Industrial Chemicals. *Current Opinion in Biotechnology* **2018,** *50*, 32-38.

179. Schujman, G. E.; Paoletti, L.; Grossman, A. D.; de Mendoza, D., Fapr, a Bacterial Transcription Factor Involved in Global Regulation of Membrane Lipid Biosynthesis. *Developmental Cell* **2003,** *4* (5), 663-672.

180. Schujman, G. E.; Guerin, M.; Buschiazzo, A.; Schaeffer, F.; Llarrull, L. I.; Reh, G.; Vila, A. J.; Alzari, P. M.; de Mendoza, D., Structural Basis of Lipid Biosynthesis Regulation in Gram-Positive Bacteria. *The EMBO journal* **2006,** *25* (17), 4074-4083.

181. Albanesi, D.; Reh, G.; Guerin, M. E.; Schaeffer, F.; Debarbouille, M.; Buschiazzo, A.; Schujman, G. E.; de Mendoza, D.; Alzari, P. M., Structural Basis for Feed-Forward Transcriptional Regulation of Membrane Lipid Homeostasis in Staphylococcus Aureus. *PLoS pathogens* **2013,** *9* (1), e1003108.

182. Xu, P.; Li, L.; Zhang, F.; Stephanopoulos, G.; Koffas, M., Improving Fatty Acids Production by Engineering Dynamic Pathway Regulation and Metabolic Control. *Proc Natl Acad Sci U S A* **2014,** *111* (31), 11299-11304.

183. Xu, P.; Wang, W.; Li, L.; Bhan, N.; Zhang, F.; Koffas, M. A., Design and Kinetic Analysis of a Hybrid Promoter-Regulator System for Malonyl-Coa Sensing in Escherichia Coli. *ACS Chem. Biol.* **2014,** *9* (2), 451-458.

184. Liu, D.; Xiao, Y.; Evans, B. S.; Zhang, F., Negative Feedback Regulation of Fatty Acid Production Based on a Malonyl-Coa Sensor–Actuator. *ACS Synth. Biol.* **2015,** *4* (2), 132-140.

185. David, F.; Nielsen, J.; Siewers, V., Flux Control at the Malonyl-Coa Node through Hierarchical Dynamic Pathway Regulation in Saccharomyces Cerevisiae. *ACS Synth Biol* **2016,** *5* (3), 224-233.

186. Li, S.; Si, T.; Wang, M.; Zhao, H., Development of a Synthetic Malonyl-Coa Sensor in Saccharomyces Cerevisiae for Intracellular Metabolite Monitoring and Genetic Screening. *ACS Synth Biol* **2015,** *4* (12), 1308-1315.

187. Ellis, J. M.; Wolfgang, M. J., A Genetically Encoded Metabolite Sensor for Malonyl-Coa. *Chem Biol* **2012,** *19* (10), 1333-1339.

188. Shaner, N. C.; Lambert, G. G.; Chammas, A.; Ni, Y.; Cranfill, P. J.; Baird, M. A.; Sell, B. R.; Allen, J. R.; Day, R. N.; Israelsson, M.; Davidson, M. W.; Wang, J., A Bright Monomeric Green Fluorescent Protein Derived from Branchiostoma Lanceolatum. *Nat. Methods* **2013,** *10* (5), 407-409.

189. Espah Borujeni, A.; Channarasappa, A. S.; Salis, H. M., Translation Rate Is Controlled by Coupled Trade-Offs between Site Accessibility, Selective Rna Unfolding and Sliding at Upstream Standby Sites. *Nucleic Acids Research* **2014,** *42* (4), 2646-2659.

190. Salis, H. M.; Mirsky, E. A.; Voigt, C. A., Automated Design of Synthetic Ribosome Binding Sites to Control Protein Expression. *Nature Biotechnology* **2009,** *27*, 946.

191. Farasat, I.; Kushwaha, M.; Collens, J.; Easterbrook, M.; Guido, M.; Salis, H. M., Efficient Search, Mapping, and Optimization of Multi-Protein Genetic Systems in Diverse Bacteria. *Molecular Systems Biology* **2014,** *10* (6).

192. Price, A. C.; Choi, K.-H.; Heath, R. J.; Li, Z.; White, S. W.; Rock, C. O., Inhibition of B-Ketoacyl-Acyl Carrier Protein Synthases by Thiolactomycin and Cerulenin: Structure and Mechanism. *Journal of Biological Chemistry* **2001,** *276* (9), 6551-6559.

193. Pizer, E. S.; Thupari, J.; Han, W. F.; Pinn, M. L.; Chrest, F. J.; Frehywot, G. L.; Townsend, C. A.; Kuhajda, F. P., Malonyl-Coenzyme-a Is a Potential Mediator of Cytotoxicity Induced by Fatty-Acid Synthase Inhibition in Human Breast Cancer Cells and Xenografts. *Cancer research* **2000,** *60* (2), 213-218.

194. Thupari, J. N.; Pinn, M. L.; Kuhajda, F. P., Fatty Acid Synthase Inhibition in Human Breast Cancer Cells Leads to Malonyl-Coa-Induced Inhibition of Fatty Acid Oxidation and Cytotoxicity. *Biochemical and biophysical research communications* **2001,** *285* (2), 217-223.

195. Rodriguez, E.; Gramajo, H., Genetic and Biochemical Characterization of the Alpha and Beta Components of a Propionyl-Coa Carboxylase Complex of Streptomyces Coelicolor A3(2). *Microbiology (Reading, England)* **1999,** *145 (Pt 11)*, 3109-3119.

196. Cheng, F.; Zhu, L.; Schwaneberg, U., Directed Evolution 2.0: Improving and Deciphering Enzyme Properties. *Chem Commun (Camb)* **2015,** *51* (48), 9760-9772.

197. Silver, P. A.; Way, J. C.; Arnold, F. H.; Meyerowitz, J. T., Synthetic Biology: Engineering Explored. *Nature* **2014,** *509* (7499), 166-167.

198. Farmer, W. R.; Liao, J. C., Precursor Balancing for Metabolic Engineering of Lycopene Production in Escherichia Coli. *Biotechnol Prog* **2001,** *17* (1), 57-61.

199. Siu, Y.; Fenno, J.; Lindle, J. M.; Dunlop, M. J., Design and Selection of a Synthetic Feedback Loop for Optimizing Biofuel Tolerance. *ACS Synth. Biol.* **2018,** *7* (1), 16-23.

200. Forster-Fromme, K.; Jendrossek, D., Atur Is a Repressor of Acyclic Terpene Utilization (Atu) Gene Cluster Expression and Specifically Binds to Two 13 Bp Inverted Repeat Sequences of the Atua-Atur Intergenic Region. *FEMS microbiology letters* **2010,** *308* (2), 166-174.

201. Forster-Fromme, K.; Hoschle, B.; Mack, C.; Bott, M.; Armbruster, W.; Jendrossek, D., Identification of Genes and Proteins Necessary for Catabolism of Acyclic Terpenes and Leucine/Isovalerate in Pseudomonas Aeruginosa. *Appl Environ Microbiol* **2006,** *72* (7), 4819-4828.

202. Forster-Fromme, K.; Jendrossek, D., Catabolism of Citronellol and Related Acyclic Terpenoids in Pseudomonads. *Appl. Microbiol. Biotechnol.* **2010,** *87* (3), 859-869.

203. Ramos, J. L.; Martinez-Bueno, M.; Molina-Henares, A. J.; Teran, W.; Watanabe, K.; Zhang, X.; Gallegos, M. T.; Brennan, R.; Tobes, R., The Tetr Family of Transcriptional Repressors. *Microbiology and molecular biology reviews : MMBR* **2005,** *69* (2), 326-356.

204. Yu, Z.; Reichheld, S. E.; Savchenko, A.; Parkinson, J.; Davidson, A. R., A Comprehensive Analysis of Structural and Sequence Conservation in the Tetr Family Transcriptional Regulators. *J Mol Biol* **2010,** *400* (4), 847-864.

205. Liu, L.; Chen, H.; Brecher, M. B.; Li, Z.; Wei, B.; Nandi, B.; Zhang, J.; Ling, H.; Winslow, G.; Braun, J.; Li, H., Pfit Is a Structurally Novel Crohn's Disease-Associated Superantigen. *PLoS pathogens* **2013,** *9* (12), e1003837.

206. Chou, H. H.; Keasling, J. D., Programming Adaptive Control to Evolve Increased Metabolite Production. *Nat Commun* **2013,** *4*, 2595.

207. Cheng, F.; Kardashliev, T.; Pitzler, C.; Shehzad, A.; Lue, H.; Bernhagen, J.; Zhu, L.; Schwaneberg, U., A Competitive Flow Cytometry Screening System for Directed Evolution of Therapeutic Enzyme. *ACS Synth Biol* **2015,** *4* (7), 768-775.

208. Gardner, L.; Zou, Y.; Mara, A.; Cropp, T. A.; Deiters, A., Photochemical Control of Bacterial Signal Processing Using a Light-Activated Erythromycin. *Molecular BioSystems* **2011,** *7* (9), 2554-2557.

209. Cuthbertson, L.; Nodwell, J. R., The Tetr Family of Regulators. *Microbiology and molecular biology reviews : MMBR* **2013,** *77* (3), 440-475.

210. Gardner, L.; Zou, Y.; Mara, A.; Cropp, T. A.; Deiters, A., Photochemical Control of Bacterial Signal Processing Using a Light-Activated Erythromycin. *Molecular bioSystems* **2011,** *7* (9), 2554-2557.

211. Nigam, A.; Almabruk, K. H.; Saxena, A.; Yang, J.; Mukherjee, U.; Kaur, H.; Kohli, P.; Kumari, R.; Singh, P.; Zakharov, L. N.; Singh, Y.; Mahmud, T.; Lal, R., Modification of Rifamycin Polyketide Backbone Leads to Improved Drug Activity against Rifampicin-

Resistant Mycobacterium Tuberculosis. *Journal of Biological Chemistry* **2014,** *289* (30), 21142-21152.

212. Ray, L.; Moore, B. S., Recent Advances in the Biosynthesis of Unusual Polyketide Synthase Substrates. *Nat. Prod. Rep.* **2016,** *33* (2), 150-161.

213. Niquille, D. L.; Hansen, D. A.; Mori, T.; Fercher, D.; Kries, H.; Hilvert, D., Nonribosomal Biosynthesis of Backbone-Modified Peptides. *Nature Chemistry* **2017,** *10*, 282.

214. Klopries, S.; Sundermann, U.; Schulz, F., Quantification of N-Acetylcysteamine Activated Methylmalonate Incorporation into Polyketide Biosynthesis. *Beilstein Journal of Organic Chemistry* **2013,** *9*, 664-674.

215. Seiple, I. B.; Zhang, Z.; Jakubec, P.; Langlois-Mercier, A.; Wright, P. M.; Hog, D. T.; Yabu, K.; Allu, S. R.; Fukuzaki, T.; Carlsen, P. N.; Kitamura, Y.; Zhou, X.; Condakes, M. L.; Szczypiński, F. T.; Green, W. D.; Myers, A. G., A Platform for the Discovery of New Macrolide Antibiotics. *Nature* **2016,** *533*, 338.

216. Reid, R.; Piagentini, M.; Rodriguez, E.; Ashley, G.; Viswanathan, N.; Carney, J.; Santi, D. V.; Hutchinson, C. R.; McDaniel, R., A Model of Structure and Catalysis for Ketoreductase Domains in Modular Polyketide Synthases. *Biochemistry* **2003,** *42* (1), 72-79.

217. Gullon, S.; Olano, C.; Abdelfattah, M. S.; Brana, A. F.; Rohr, J.; Mendez, C.; Salas, J. A., Isolation, Characterization, and Heterologous Expression of the Biosynthesis Gene Cluster for the Antitumor Anthracycline Steffimycin. *Appl Environ Microbiol* **2006,** *72* (6), 4172-4183.

218. Melançon, C. E.; Schultz, P. G., One Plasmid Selection System for the Rapid Evolution of Aminoacyl-Trna Synthetases. *Bioorg. Med. Chem. Lett.* **2009,** *19* (14), 3845-3847.

219. Liu, Y.; Yan, Z.; Lu, X.; Xiao, D.; Jiang, H., Improving the Catalytic Activity of Isopentenyl Phosphate Kinase through Protein Coevolution Analysis. *Scientific reports* **2016,** *6*, 24117.

220. Reiling, K. K.; Yoshikuni, Y.; Martin, V. J.; Newman, J.; Bohlmann, J.; Keasling, J. D., Mono and Diterpene Production in Escherichia Coli. *Biotechnol Bioeng* **2004,** *87* (2), 200-212.

221. Zhao, J.; Bao, X.; Li, C.; Shen, Y.; Hou, J., Improving Monoterpene Geraniol Production through Geranyl Diphosphate Synthesis Regulation in Saccharomyces Cerevisiae. *Appl. Microbiol. Biotechnol.* **2016,** *100* (10), 4561-4571.

222. Zhang, B.; Watts, K. M.; Hodge, D.; Kemp, L. M.; Hunstad, D. A.; Hicks, L. M.; Odom, A. R., A Second Target of the Antimalarial and Antibacterial Agent Fosmidomycin Revealed by Cellular Metabolic Profiling. *Biochemistry* **2011,** *50* (17), 3570-3577.

APPENDICES

Appendix A

Table A1. Rationale and Analysis of All Simulated EryAT6 Mutants.

Mutation(s)	Rationale	MD Result	Agree?	Reference
A677G	emCoA ATs; increase hinge flexibility	Nearly closed active site	✘	This study
A677V	Decrease hinge flexibility	Slightly narrowed active site	✓	This study
A707G	Increase flexibility in small subunit	More flexible small subunit	✓	This study
A745G	Increase YGSH loop flexibility	Similarly-sized active site	✘	This study
D613A	Eliminate H-bond with Arg674 for larger active site	Active site closes in back; no access to Arg669	✘	This study
D613E	emCoA ATs	Very large active site	✓	This study
D743A	Eliminate H-bond with Arg674 for more flexible small subunit	Slight disturbance in small subunit; Arg674 H-bonds with Asp613 exclusively	✓	This study
D743E	Increase residue length for more H-bonding with Arg674	Significant disturbance in small subunit; Arg674 H-bonds with Glu613 exclusively	✘	This study
E647A	Eliminate negative charge near Arg669 to disrupt active site	Active site closes in back; no access to Arg669	✓	This study
L673A	Decrease steric bulk	Larger active site; YASH motif moves into empty space provided by Ala673	✓	This study
L673C	Shown to be more active	Active site remains similar	✘	Koryakina, *ACS Chem Biol* **2017**
L673E	Introduce negative charge in back of active site	Slight disruption in large subunit; Glu673 H-bonds with Asn616	✓	This study
L673G	Shown to be less promiscuous	Larger active site; YASH motif moves into empty space provided by Gly673	✘	Koryakina, *ACS Chem Biol* **2017**
L673I	Minimal change expected	Similarly-sized active site	✓	This study

L673K	Based on similar L673R mutant	Significantly smaller active site	✓	This study
L673M	Minimal change expected	Similarly-sized active site	✓	This study
L673R	Very low activity; only active with methyl	Larger and very unstable active site	✓	Koryakina, *ACS Chem Biol* **2017**
L673S	Introduce uncharged H-bond in back of active site	Loss of access to Arg669 due to H-bonding with Ser673	✓	This study
L673W	Introduce more steric bulk and π-stack with Y744	Smaller active site; no π-stacking	✓/✗	This study
L673Y	Introduce more steric bulk and π-stack with Y744	Tyr673 partially blocks access to Arg669	✓/✗	This study
L740A	Remove steric bulk and increase flexibility of small subunit	Small subunit loop moves inward and rotates; ACP binding might be affected	✓	This study
L740V	Decrease steric bulk at opening	Opened up entrance to active site	✓	This study
L797A	Decrease steric bulk so that residues can move towards back of active site	Smaller pocket; Ser746 moves in front of Ala797	✗/✓	This study
L797H	Introduce polar residue to hydrophobic region of active site	Similarly-sized active site	✗	This study
M682A	Remove steric bulk and increase flexibility of small subunit	Pocket in back of active site	✓	This study
M682C	Remove steric bulk and increase flexibility of small subunit	Large pocket on small subunit side of active site	✓	This study
M682G	Remove steric bulk and increase flexibility of small subunit	Very flexible small subunit; smaller active site	✗/✓	This study
M682H	Remove steric bulk without increasing flexibility	Additional pocket on small subunit side of active site	✓	This study

M682I	Minimal change expected	Similarly-sized active site	✓	This study
M682L	emCoA ATs	Active site becomes much larger	✓	This study
M682S	Remove steric bulk and increase flexibility of small subunit	Unstable; very large active site	✘	This study
M682T	Remove steric bulk and increase flexibility of small subunit	Very large active site; slightly unstable	✘	This study
M682V	Remove steric bulk and increase flexibility of small subunit	Similarly-sized active site	✘	This study
P741I	emCoA ATs	Similarly-sized active site	✘	This study
R674L	emCoA ATs	More flexible small subunit; Leu674 interacts hydrophobically with Val612 and Asp613	✓	This study
S746A	emCoA ATs	Small subunit loop moves towards top of active site	✓	
S746G	emCoA ATs; shown to accept larger substrates	More space near catalytic dyad; minimal change in small subunit	✓	Bravo-Rodriguez, *Chem Biol* **2015**; Koryakina, *ACS Chem Biol* **2017**
S746T	emCoA ATs	Similarly-sized active site	✘	This study
V612A	Remove steric bulk for a tighter fit between subunits	Active site becomes much smaller	✓	This study
V612D	Introduce negative charge to push subunits apart	Subunits separate significantly; potentially unstable	✓	This study
V612G	Increase flexibility in large subunit	Significantly narrows active site	✓	This study
V742G	Shown to be inactive	Unstable; very large active site	✓	Sundermann, *ACS Chem Biol* **2013**
V742P	Decrease flexibility	Narrower active site	✓	This study

V742S	emCoA ATs	Similarly-sized active site	✘	This study
V742T	emCoA ATs	Active site becomes slightly smaller	✘	This study
Y744C	emCoA ATs; shown to be less promiscuous	CASH motif rotates away from active site; larger active site	✓/✘	Koryakina, *ACS Chem Biol* **2017**
Y744C/A745P/S746T	CPTH motif from FkbAT4	More open active site, especially near top of active site	✓	Mo, *J Am Chem Soc* **2011**
Y744E	Reverse charge of Y744R	Similarly-sized active site	✘	This study
Y744F	Remove H-bonding effect from Tyr744	Similarly-sized active site	✓	This study
Y744H	mCoA ATs; more mCoA-permissive	HASH loop moves towards top of active site	✘	Del Vecchio, *J Ind Microbiol Biotechnol* **2003**
Y744H/S746F	mCoA ATs	Phe746 significantly hinders access to His747; potential π-stacking between Phe746 and His744	✓	Del Vecchio, *J Ind Microbiol Biotechnol* **2003**
Y744K	Introduce new H-bond to mimic Y744R	Narrower active site, but Lys744 does H-bond with Asp743	✓/✘	This study
Y744N	Shown to be similar to wild-type	Slightly narrower active site; Asn744 H-bonds with Arg674 backbone	✓	Koryakina, *ACS Chem Biol* **2017**
Y744Q	Less steric hindrance and introduce H-bond	Larger active site; Gln744 H-bonds with Ser746	✓	This study
Y744S	Shown as inactive	Larger active site; mostly undisturbed small subunit	✘	Koryakina, *ACS Chem Biol* **2017**
Y744T	emCoA ATs	TASH loop moves towards top of active site	✓	This study
Y744W	Increase steric hindrance	Disrupts salt bridge between subunits; less room in active site	✓	This study

Table A2. Wild-Type EryAT6 Construct Amino Acid Sequences for MD Simulations.

EryAT6 Construct	Amino Acid FASTA Sequences for MD Simulations
Wild-Type with Both Linkers	AEPPEPEPLPEPGPVGVLAAANSVPVLLSARTETALAAQARLLESAVDDS VPLTALASALATGRAHLPRRAALLAGDHEQLRGQLRAVAEGVAAPGATTG TASAGGVVFVFPGQGAQWEGMARGLLSVPVFAESIAECDAVLSEVAGFSA SEVLEQRPDAPSLERVDVVQPVLFSVMVSLARLWGACGVSPSAVIGHSQG EIAAAVVAGVLSLEDGVRVVALRAKALRALAGKGGMVSLAAPGERARALI APWEDRISVAAVNSPSSVVVSGDPEALAELVARCEDEGVRAKTLPVDYAS HSRHVEEIRETILADLDGISARRAAIPLYSTLHGERRDGADMGPRYWYDN LRSQVRFDEAVSAAVADGHATFVEMSPHPVLTAAVQEIAADAVAIGSLHR DTAEEHLIAELARAHVHGVAVDWRNVFPAA
Wild-Type with N-Terminal Linker	AEPPEPEPLPEPGPVGVLAAANSVPVLLSARTETALAAQARLLESAVDDS VPLTALASALATGRAHLPRRAALLAGDHEQLRGQLRAVAEGVAAPGATTG TASAGGVVFVFPGQGAQWEGMARGLLSVPVFAESIAECDAVLSEVAGFSA SEVLEQRPDAPSLERVDVVQPVLFSVMVSLARLWGACGVSPSAVIGHSQG EIAAAVVAGVLSLEDGVRVVALRAKALRALAGKGGMVSLAAPGERARALI APWEDRISVAAVNSPSSVVVSGDPEALAELVARCEDEGVRAKTLPVDYAS HSRHVEEIRETILADLDGISARRAAIPLYSTLHGERRDGADMGPRYWYDN LRSQVRFDEAVSAAVADGHATFVEMSPHPVLTAAVQE
Wild-Type with No Linkers	VFPGQGAQWEGMARGLLSVPVFAESIAECDAVLSEVAGFSASEVLEQRPD APSLERVDVVQPVLFSVMVSLARLWGACGVSPSAVIGHSQGEIAAAVVAG VLSLEDGVRVVALRAKALRALAGKGGMVSLAAPGERARALIAPWEDRISV AAVNSPSSVVVSGDPEALAELVARCEDEGVRAKTLPVDYASHSRHVEEIR ETILADLDGISARRAAIPLYSTLHGERRDGADMGPRYWYDNLRSQVRFDE AVSAAVADGHATFVEMSPHPVLTAAVQE

Appendix B

Table B1. DEBS PKS DNA FASTA Sequences. AT domains in modules are in bold. Pik docking domain is italicized.

Construct	Nucleotide Sequences
ThaAT13 Motifs	GACCCAATACGCGCCGTCGCTGGAGCGTCTGGATGTGAATCAGCCGGTTC TGTTTAGTGTTATGGTGAGTCTGGCCCGTCTGTGGGGTGCCTGTGGTGTG AGTCCGAGTGCCGTTATTGGCCATAGCCAGGGTGAAATTGCAGCAGCAGT TGTTGCAGGCGTGCTGAGTCTGGAAGATGGTGTGCGCGTGGTTGCACTGC GTGCCAAAGCACTGCGTGCGCTGGCAGGCAAAGGTGGCATGGTTAGCCTG GCCGCACCGGGCGAACGCGCTAGAGCTCTGATTGCCCCGTGGGAAGATCG TATTAGCGTTGCCGCAGTTAATAGTCCGAGTAGCGTGGTGGTGAGTGGTG ACCCGGAAGCCCTGGCCGAACTGGTGGCACGTTGTGAAGATGAAGGTGTT CGCGCAAAAGTTCTGCCGGGTGCCGATGCCGCAGGCCACTCCCGCCACGT CGAGACCCAAT
CinAT1 Motifs	ACGCGCCGTCGCTGGAGGAAGGCGATATTCAGCAGCCGGTTCTGTTTAGT GTGATGGTTAGTCTGGCACGCCTGTGGGGCGCCTGCGGTGTTAGTCCGAG TGCAGTTATTGGCCATAGCCAGGGCGAAATTGCAGCCGCCGTTGTTGCAG GCGTTCTGAGCCTGGAAGATGGCGTTCGTGTTGTGGCCCTGCGCGCCAAA GCACTGCGCGCACTGGCAGGCAAAGGTGGTATGGTGAGTCTGGCAGCACC GGGTGAACGCGCCCGTGCTCTGATTGCACCGTGGGAAGATCGCATTAGCG TGGCCGCAGTGAATAGCCCGAGCAGCGTTGTTGTTAGTGGCGATCCGGAA GCCCTGGCAGAACTGGTGGCCCGTTGCGAAGATGAAGGCGTTCGTGCCAA AGCCCTGCGCGTGGAACGCGCAGGCCACTCCCGCCACGTCGA
Ery6 WT	ATGGTCGGCGCAGCAGAGGCGGAGCAAGCCCCGGCGCTCGTGCGCGAGGT GCCGAAGGATGCCGACGACCCGATCGCGATCGTCGGCATGGCCTGCCGCT TCCCCGGCGGCGTGCACAACCCCGGTGAGCTGTGGGAGTTCATCGTCGGC GGCGGAGACGCCGTGACGGAGATGCCCACCGACCGCGGCTGGGACCTCGA CGCGCTGTTCGACCCCGACCCGCAGCGCCACGGAACCAGCTACTCGCGAC ACGGCGCGTTCCTCGACGGGGCCGCCGACTTCGACGCGGCGTTCTTCGGG ATCTCGCCGCGCGAGGCGCTGGCGATGGACCCGCAGCAGCGCCAGGTCCT GGAAACGACGTGGGAGCTGTTCGAGAACGCCGGCATCGACCCGCACTCGC TGCGGGGCAGCGACACCGGCGTCTTCCTCGGCGCCGCGTACCAGGGCTAC GGCCAGGACGCGGTGGTGCCCGAGGACAGCGAGGGCTACCTGCTCACCGG CAACTCCTCCGCCGTGGTGTCCGGCCGGGTCGCCTACGTGCTGGGGCTGG AAGGCCCCGCGGTCACGGTGGACACGGCGTGTTCGTCGTCGTTGGTGGCC TTGCATTCGGCGTGTGGGTCGTTGCGTGACGGTGACTGCGGTCTTGCGGT GGCCGGTGGTGTGTCGGTGATGGCGGGCCCGGAGGTGTTCACCGAGTTCT CCCGCCAGGGCGGCTTGGCCGTGGACGGGCGCTGCAAGGCGTTCTCCGCG GAGGCCGACGGCTTCGGTTTCGCCGAGGGCGTCGCGGTGGTCCTGCTCCA GCGGTTGTCCGACGCCCGCAGGGCGGGTCGCCAGGTGCTCGGCGTGGTCG CGGGCTCGGCGATCAACCAGGACGGCGCGAGCAACGGTCTCGCGGCGCCG AGCGGCGTCGCCCAGCAGCGCGTGATCCGCAAGGCGTGGGCGCGTGCGGG GATCACGGGCGCGGATGTGGCCGTGGTGGAGGCGCATGGGACCGGTACGC

GGCTGGGCGATCCGGTGGAGGCGTCGGCGTTGCTGGCTACTTACGGCAAG
TCGCGCGGGTCGTCGGGCCCGGTGCTGCTGGGTTCGGTGAAGTCGAACAT
CGGTCACGCGCAGGCGGCCGCGGGTGTCGCGGGCGTGATCAAGGTGGTCC
TGGGGTTGAACCGCGGCCTGGTGCCGCCGATGCTCTGCCGCGGCGAGCGG
TCGCCGCTGATCGAATGGTCCTCGGGTGGTGTGGAACTTGCCGAGGCCGT
GAGCCCGTGGCCTCCGGCCGCGGACGGGGTGCGCCGGGCCGGTGTGTCGG
CGTTCGGGGTGAGC**GGGACGAACGCGCACGTGATCATCGCCGAGCCCCCG**
GAGCCCGAGCCGCTGCCGGAACCCGGACCGGTGGGCGTGCTGGCCGCTGC
GAACTCGGTGCCCGTACTGCTGTCGGCCAGGACCGAGACCGCGTTGGCAG
CGCAGGCGCGGCTCCTGGAGTCCGCAGTGGACGACTCGGTTCCGTTGACG
GCATTGGCTTCCGCGCTGGCCACCGGACGCGCCCACCTGCCGCGTCGTGC
GGCGTTGCTGGCAGGCGACCACGAACAGCTCCGCGGGCAGTTGCGAGCGG
TCGCCGAGGGCGTTGCGGCTCCCGGTGCCACCACCGGAACCGCCTCCGCC
GGCGGCGTGGTTTTCGTCTTCCCAGGTCAGGGTGCTCAGTGGGAGGGCAT
GGCCCGGGGCTTGCTCTCGGTCCCCGTCTTCGCCGAGTCGATCGCCGAGT
GCGATGCGGTGTTGTCGGAGGTGGCCGGGTTCTCGGCCTCCGAAGTGCTG
GAGCAGCGTCCGGACGCGCCGTCGCTGGAGCGGGTCGACGTCGTACAGCC
GGTGTTGTTCTCCGTGATGGTGTCGCTGGCGCGGCTGTGGGGCGCTTGCG
GAGTCAGCCCCTCGGCCGTCATCGGCCATTCGCAGGGCGAGATCGCCGCC
GCGGTGGTGGCCGGGGTGTTGTCGCTGGAGGACGGCGTGCGCGTCGTGGC
CCTGCGCGCGAAGGCGTTGCGTGCGCTGGCGGGCAAGGGCGGCATGGTCT
CGTTGGCGGCTCCCGGTGAACGCGCCCGCGCGCTGATCGCACCGTGGGAG
GACCGGATCTCCGTCGCGGCGGTCAACTCCCCGTCCTCGGTCGTGGTCTC
CGGCGATCCGGAGGCGCTGGCCGAACTCGTCGCACGTTGCGAGGACGAGG
GCGTGCGCGCCAAGACGCTCCCGGTGGACTACGCCTCGCACTCCCGCCAC
GTCGAGGAGATCCGCGAGACGATCCTCGCCGACCTCGACGGCATCTCCGC
GCGGCGTGCCGCCATCCCGCTCTACTCCACGCTGCACGGCGAACGGCGCG
ACGGCGCCGACATGGGTCCGCGGTACTGGTACGACAACCTGCGCTCCCAG
GTGCGCTTCGACGAGGCGGTCTCGGCCGCCGTCGCCGACGGTCACGCCAC
CTTCGTCGAGATGAGCCCGCACCCGGTGCTCACCGCGGCGGTGCAGGAGA
TCGCCGCGGACGCCGTGGCCATCGGGTCGCTGCACCGCGACACCGCGGAG
GAGCACCTGATCGCCGAGCTCGCCCGGGCGCACGTGCACGGCGTGGCCGT
GGACTGGCGGAACGTCTTCCCGGCGGCACCTCCGGTGGCGCTGCCCAACT
ACCCGTTCGAGCCCCAGCGGTACTGGCTCGCGCCGGAGGTGTCCGACCAG
CTCGCCGACAGCCGCTACCGCGTCGACTGGCGACCGCTGGCCACCACGCC
GGTGGACCTGGAAGGCGGCTTCCTGGTCCACGGGTCCGCACCGGAGTCGC
TGACCAGCGCAGTCGAGAAGGCCGGAGGCCGCGTCGTGCCGGTCGCCTCG
GCCGACCGCGAAGCGCTCGCGGCGGCCCTGCGGGAGGTGCCGGGCGAGGT
CGCCGGCGTGCTCTCGGTCCACACCGGCGCCGCAACGCACCTCGCCCTGC
ACCAGTCGCTGGGTGAGGCCGGCGTGCGGGCCCCGCTCTGGCTGGTCACC
AGCCGAGCGGTCGCGCTCGGGGAGTCCGAGCCGGTCGATCCCGAGCAGGC
GATGGTGTGGGGTCTCGGGCGCGTCATGGGCCTGGAGACCCCGGAACGGT
GGGGCGGTCTGGTGGACCTGCCCGCCGAACCCGCGCCGGGGGACGGCGAG
GCGTTCGTCGCCTGCCTCGGCGCGGACGGCCACGAGGACCAGGTCGCGAT
CCGTGACCACGCCCGCTACGGCCGCCGCCTCGTCCGCGCCCCGCTGGGCA
CCCGCGAGTCGAGCTGGGAGCCGGCGGGCACGGCGCTGGTCACCGGCGGC
ACCGGTGCGCTCGGCGGCCACGTCGCCCGCCACCTCGCCAGGTGCGGGGT

	GGAGGACCTGGTGCTGGTCAGCAGGCGCGGCGTCGACGCTCCCGGCGCGG CCGAGCTGGAAGCCGAACTGGTCGCCCTCGGCGCGAAGACGACCATCACC GCCTGCGACGTGGCCGACCGCGAGCAGCTCTCCAAGCTGCTGGAAGAACT GCGCGGGCAGGGACGTCCGGTGCGGACCGTCGTGCACACCGCCGGGGTGC CCGAATCGAGGCCGCTGCACGAGATCGGCGAGCTGGAGTCGGTCTGCGCG GCGAAGGTGACCGGGGCCCGGCTGCTCGACGAGCTGTGCCCGGACGCCGA GACCTTCGTCCTGTTCTCGTCCGGAGCGGGGGTGTGGGGCAGTGCGAACC TCGGCGCCTACTCCGCGGCCAACGCCTACCTCGACGCGCTGGCCCACCGC CGCCGTGCGGAAGGCCGTGCGGCGACGTCCGTCGCGTGGGGCGCCTGGGC GGGCGAGGGCATGGCCACCGGCGACCTCGAGGGGCTCACCCGGCGCGGCC TGCGCCCGATGGCGCCCGAGCGCGCGATCCGCGCGCTGCACCAGGCGCTG GACAACGGCGACACGTGCGTTTCGATCGCCGACGTCGACTGGGAGCGCTT CGCGGTCGGCTTCACCGCCGCCCGGCCGCGTCCGCTGCTGGACGAGCTCG TCACGCCGGCGGTGGGGGCCGTCCCCGCGGTGCAGGCGGCCCCGGCGCGG GAGATGACGTCGCAGGAGTTGCTGGAGTTCACGCACTCGCACGTCGCGGC GATCCTCGGGCATTCCAGCCCGGACGCGGTCGGGCAGGACCAGCCGTTCA CCGAGCTCGGCTTCGACTCGCTGACCGCGGTCGGGCTGCGCAACCAGCTC CAGCAGGCCACCGGGCTCGCGCTGCCCGCGACCCTGGTGTTCGAGCACCC CACGGTCCGCAGGTTGGCCGACCACATAGGACAGCAGCTCTGA
PDDEry6TE **WT**	*ATGACGAGTTCCAACGAACAGTTGGTGGACGCTCTGCGCGCCTCTCTCAA* *GGAGAACGAAGAACTCCGGAAAGAGAGCCGTCGCCGGGCCGACCGTCGGC* *AGGAGGAG*ATCGCGATCGTCGGCATGGCCTGCCGCTTCCCCGGCGGCGTG CACAACCCCGGTGAGCTGTGGGAGTTCATCGTCGGCGGCGGAGACGCCGT GACGGAGATGCCCACCGACCGCGGCTGGGACCTCGACGCGCTGTTCGACC CCGACCCGCAGCGCCACGGAACCAGCTACTCGCGACACGGCGCGTTCCTC GACGGGGCCGCCGACTTCGACGCGGCGTTCTTCGGGATCTCGCCGCGCGA GGCGCTGGCGATGGACCCGCAGCAGCGCCAGGTCCTGGAAACGACGTGGG AGCTGTTCGAGAACGCCGGCATCGACCCGCACTCGCTGCGGGGCAGCGAC ACCGGCGTCTTCCTCGGCGCCGCGTACCAGGGCTACGGCCAGGACGCGGT GGTGCCCGAGGACAGCGAGGGCTACCTGCTCACCGGCAACTCCTCCGCCG TGGTGTCCGGCCGGGTCGCCTACGTGCTGGGGCTGGAAGGCCCCGCGGTC ACGGTGGACACGGCGTGTTCGTCGTCGTTGGTGGCCTTGCATTCGGCGTG TGGGTCGTTGCGTGACGGTGACTGCGGTCTTGCGGTGGCCGGTGGTGTGT CGGTGATGGCGGGCCCGGAGGTGTTCACCGAGTTCTCCCGCCAGGGCGGC TTGGCCGTGGACGGGCGCTGCAAGGCGTTCTCCGCGGAGGCCGACGGCTT CGGTTTCGCCGAGGGCGTCGCGGTGGTCCTGCTCCAGCGGTTGTCCGACG CCCGCAGGGCGGGTCGCCAGGTGCTCGGCGTGGTCGCGGGCTCGGCGATC AACCAGGACGGCGCGAGCAACGGTCTCGCGGCGCCGAGCGGCGTCGCCCA GCAGCGCGTGATCCGCAAGGCGTGGGCGCGTGCGGGGATCACGGGCGCGG ATGTGGCCGTGGTGGAGGCGCATGGGACCGGTACGCGGCTGGGCGATCCG GTGGAGGCGTCGGCGTTGCTGGCTACTTACGGCAAGTCGCGCGGGTCGTC GGGCCCGGTGCTGCTGGGTTCGGTGAAGTCGAACATCGGTCACGCGCAGG CGGCCGCGGGTGTCGCGGGCGTGATCAAGGTGGTCCTGGGGTTGAACCGC GGCCTGGTGCCGCCGATGCTCTGCCGCGGCGAGCGGTCGCCGCTGATCGA ATGGTCCTCGGGTGGTGTGGAACTTGCCGAGGCCGTGAGCCCGTGGCCTC CGGCCGCGGACGGGGTGCGCCGGGCCGGTGTGTCGGCGTTCGGGGTGAGC **GGGACGAACGCGCACGTGATCATCGCCGAGCCCCCGGAGCCCGAGCCGCT**

GCCGGAACCCGGACCGGTGGGCGTGCTGGCCGCTGCGAACTCGGTGCCCG
TACTGCTGTCGGCCAGGACCGAGACCGCGTTGGCAGCGCAGGCGCGGCTC
CTGGAGTCCGCAGTGGACGACTCGGTTCCGTTGACGGCATTGGCTTCCGC
GCTGGCCACCGGACGCGCCCACCTGCCGCGTCGTGCGGCGTTGCTGGCAG
GCGACCACGAACAGCTCCGCGGGCAGTTGCGAGCGGTCGCCGAGGGCGTT
GCGGCTCCCGGTGCCACCACCGGAACCGCCTCCGCCGGCGGCGTGGTTTT
CGTCTTCCCAGGTCAGGGTGCTCAGTGGGAGGGCATGGCCCGGGGCTTGC
TCTCGGTCCCCGTCTTCGCCGAGTCGATCGCCGAGTGCGATGCGGTGTTG
TCGGAGGTGGCCGGGTTCTCGGCCTCCGAAGTGCTGGAGCAGCGTCCGGA
CGCGCCGTCGCTGGAGCGGGTCGACGTCGTACAGCCGGTGTTGTTCTCCG
TGATGGTGTCGCTGGCGCGGCTGTGGGGCGCTTGCGGAGTCAGCCCCTCG
GCCGTCATCGGCCATTCGCAGGGCGAGATCGCCGCCGCGGTGGTGGCCGG
GGTGTTGTCGCTGGAGGACGGCGTGCGCGTCGTGGCCCTGCGCGCGAAGG
CGTTGCGTGCGCTGGCGGGCAAGGGCGGCATGGTCTCGTTGGCGGCTCCC
GGTGAACGCGCCCGCGCGCTGATCGCACCGTGGGAGGACCGGATCTCCGT
CGCGGCGGTCAACTCCCCGTCCTCGGTCGTGGTCTCCGGCGATCCGGAGG
CGCTGGCCGAACTCGTCGCACGTTGCGAGGACGAGGGCGTGCGCGCCAAG
ACGCTCCCGGTGGACTACGCCTCGCACTCCCGCCACGTCGAGGAGATCCG
CGAGACGATCCTCGCCGACCTCGACGGCATCTCCGCGCGGCGTGCCGCCA
TCCCGCTCTACTCCACGCTGCACGGCGAACGGCGCGACGGCGCCGACATG
GGTCCGCGGTACTGGTACGACAACCTGCGCTCCCAGGTGCGCTTCGACGA
GGCGGTCTCGGCCGCCGTCGCCGACGGTCACGCCACCTTCGTCGAGATGA
GCCCGCACCCGGTGCTCACCGCGGCGGTGCAGGAGATCGCCGCGGACGCC
GTGGCCATCGGGTCGCTGCACCGCGACACCGCGGAGGAGCACCTGATCGC
CGAGCTCGCCCGGGCGCACGTGCACGGCGTGGCCGTGGACTGGCGGAACG
TCTTCCCGGCGGCACCTCCGGTGGCGCTGCCCAACTACCCGTTCGAGCCC
CAGCGGTACTGGCTCGCGCCGGAGGTGTCCGACCAGCTCGCCGACAGCCG
CTACCGCGTCGACTGGCGACCGCTGGCCACCACGCCGGTGGACCTGGAAG
GCGGCTTCCTGGTCCACGGGTCCGCACCGGAGTCGCTGACCAGCGCAGTC
GAGAAGGCCGGAGGCCGCGTCGTGCCGGTCGCCTCGGCCGACCGCGAAGC
GCTCGCGGCGGCCCTGCGGGAGGTGCCGGGCGAGGTCGCCGGCGTGCTCT
CGGTCCACACCGGCGCCGCAACGCACCTCGCCCTGCACCAGTCGCTGGGT
GAGGCCGGCGTGCGGGCCCCGCTCTGGCTGGTCACCAGCCGAGCGGTCGC
GCTCGGGGAGTCCGAGCCGGTCGATCCCGAGCAGGCGATGGTGTGGGGTC
TCGGGCGCGTCATGGGCCTGGAGACCCCGGAACGGTGGGGCGGTCTGGTG
GACCTGCCCGCCGAACCCGCGCCGGGGGACGGCGAGGCGTTCGTCGCCTG
CCTCGGCGCGGACGGCCACGAGGACCAGGTCGCGATCCGTGACCACGCCC
GCTACGGCCGCCGCCTCGTCCGCGCCCCGCTGGGCACCCGCGAGTCGAGC
TGGGAGCCGGCGGGCACGGCGCTGGTCACCGGCGGCACCGGTGCGCTCGG
CGGCCACGTCGCCCGCCACCTCGCCAGGTGCGGGGTGGAGGACCTGGTGC
TGGTCAGCAGGCGCGGCGTCGACGCTCCCGGCGCGGCCGAGCTGGAAGCC
GAACTGGTCGCCCTCGGCGCGAAGACGACCATCACCGCCTGCGACGTGGC
CGACCGCGAGCAGCTCTCCAAGCTGCTGGAAGAACTGCGCGGGCAGGGAC
GTCCGGTGCGGACCGTCGTGCACACCGCCGGGGTGCCCGAATCGAGGCCG
CTGCACGAGATCGGCGAGCTGGAGTCGGTCTGCGCGGCGAAGGTGACCGG
GGCCCGGCTGCTCGACGAGCTGTGCCCGGACGCCGAGACCTTCGTCCTGT
TCTCGTCCGGAGCGGGGGTGTGGGGCAGTGCGAACCTCGGCGCCTACTCC

	GCGGCCAACGCCTACCTCGACGCGCTGGCCCACCGCCGCCGTGCGGAAGG CCGTGCGGCGACGTCCGTCGCGTGGGGCGCCTGGGCGGGCGAGGGCATGG CCACCGGCGACCTCGAGGGGCTCACCCGGCGCGGCCTGCGCCCGATGGCG CCCGAGCGCGCGATCCGCGCGCTGCACCAGGCGCTGGACAACGGCGACAC GTGCGTTTCGATCGCCGACGTCGACTGGGAGCGCTTCGCGGTCGGCTTCA CCGCCGCCCGGCCGCGTCCGCTGCTGGACGAGCTCGTCACGCCGGCGGTG GGGGCCGTCCCCGCGGTGCAGGCGGCCCCGGCGCGGGAGATGACGTCGCA GGAGTTGCTGGAGTTCACGCACTCGCACGTCGCGGCGATCCTCGGGCATT CCAGCCCGGACGCGGTCGGGCAGGACCAGCCGTTCACCGAGCTCGGCTTC GACTCGCTGACCGCGGTCGGGCTGCGCAACCAGCTCCAGCAGGCCACCGG GCTCGCGCTGCCCGCGACCCTGGTGTTCGAGCACCCCACGGTCCGCAGGT TGGCCGACCACATAGGACAGCAGCTCGACAGCGGGACTCCCGCCCGGGAA GCGAGCAGCGCTCTTCGCGACGGCTACCGGCAGGCGGGCGTGTCGGGCAG GGTCCGGTCCTACCTCGACCTGCTGGCGGGGCTGTCGGACTTCCGCGAGC ACTTCGACGGCTCCGACGGGTTCTCCCTCGATCTCGTGGACATGGCCGAC GGTCCCGGAGAGGTCACGGTGATCTGCTGCGCGGGAACGGCGGCGATCTC CGGTCCGCACGAGTTCACCCGGCTCGCCGGGGCGCTGCGCGGAATCGCTC CGGTTCGGGCCGTGCCCCAGCCCGGCTACGAGGAGGGCGAACCTCTGCCG TCGTCGATGGCGGCGGTGGCGGCGGTGCAGGCCGATGCGGTCATCAGGAC ACAGGGGGACAAGCCGTTCGTGGTGGCCGGTCACTCCGCGGGGGCACTGA TGGCCTACGCGCTGGCGACCGAACTGCTCGATCGCGGGCACCCGCCACGC GGTGTCGTCCTGATCGACGTCTACCCGCCCGGTCACCAGGACGCGATGAA CGCCTGGCTGGAGGAGCTGACCGCCACGCTGTTCGACCGCGAGACGGTGC GGATGGACGACACCAGGCTCACCGCCCTGGGCGCCTACGACCGCCTCACC GGTCAGTGGCGACCCCGGGAAACCGGGCTGCCGACGCTGCTGGTCAGCGC CGGCGAGCCGATGGGTCCGTGGCCCGACGACAGCTGGAAGCCGACGTGGC CCTTCGAGCACGACACCGTCGCCGTCCCCGGCGACCACTTCACGATGGTG CAGGAACACGCCGACGCGATCGCGCGGCACATCGACGCCTGGCTGGGCGG AGGGAATTCAAGA

Table B2. Primers for DEBS PKS Construction and Mutagenesis. GA = Gibson Assembly. RTH = 'Round-the-horn mutagenesis. SDM = Site-directed mutagenesis. RTH primers have 5' phosphorylated primers.

Primer Name	5'→3' Primer Sequence	Primer Function
AT_Motif.Gib1	CACTCCCGCCACGTCGAG	Amplifies Ery6TE without AT motifs for GA
AT_Motif.Gib2	CTCCAGCGACGGCGCG	
AT_Motif.FOR	ACGCGCCGTCGCTGG	Amplifies motif fragments for GA
AT_Motif.REV	TCGACGTGGCGGGAGTG	
Ery6V742A.FOR	GCGGACTACGCCTCGCACTCCC	Ery6 V742A RTH
Ery6V742A.REV	CGGGAGCGTCTTGGCGC	
Ery6Y744RGV.RTH1	SKGGCCTCGCACTCCCGC	Ery6TE Y744R/G RTH
Ery6Y744RGV.RTH2	GTCCACCGGGAGCGTCTTG	

Ery6M682X.FOR	NNKGTCTCGTTGGCGGCT	Ery6 M682X RTH
Ery6M682X.REV	GCCGCCCTTGCCCGC	
Ery6L676V.FOR	GTGGCGGGCAAGGGCG	Ery6 L676V RTH
Ery6L676V.REV	CGCACGCAACGCCTTCG	
Ery6L673G.FOR	GGACGTGCGCTGGCGGG	Ery6 L673G RTH
Ery6L673G.REV	CGCCTTCGCGCGCAG	
Ery6E647D.FOR	GATATCGCCGCCGCGG	Ery6 E647D RTH
Ery6E647D.REV	GCCCTGCGAATGGCCGAT	
Ery6L797X.FOR	NNKCGCTCCCAGGTGCGC	Ery6 L797X RTH
Ery6L797X.REV	GTTGTCGTACCAGTACCGCGGA	
Ery6I648A.FOR	GCCGCCGCGGTGGTGG	Ery6 I648G/A/V RTH
Ery6I648X.REV	GNSCTCGCCCTGCGAATGGC	
Ery6H643F.FOR	TTTTCGCAGGGCGAGATCGC	Ery6 H643F RTH
Ery6H643F.REV	GCCGATGACGGCCGAGG	
Ery6H643S.FOR	TCTTCGCAGGGCGAGATCGC	Ery6 H643S RTH
Ery6H643F.REV	GCCGATGACGGCCGAGG	
Ery6L740V.FOR	GTCCCGGTGGACTACGCCTC	Ery6 L740V RTH
Ery6L740V.REV	CGTCTTGGCGCGCACG	
Ery6A707G.FOR	GGGGTCAACTCCCCGTCCTCG	Ery6 A707G RTH
Ery6A707G.REV	CGCGACGGAGATCCGGTC	
Ery6G681A.FOR	GCAATGGTCTCGTTGGCGGC	Ery6 G681A RTH
Ery6G681A.REV	GCCCTTGCCCGCCAGC	
Ery6V751A.FOR	GCCGAGGAGATCCGCGAGAC	Ery6 V751A RTH
Ery6V751A.REV	GTGGCGGGAGTGCGAGG	
Ery6S644A.FOR	GTCAGCCGTTATCGGTCATGCT	Ery6 S644A RTH
Ery6S644A.REV	CCGCAATTTCGCCCTG	
Ery6D613E.FOR	GAAGTCGTACAGCCGGTGTTG	Ery6 D613E RTH
Ery6D613E.REV	GACCCGCTCCAGCGACG	
Ery6V751A.FOR	GCCGAGGAGATCCGCGAGAC	Ery6 V751A RTH
Ery6V751A.REV	GTGGCGGGAGTGCGAGG	

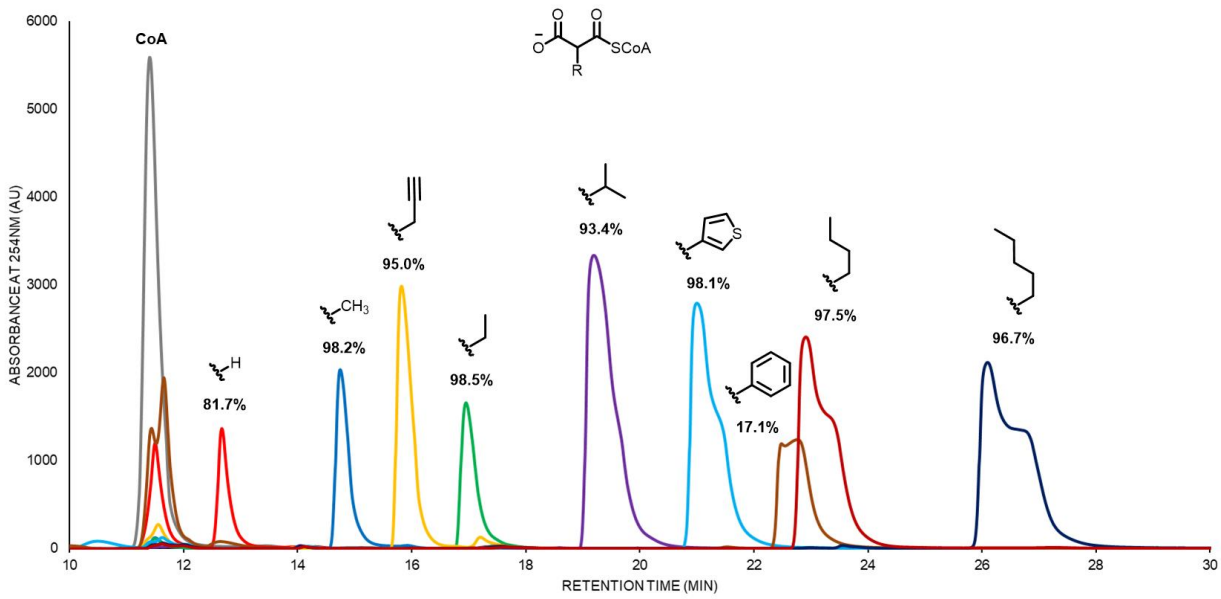


Figure B1. Representative HPLC Traces for MatB-synthesized Malonyl-CoAs. Percent conversions were determined by remaining CoA peak.

Table B3. Distributions of Reduced and Non-Reduced Products by Substrate. Module lysates were incubated with and products were derived the following substrates: **1a**, **1b**, **1c**, **1d**, **1e**, **1g**, and **1i**. Products were analyzed by HR-LCMS. Reduced products are **4x** and non-reduced products are **5x**.

Substrate	Percent Product 5x
1a	0.0 ± 0.0%
1b	49.1 ± 14.1%
1c	44.9 ± 15.2%
1d	80.0 ± 23.4%
1e	100.0 ± 0.0%
1g	100.0 ± 0.0%
1i	69.1 ± 5.9%

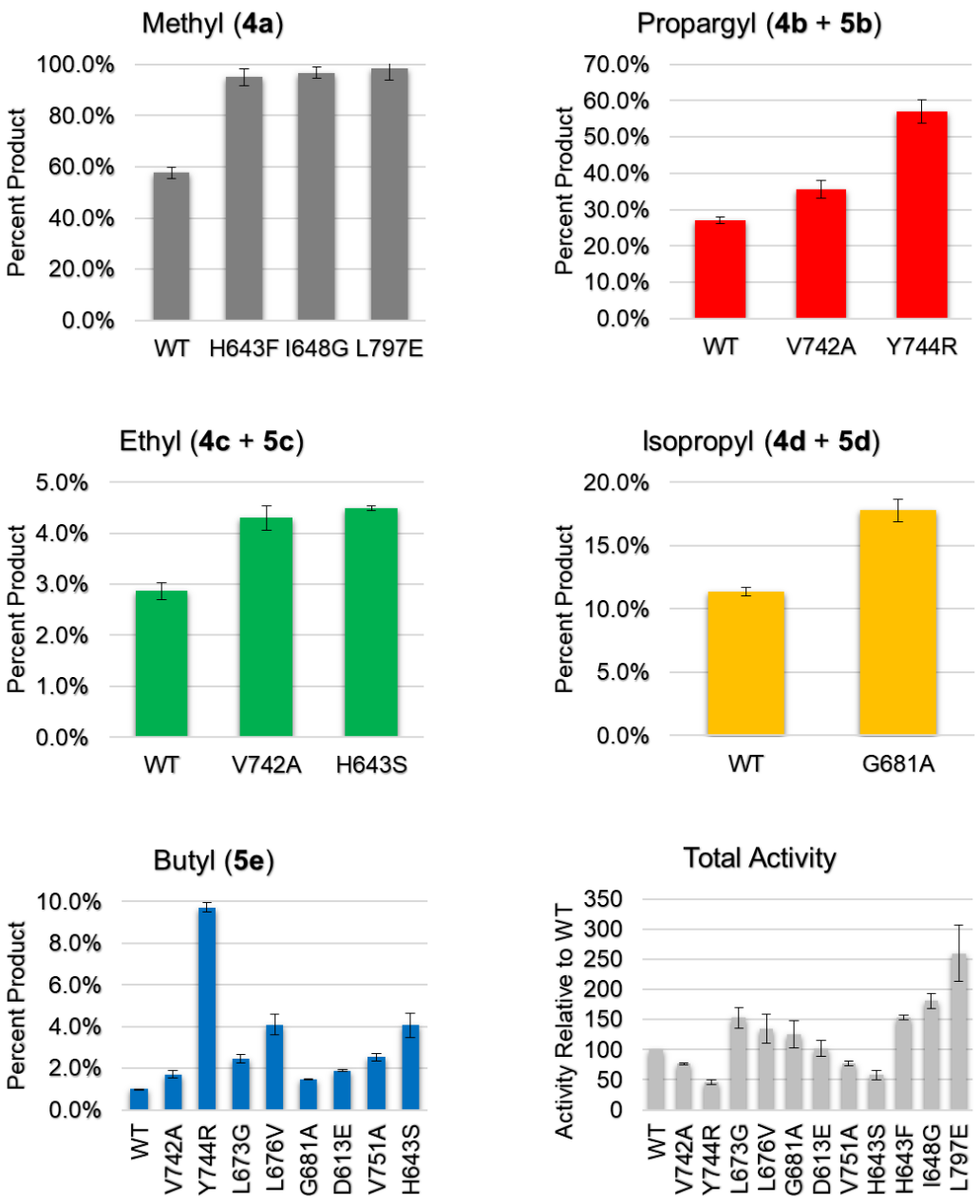


Figure B2. Assaying Ery6TE Mutants with Non-Natural Extender Unit Panel. Module lysates were incubated with 500 μM each of **1a**, **1b**, **1c**, **1d**, **1e**, and **1h** simultaneously. Products were analyzed by HR-LCMS and notable results are shown. Total activities were compared to the average wild-type activity. Full results shown in **Table B3**.

Table B4. Product Distributions of Module Lysates with Six Substrates. Module lysates were incubated with 500 μM each of **1a**, **1b**, **1c**, **1d**, **1e**, and **1h** simultaneously. Products were analyzed by HR-LCMS and notable results are shown. Total activities were compared to the average wild-type activity.

Ery6TE Variant	**4a**	**4b** + **5b**	**4c** + **5c**	**4d** + **5d**	**5e**	Relative Activity
WT	57.7 ± 2.2%	27.1 ± 0.9%	2.9 ± 0.2%	11.4 ± 0.4%	1.0 ± 0.0%	100.0
V742A	56.0 ± 2.4%	35.6 ± 2.4%	4.3 ± 0.2%	2.4 ± 0.2%	1.7 ± 0.2%	76.8 ± 1.8
Y744R	17.7 ± 1.2%	57.0 ± 3.2%	2.5 ± 0.1%	13.0 ± 1.9%	9.7 ± 0.2%	46.2 ± 4.1
L673G	65.7 ± 2.2%	22.9 ± 1.6%	2.5 ± 0.4%	6.5 ± 0.9%	2.4 ± 0.2%	152.6 ± 17.6
L676V	73.4 ± 1.6%	7.8 ± 0.9%	3.1 ± 0.1%	11.7 ± 1.0%	4.1 ± 0.5%	135.1 ± 23.9
G681A	59.5 ± 2.0%	18.6 ± 0.8%	2.7 ± 0.4%	17.8 ± 1.2%	1.5 ± 0.0%	124.6 ± 22.6
D613E	56.4 ± 4.6%	25.9 ± 3.1%	2.5 ± 0.4%	13.4 ± 1.1%	1.9 ± 0.0%	101.9 ± 13.3
V751A	54.3 ± 2.8%	29.7 ± 2.3%	2.7 ± 0.3%	10.7 ± 0.2%	2.5 ± 0.2%	76.5 ± 3.7
H643S	52.9 ± 2.7%	25.5 ± 2.3%	4.5 ± 0.1%	13.1 ± 1.5%	4.1 ± 0.1%	57.6 ± 7.0
H643F	95.0 ± 2.9%	1.9 ± 0.2%	1.1 ± 0.0%	1.7 ± 0.2%	0.3 ± 0.0%	153.5± 4.3
I648G	96.7 ± 2.1%	0.8 ± 0.1%	0.4 ± 0.1%	1.7 ± 0.2%	0.4 ± 0.0%	180.8 ± 12.8
L797E	98.4 ± 1.3%	0.5 ± 0.0%	0.5 ± 0.1%	0.3 ± 0.0%	0.3 ± 0.0%	259.6 ± 47.0
Cin1	16.3 ± 0.2%	55.8 ± 5.9%	0.6 ± 0.1%	6.9 ± 0.5%	20.3 ± 2.7%	35.9 ± 5.7
Tha13	11.2 ± 0.2%	18.3 ± 0.8%	0.9 ± 0.0%	23.2 ± 3.0%	46.5 ± 2.9%	30.1 ± 1.4

Table B5. High-Resolution LC-MS Parameters and Gradient.

HESI Source Parameters	
Spray Voltage	3.5 kV
Capillary Temperature	350 °C
Heater Temperature	300 °C
S Lens RF Level	70 V
Sheath Gas Flow Rate	60 au
Resolution	70,000 FWHM
Scan Range	100-1000 m/z

LC Gradient

Time (min)	% B
0.0	25.0
1.0	25.0
10.0	59.0
11.0	85.0
12.0	85.0
12.5	25.0
16.5	25.0

Table B6. High-Resolution LC-MS Retention Times, Calculated Masses, and Observed Masses for DEBS PKS-Catalyzed Reaction Products. N.D. = Not Detected.

Compound & Retention Time (min)		Calculated Mass ($[M-H_2O+H]^+$)	Observed Mass ($[M-H_2O+H]^+$)	Calculated Mass ($[M+H]^+$)	Observed Mass ($[M+H]^+$)	Calculated Mass ($[M+Na]^+$)	Observed Mass ($[M+Na]^+$)
4a	8.03	279.1955	279.1962	297.2060	297.2069	319.1880	319.1904
4b	8.96	303.1955	303.1963	321.2060	321.2068	343.1880	343.1890
4c	9.68	293.2111	293.2119	311.2217	311.2227	333.2036	333.2044
4d	10.63	307.2268	307.2274	325.2373	325.2378	347.2193	347.2198
4e	N.D.	321.2424	N.D.	339.2530	N.D.	361.2349	N.D.
4f	N.D.	347.1675	N.D.	365.1781	N.D.	387.1601	N.D.
4g	N.D.	335.2581	N.D.	353.2686	N.D.	375.2506	N.D.
4h	N.D.	361.1832	N.D.	379.1938	N.D.	401.1757	N.D.
4i	10.12	305.2111	305.2105	323.2217	323.2201	345.2036	345.2035
5a	N.D.	--	--	295.1904	N.D.	317.1723	N.D.
5b	11.80	--	--	319.1904	319.1911	341.1723	341.1732
5c	12.39	--	--	309.2060	309.2067	331.1880	331.1886
5d	11.79	--	--	323.2217	323.2224	345.2036	345.2044

5e	14.56	--	--	337.2373	337.2380	359.2193	359.2200
5f	N.D.	--	--	363.1625	N.D.	385.1444	N.D.
5g	15.07	--	--	351.2530	351.2525	373.2349	373.2345
5h	N.D.	--	--	377.1781	N.D.	399.1601	N.D.
5i	12.67	--	--	321.2060	321.2055	343.1880	343.1876

Table B7. High-Resolution LC-MS Peak Areas for PKS-Catalyzed Reaction Products with 6 Non-Natural Extenders. Extracted ion count peak areas from one replicate are shown for each reaction condition. N.D. = Not Detected.

Ery6TE Variant	**Representative EIC Peak Areas**				
	4a	**4b**	**5b**	**4c**	**5c**
WT	211,060,257	60,428,975	33,336,885	8,350,707	3,201,023
V742A	1,074,530,536	621,019,227	62,604,070	74,181,724	8,289,132
Y744R	201,560,426	425,370,308	221,736,187	18,273,398	10,450,848
L673G	412,033,607	96,211,067	47,618,973	11,485,854	3,946,441
A672S	918,117,379	174,246,461	190,537,319	17,567,468	16,886,399
L676V	913,134,214	202,357,895	254,127,057	20,437,965	21,610,476
G681A	258,612,993	42,211,949	38,450,516	7,211,911	4,379,076
D613E	1,112,681,260	246,571,818	263,851,451	25,901,052	23,610,928
R674A	1,137,937,000	178,486,582	202,512,483	22,945,009	25,007,975
H643S	244,645,247	70,681,023	47,042,985	12,743,255	8,023,759
H643F	215,279,432	2,131,336	2,185,209	1,629,754	960,271
I648G	365,230,755	1,770,386	1,262,385	869,632	475,403
L797E	565,835,459	829,163	1,769,043	782,613	2,352,056
Cin1	1,122,551	1,093,275	2,750,716	14,392	29,947
Tha13	29,754,912	31,038,395	17,687,295	1,565,605	928,229

Ery6TE Variant	Representative EIC Peak Areas					
	4d	**5d**	**4e**	**5e**	**4h**	**5h**
WT	11,975,213	22,923,302	N.D.	3,401,711	N.D.	N.D.
V742A	41,645,025	4,688,419	N.D.	32,702,810	N.D.	N.D.
Y744R	87,834,327	60,330,553	N.D.	110,340,604	N.D.	N.D.
L673G	12,692,412	27,903,903	N.D.	15,313,652	N.D.	N.D.
A672S	7,799,155	195,172,467	N.D.	18,325,167	N.D.	N.D.
L676V	2,065,221	161,120,479	N.D.	28,957,619	N.D.	N.D.
G681A	13,320,019	63,892,775	N.D.	6,301,166	N.D.	N.D.
D613E	20,124,173	244,569,998	N.D.	37,219,146	N.D.	N.D.
R674A	16,724,087	256,290,037	N.D.	19,347,354	N.D.	N.D.
H643S	25,063,933	35,489,999	N.D.	18,784,167	N.D.	N.D.
H643F	74,197	3,727,611	N.D.	589,128	N.D.	N.D.
I648G	595,456	6,006,129	N.D.	1,376,049	N.D.	N.D.
L797E	128,196	1,623,863	N.D.	1,573,907	N.D.	N.D.
Cin1	39,917	439,077	N.D.	1,402,533	N.D.	N.D.
Tha13	22,492,383	39,231,371	N.D.	123,779,145	N.D.	N.D.

Table B8. High-Resolution LC-MS Peak Areas for PKS-Catalyzed Reaction Products with 2 Extenders. Extracted ion count peak areas from one replicate are shown for each reaction condition. N.D. = Not Detected.

		Representative EIC Peak Areas					
Substrates	Product	WT	Y744G	Y744R	I648G	Tha13	Cin1
1a/**1i**	**4a**	1,263,268,254	305,773,861	122,464,111	1,467,946,423	43,610,053	45,392,299
	4i	7,647,855	16,447,345	28,872,748	N.D.	10,197,042	964,634
	5i	21,784,302	25,330,603	53,689,472	N.D.	33,024,254	17,802,789
1a/**1b**	**4a**	532,578,041	191,683,831	78,684,251	799,908,718	26,068,751	29,848,810
	4b	31,090,929	286,056,727	320,918,411	2,583,102	45,737,083	20,904,574
	5b	75,186,842	135,129,518	109,088,198	N.D.	37,497,560	44,368,738
1a/**1e**	**4a**	1,241,364,749	40,688,455	230,399,192	650,111,454	29,616,734	478,245
	4e	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
	5e	6,953,819	21,957,829	88,178,801	N.D.	193,602,896	4,152,337

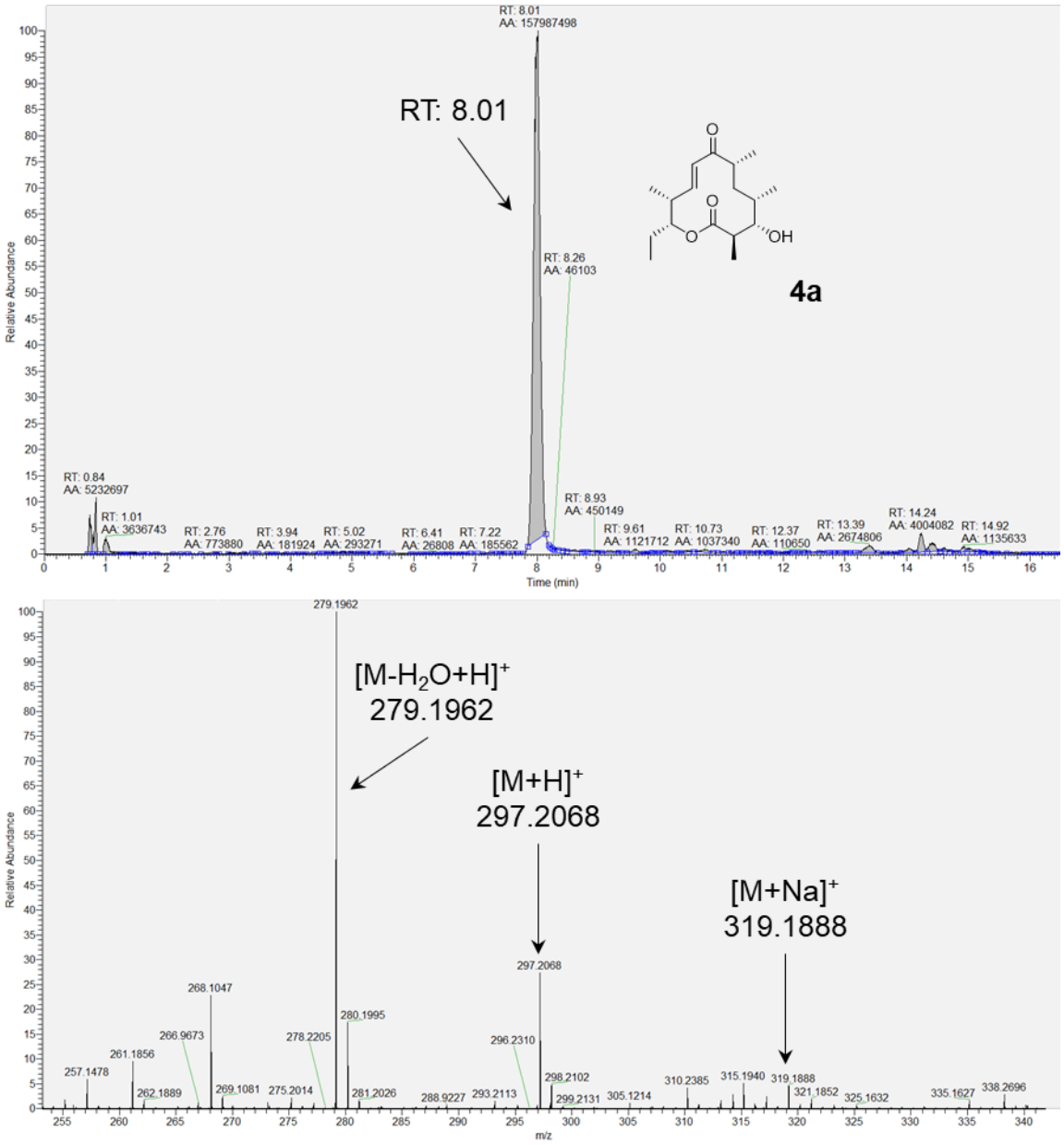
1a/**1g**	**4a**	1,144,743,869	905,709,691	622,633,725	1,115,568,669	30,377,170	1,042,802,675
	4g	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
	5g	1,583,892	263,648,315	217,518,820	N.D.	473,073,518	4,458,627,491

Table B9. High-Resolution LC-MS Peak Areas for PKS-Catalyzed Reaction Products with 2 Non-Natural Extenders. Extracted ion count peak areas from one replicate are shown for each reaction condition. N.D. = Not Detected.

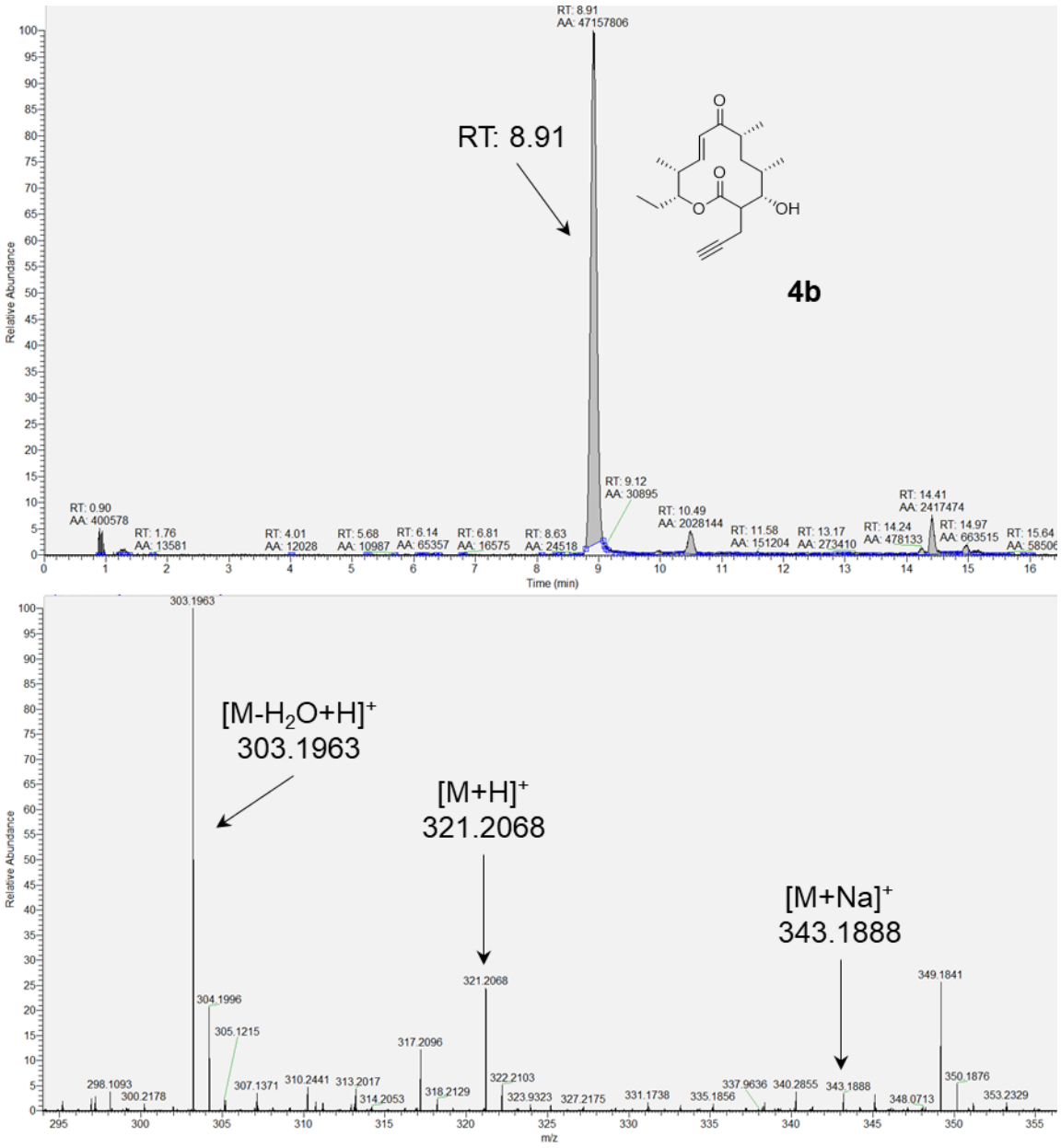
		Representative EIC Peak Areas		
Enzyme	**1a/1b Concentrations**	**4a**	**4b**	**5b**
Ery6TE WT	0 μM	N.D.	N.D.	N.D.
	25 μM	14,110,402	18,677,899	5,072,464
	75 μM	51,916,999	65,440,367	15,171,009
	150 μM	130,676,020	22,705,683	56,951,555
	300 μM	330,487,368	360,107,712	79,984,554
	1500 μM	1,383,958,399	328,868,153	106,042,033

Figure B3. Representative LC-MS Chromatograms of Products from Lysate Module-Catalyzed Reactions. Top panel for each compound shows the extracted ion chromatogram. Bottom panel shows the total ion spectra.

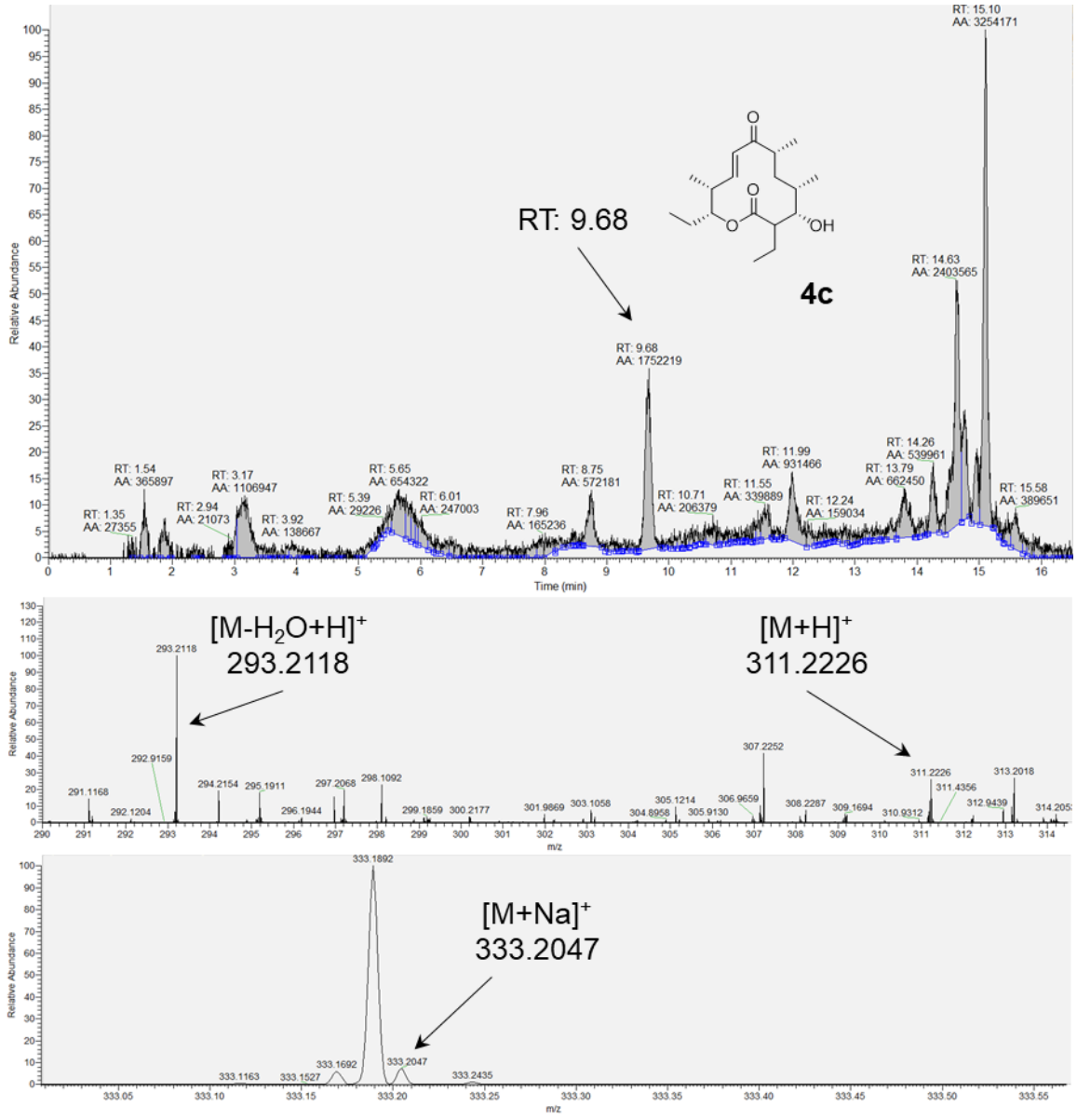
A



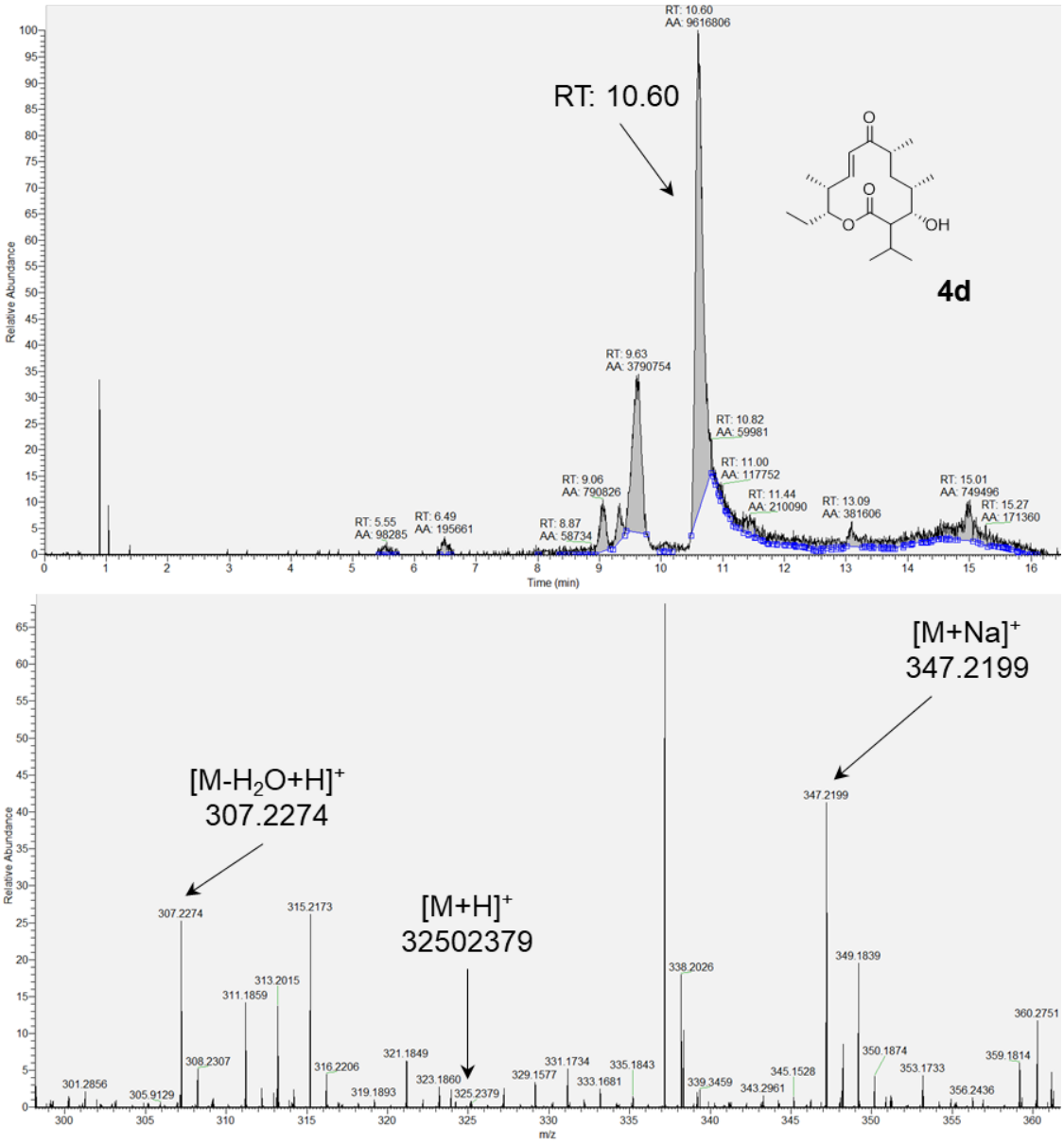
B


RT: 8.91
AA: 47157806
RT: 8.91
4b
Relative Abundance
Time (min)
RT: 0.90
AA: 400578
RT: 1.76
AA: 13581
RT: 4.01
AA: 12028
RT: 5.68
AA: 10987
RT: 6.14
AA: 65357
RT: 6.81
AA: 16575
RT: 8.63
AA: 24518
RT: 9.12
AA: 30895
RT: 10.49
AA: 2028144
RT: 11.58
AA: 151204
RT: 13.17
AA: 273410
RT: 14.24
AA: 478133
RT: 14.41
AA: 2417474
RT: 14.97
AA: 663515
[M-H2O+H]+
303.1963
[M+H]+
321.2068
[M+Na]+
343.1888
m/z

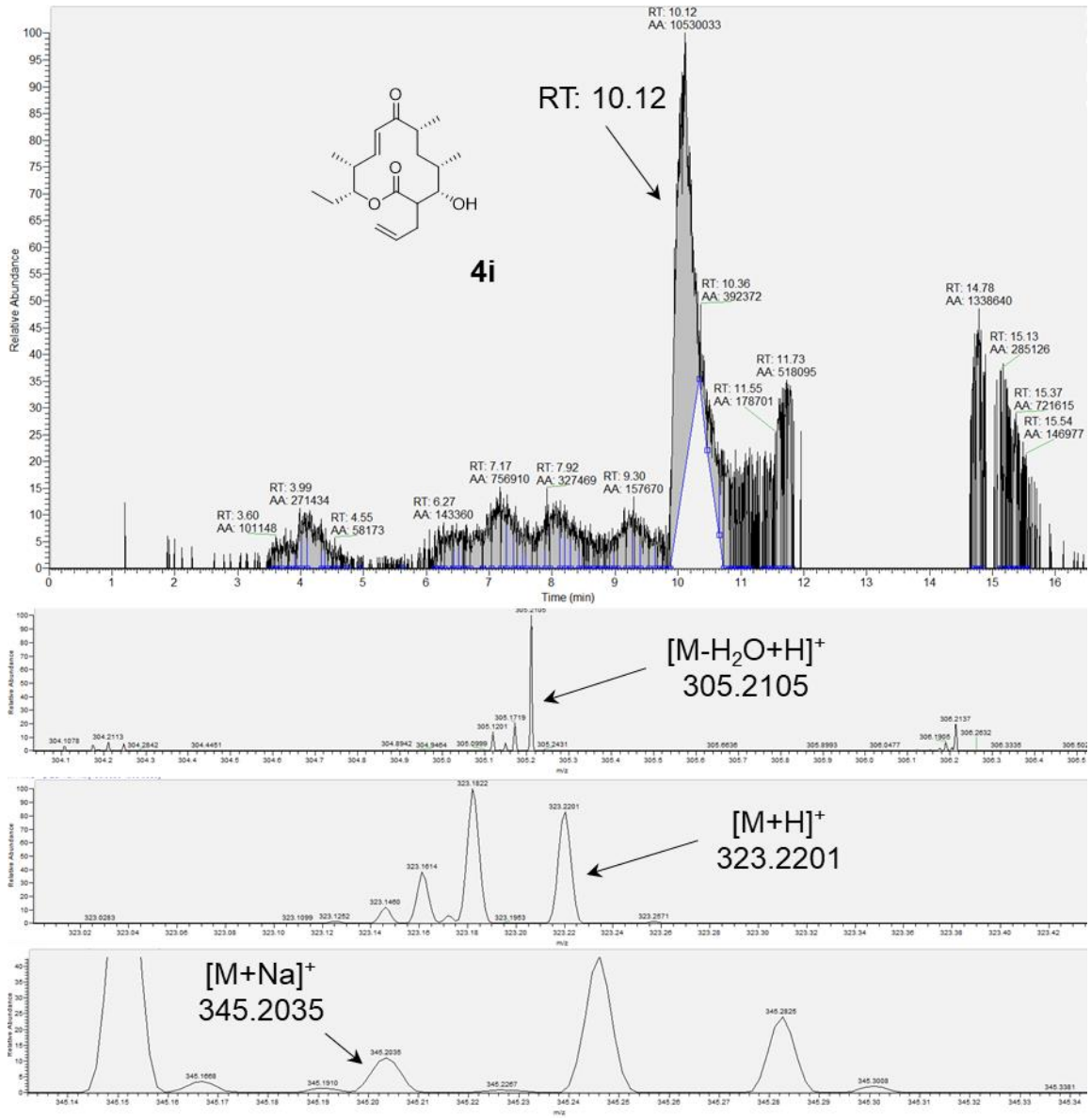
C


RT: 9.68
4c
RT: 15.10
AA: 3254171
RT: 14.63
AA: 2403565
RT: 9.68
AA: 1752219
RT: 1.54
AA: 365897
RT: 3.17
AA: 1106947
RT: 5.65
AA: 654322
RT: 8.75
AA: 572181
RT: 11.99
AA: 931466
RT: 14.26
AA: 539961
RT: 13.79
AA: 662450
RT: 15.58
AA: 389651
RT: 1.35
AA: 27355
RT: 2.94
AA: 21073
RT: 3.92
AA: 138667
RT: 5.39
AA: 29226
RT: 6.01
AA: 247003
RT: 7.96
AA: 165236
RT: 10.71
AA: 206379
RT: 11.55
AA: 339889
RT: 12.24
AA: 159034
Relative Abundance
Time (min)
[M-H2O+H]+
293.2118
[M+H]+
311.2226
293.2118
292.9159
291.1168
292.1204
294.2154
295.1911
296.1944
297.2068
298.1092
299.1859
300.2177
301.9869
303.1058
304.8958
305.1214
305.9130
306.9659
307.2252
308.2287
309.1694
310.9312
311.2226
311.4356
312.9439
313.2018
314.205
m/z
[M+Na]+
333.2047
333.1892
333.1163
333.1527
333.1692
333.2047
333.2435
m/z

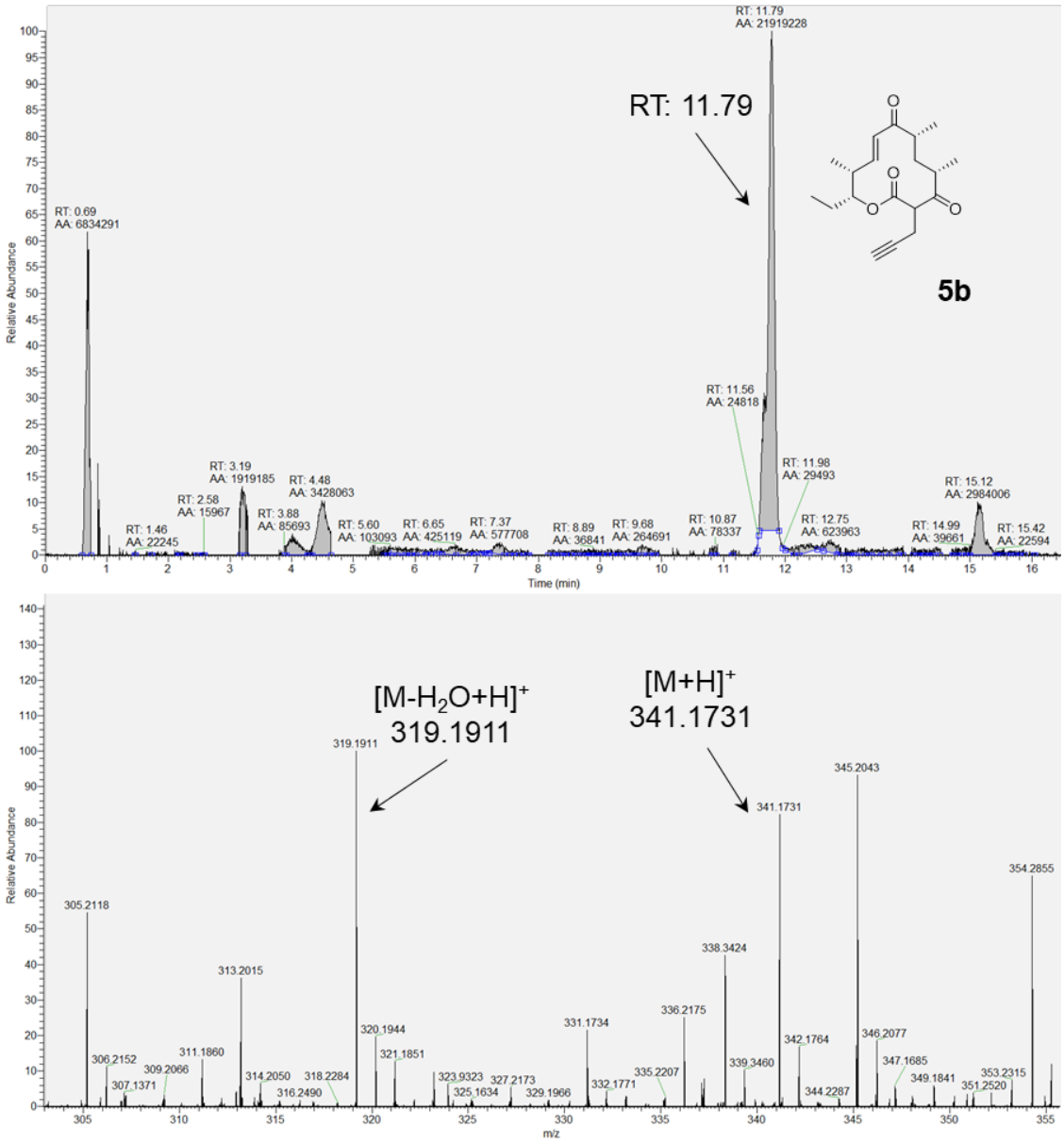
D


RT: 10.60
AA: 9616806
RT: 10.60
4d
RT: 9.63
AA: 3790754
RT: 10.82
AA: 59981
RT: 11.00
AA: 117752
RT: 9.06
AA: 790826
RT: 11.44
AA: 210090
RT: 13.09
AA: 381606
RT: 15.01
AA: 749496
RT: 15.27
AA: 171360
RT: 5.55
AA: 98285
RT: 6.49
AA: 195661
RT: 8.87
AA: 58734
Relative Abundance
Time (min)
[M+Na]+
347.2199
[M-H2O+H]+
307.2274
[M+H]+
32502379
347.2199
307.2274
315.2173
313.2015
311.1859
338.2026
349.1839
360.2751
308.2307
316.2206
321.1849
323.1860
331.1734
335.1843
329.1577
333.1681
339.3459
345.1528
343.2961
350.1874
353.1733
359.1814
356.2436
301.2856
305.9129
319.1893
325.2379
m/z

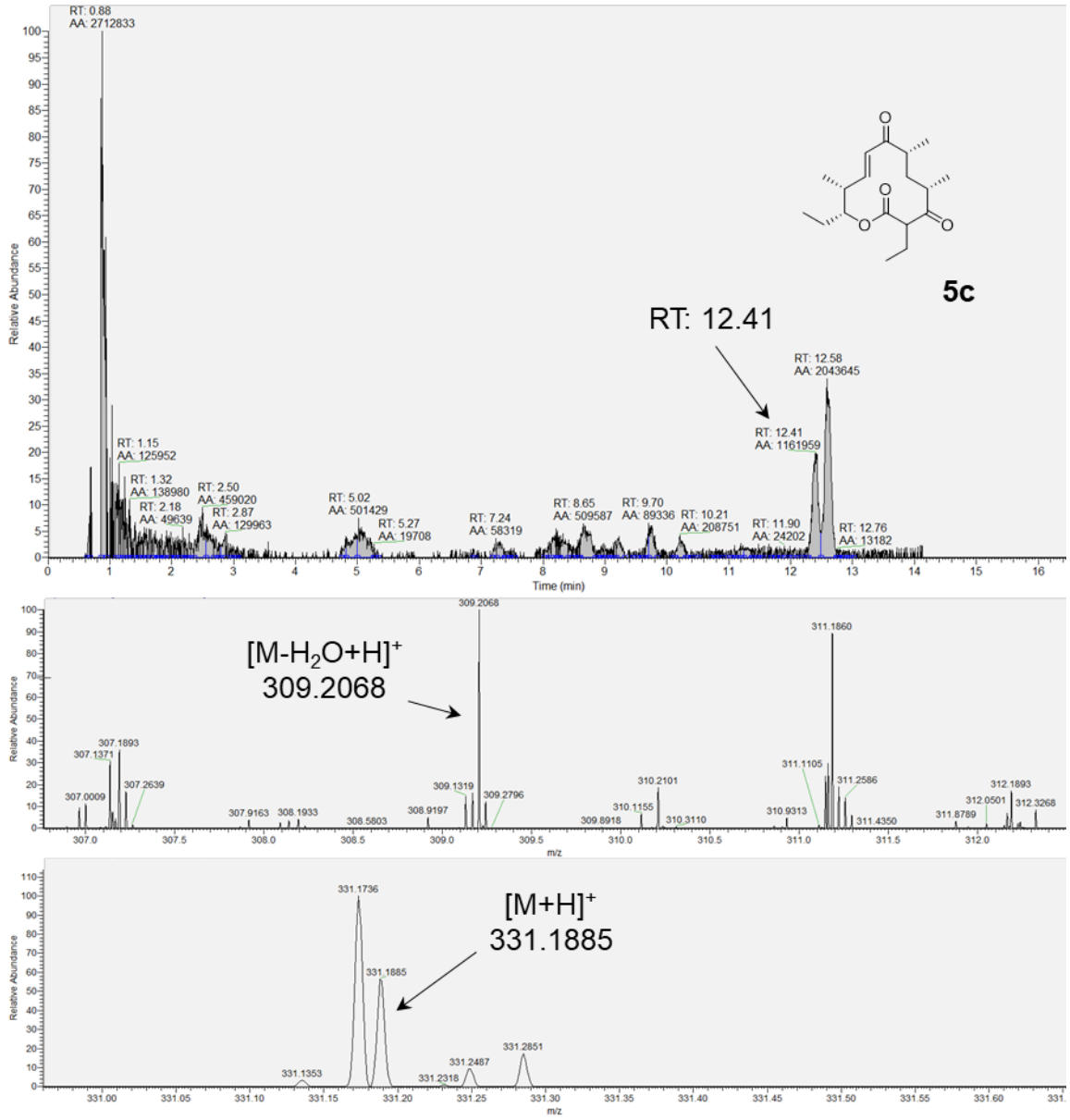
E


RT: 10.12
AA: 10530033
RT: 10.12
4i
RT: 10.36
AA: 392372
RT: 14.78
AA: 1338640
RT: 15.13
AA: 285126
RT: 11.73
AA: 518095
RT: 11.55
AA: 178701
RT: 15.37
AA: 721615
RT: 15.54
AA: 146977
RT: 7.17
AA: 756910
RT: 7.92
AA: 327469
RT: 9.30
AA: 157670
RT: 3.99
AA: 271434
RT: 6.27
AA: 143360
RT: 3.60
AA: 101148
RT: 4.55
AA: 58173
Relative Abundance
Time (min)
[M-H2O+H]+
305.2105
[M+H]+
323.2201
[M+Na]+
345.2035

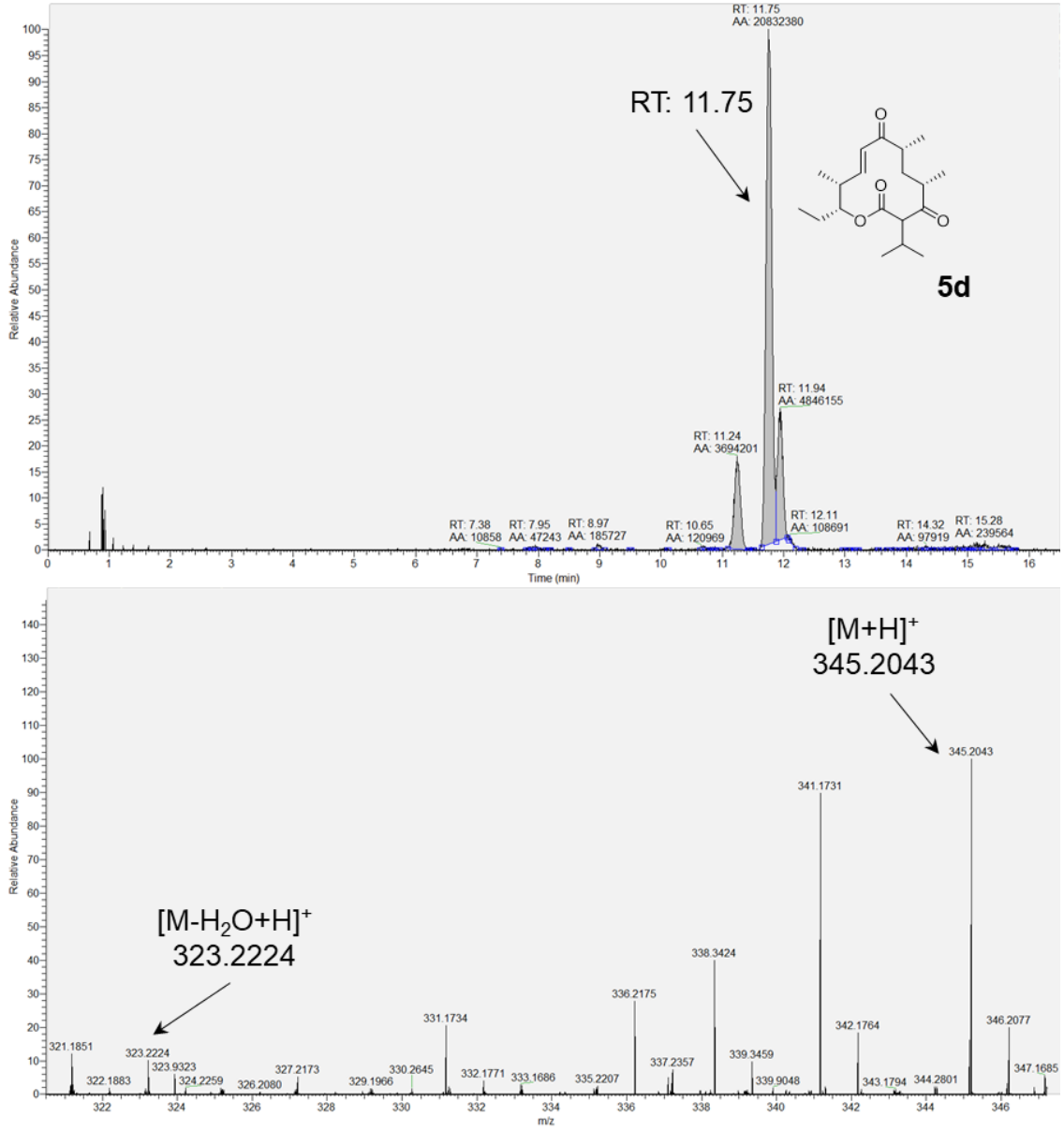
F


RT: 11.79
AA: 21919228
RT: 11.79
5b
RT: 0.69
AA: 6834291
RT: 11.56
AA: 24818
RT: 11.98
AA: 29493
RT: 3.19
AA: 1919185
RT: 4.48
AA: 3428063
RT: 2.58
AA: 15967
RT: 3.88
AA: 85693
RT: 15.12
AA: 2984006
RT: 1.46
AA: 22245
RT: 5.60
AA: 103093
RT: 6.65
AA: 425119
RT: 7.37
AA: 577708
RT: 8.89
AA: 36841
RT: 9.68
AA: 264691
RT: 10.87
AA: 78337
RT: 12.75
AA: 623963
RT: 14.99
AA: 39661
RT: 15.42
AA: 22594
Relative Abundance
Time (min)
[M-H2O+H]+
319.1911
[M+H]+
341.1731
319.1911
341.1731
345.2043
354.2855
305.2118
338.3424
313.2015
336.2175
320.1944
331.1734
342.1764
346.2077
311.1860
321.1851
306.2152
339.3460
309.2066
347.1685
314.2050
318.2284
323.9323
327.2173
335.2207
349.1841
353.2315
307.1371
316.2490
325.1634
329.1966
332.1771
344.2287
351.2520
Relative Abundance
m/z

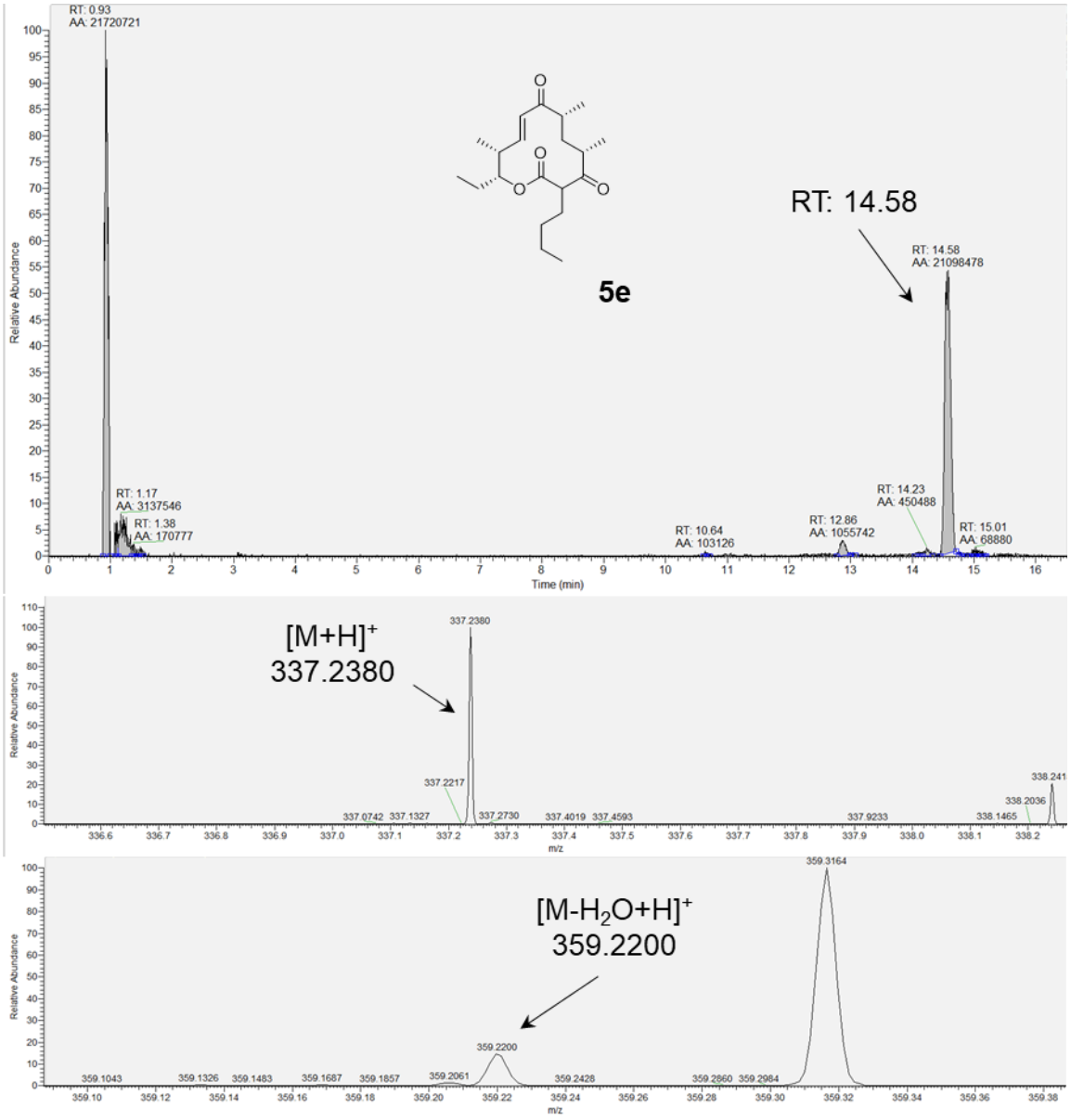
G


RT: 0.88
AA: 2712833
5c
RT: 12.41
RT: 12.58
AA: 2043645
RT: 12.41
AA: 1161959
RT: 1.15
AA: 125952
RT: 1.32
AA: 138980
RT: 2.18
AA: 49639
RT: 2.50
AA: 459020
RT: 2.87
AA: 129963
RT: 5.02
AA: 501429
RT: 5.27
AA: 19708
RT: 7.24
AA: 58319
RT: 8.65
AA: 509587
RT: 9.70
AA: 89336
RT: 10.21
AA: 208751
RT: 11.90
AA: 24202
RT: 12.76
AA: 13182
Relative Abundance
Time (min)
[M-H2O+H]+
309.2068
309.2068
311.1860
307.1893
307.1371
307.2639
307.0009
309.1319
309.2796
310.2101
310.1155
311.1105
311.2586
312.1893
312.0501
312.3268
[M+H]+
331.1885
331.1736
331.1885
331.1353
331.2318
331.2487
331.2851
m/z

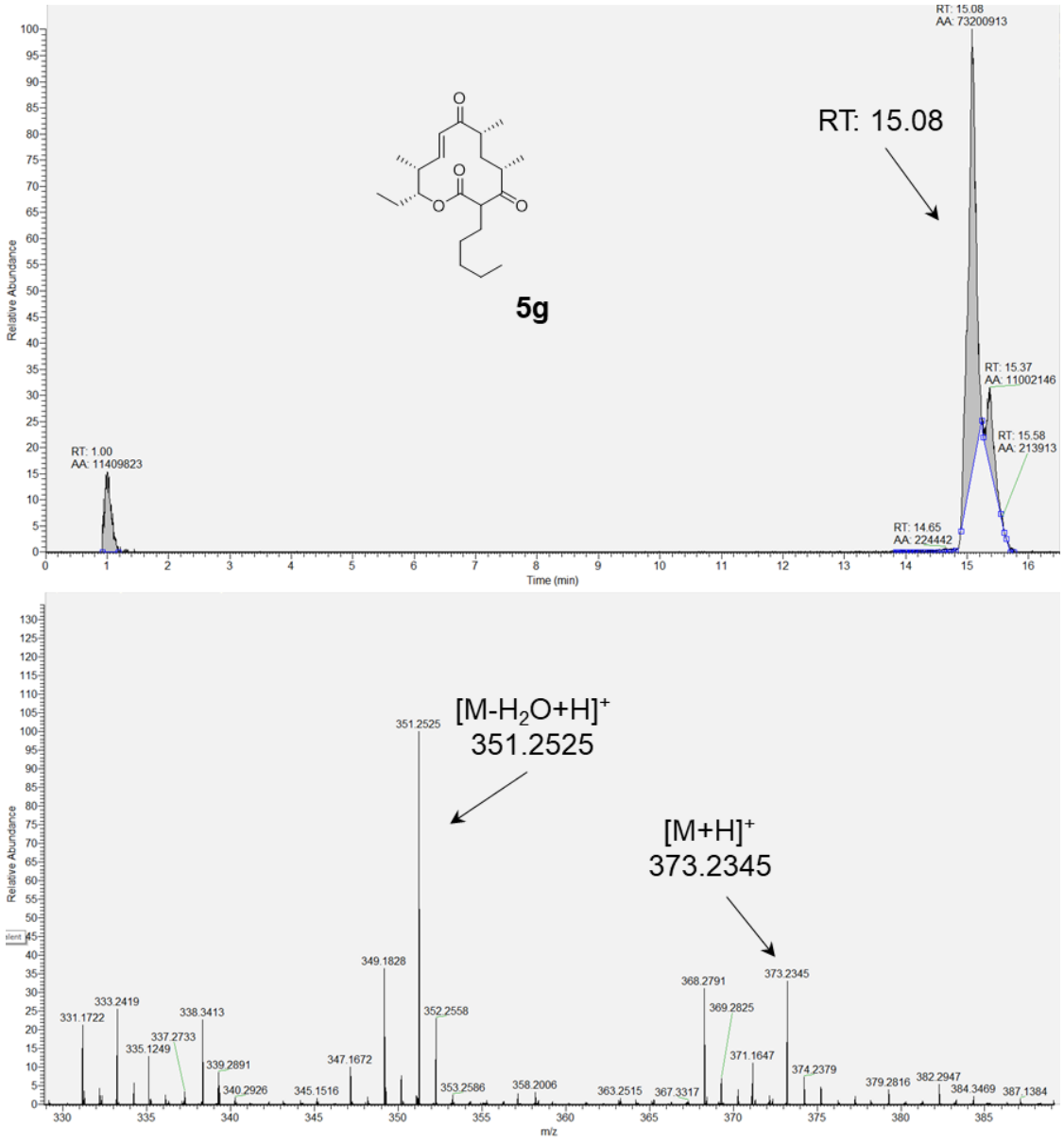
H


RT: 11.75
AA: 20832380
RT: 11.75
5d
RT: 11.94
AA: 4846155
RT: 11.24
AA: 3694201
RT: 12.11
AA: 108691
RT: 7.38
AA: 10858
RT: 7.95
AA: 47243
RT: 8.97
AA: 185727
RT: 10.65
AA: 120969
RT: 14.32
AA: 97919
RT: 15.28
AA: 239564
Relative Abundance
Time (min)
[M+H]+
345.2043
[M-H2O+H]+
323.2224
m/z

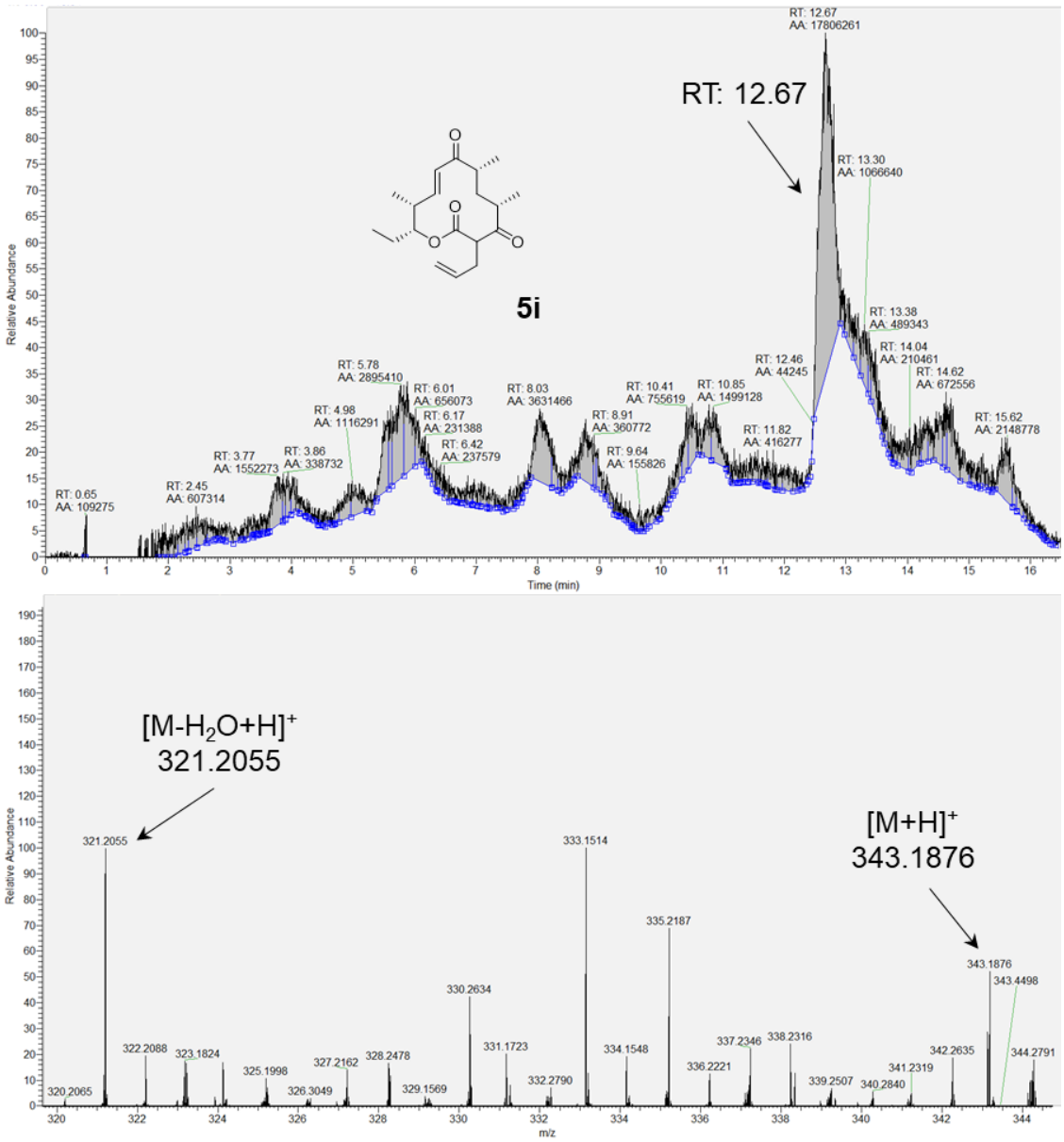
I


RT: 0.93
AA: 21720721
5e
RT: 14.58
RT: 14.58
AA: 21098478
RT: 1.17
AA: 3137546
RT: 1.38
AA: 170777
RT: 10.64
AA: 103126
RT: 12.86
AA: 1055742
RT: 14.23
AA: 450488
RT: 15.01
AA: 68880
Relative Abundance
Time (min)
[M+H]+
337.2380
337.2380
337.2217
337.2730
338.241
338.2036
m/z
[M-H2O+H]+
359.2200
359.3164
359.2200
359.2061
359.2428
359.2860
359.2984
m/z

J


RT: 15.08
AA: 73200913
RT: 15.08
5g
RT: 15.37
AA: 11002146
RT: 15.58
AA: 213913
RT: 1.00
AA: 11409823
RT: 14.65
AA: 224442
Relative Abundance
Time (min)
[M-H2O+H]+
351.2525
[M+H]+
373.2345
m/z

K



Appendix C

Table C1. Pik and DEBS PKS DNA FASTA Sequences. AT domains are in bold. Pik docking domain is italicized.

Construct	Nucleotide Sequences
PikAIII WT	ATGGCGAACAACGAAGACAAGCTCCGCGACTACCTCAAGCGCGTCACCGC CGAGCTGCAGCAGAACACCAGGCGTCTGCGCGAGATCGAGGGACGCACGC ACGAGCCGGTGGCGATCGTGGGCATGGCCTGCCGCCTGCCGGGCGGTGTC GCCTCGCCCGAGGACCTGTGGCAGCTGGTGGCCGGGGACGGGGACGCGAT CTCGGAGTTCCCGCAGGACCGCGGCTGGGACGTGGAGGGGCTGTACGACC CCGACCCGGACGCGTCCGGCAGGACGTACTGCCGGTCCGGCGGATTCCTG CACGACGCCGGCGAGTTCGACGCCGACTTCTTCGGGATCTCGCCGCGCGA

GGCCCTCGCCATGGACCCGCAGCAGCGACTGTCCCTCACCACCGCGTGGG
AGGCGATCGAGAGCGCGGGCATCGACCCGACGGCCCTGAAGGGCAGCGGC
CTCGGCGTCTTCGTCGGCGGCTGGCACACCGGCTACACCTCGGGGCAGAC
CACCGCCGTGCAGTCGCCCGAGCTGGAGGGCCACCTGGTCAGCGGCGCGG
CGCTGGGCTTCCTGTCCGGCCGTATCGCGTACGTCCTCGGTACGGACGGA
CCGGCCCTGACCGTGGACACGGCCTGCTCGTCCTCGCTGGTCGCCCTGCA
CCTCGCCGTGCAGGCCCTCCGCAAGGGCGAGTGCGACATGGCCCTCGCCG
GTGGTGTCACGGTCATGCCCAACGCGGACCTGTTCGTGCAGTTCAGCCGG
CAGCGCGGGCTGGCCGCGGACGGCCGGTCGAAGGCGTTCGCCACCTCGGC
GGACGGCTTCGGCCCCGCGGAGGGCGCCGGAGTCCTGCTGGTGGAGCGCC
TGTCGGACGCCCGCCGCAACGGACACCGGATCCTCGCGGTCGTCCGCGGC
AGCGCGGTCAACCAGGACGGCGCCAGCAACGGCCTCACGGCTCCGCACGG
GCCCTCCCAGCAGCGCGTCATCCGACGGGCCCTGGCGGACGCCCGGCTCG
CGCCGGGTGACGTGGACGTCGTCGAGGCGCACGGCACGGGCACGCGGCTC
GGCGACCCGATCGAGGCGCAGGCCCTCATCGCCACCTACGGCCAGGAGAA
GAGCAGCGAACAGCCGCTGAGGCTGGGCGCGTTGAAGTCGAACATCGGGC
ACACGCAGGCCGCGGCCGGTGTCGCAGGTGTCATCAAGATGGTCCAGGCG
ATGCGCCACGGACTGCTGCCGAAGACGCTGCACGTCGACGAGCCCTCGGA
CCAGATCGACTGGTCGGCGGGCACGGTGGAACTCCTCACCGAGGCCGTCG
ACTGGCCGGAGAAGCAGGACGGCGGGCTGCGCCGCGCGGCTGTCTCCTCC
TTCGGCATCAGCGGGACGAACGCGCACGTCGTCCTGGAGGAGGCCCCGGC
GGTCGAGGACTCCCCGGCCGTCGAGCCGCCGGCCGGTGGCGGTGTGGTGC
CGTGGCCGGTGTCCGCGAAGACTCCGGCCGCGCTGGACGCCCAGATCGGG
CAGCTCGCCGCGTACGCGGACGGTCGTACGGACGTGGATCCGGCGGTGGC
CGCCCGCGCCCTGGTCGACAGCCGTACGGCGATGGAGCACCGCGCGGTCG
CGGTCGGCGACAGCCGGGAGGCACTGCGGGACGCCCTGCGGATGCCGGAA
GGACTGGTACGCGGCACGTCCTCGGACGTGGGCCGGGTGGCGTTCGTCTT
CCCCGGCCAGGGCACGCAGTGGGCCGGCATGGGCGCCGAACTCCTTGACA
GCTCACCGGAGTTCGCTGCCTCGATGGCCGAATGCGAGACCGCGCTCTCC
CGCTACGTCGACTGGTCTCTTGAAGCCGTCGTCCGACAGGAACCCGGCGC
ACCCACGCTCGACCGCGTCGACGTCGTCCAGCCCGTGACCTTCGCTGTCA
TGGTCTCGCTGGCGAAGGTCTGGCAGCACCACGGCATCACCCCCCAGGCC
GTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTCGCCGGTGC
ACTCACCCTCGACGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCA
TCGCCGCCCACCTCGCCGGCAAGGGCGGCATGATCTCCCTCGCCCTCGAC
GAGGCGGCCGTCCTGAAGCGACTGAGCGACTTCGACGGACTCTCCGTCGC
CGCCGTCAACGGCCCCACCGCCACCGTCGTCTCCGGCGACCCGACCCAGA
TCGAGGAACTCGCCCGCACCTGCGAGGCCGACGGCGTCCGTGCGCGGATC
ATCCCGGTCGACTACGCCTCCCACAGCCGGCAGGTCGAGATCATCGAGAA
GGAGCTGGCCGAGGTCCTCGCCGGACTCGCCCCGCAGGCTCCGCACGTGC
CGTTCTTCTCCACCCTCGAAGGCACCTGGATCACCGAGCCGGTGCTCGAC
GGCACCTACTGGTACCGCAACCTGCGCCATCGCGTGGGCTTCGCCCCCGC
CGTGGAGACCTTGGCGGTTGACGGCTTCACCCACTTCATCGAGGTCAGCG
CCCACCCCGTCCTCACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGC
ACCCTCCGCCGCGAACAGGGAGGCCAGGAGCGTCTGGTCACCTCACTCGC
CGAAGCCTGGGCCAACGGCCTCACCATCGACTGGGCGCCCATCCTCCCCA
CCGCAACCGGCCACCACCCCGAGCTCCCCACCTACGCCTTCCAGACCGAG

	CGCTTCTGGCTGCAGAGCTCCGCGCCCACCAGCGCCGCCGACGACTGGCG TTACCGCGTCGAGTGGAAGCCGCTGACGGCCTCCGGCCAGGCGGACCTGT CCGGGCGGTGGATCGTCGCCGTCGGGAGCGAGCCAGAAGCCGAGCTGCTG GGCGCGCTGAAGGCCGCGGGAGCGGAGGTCGACGTACTGGAAGCCGGGGC GGACGACGACCGTGAGGCCCTCGCCGCCCGGCTCACCGCACTGACGACCG GCGACGGCTTCACCGGCGTGGTCTCGCTCCTCGACGACCTCGTGCCACAG GTCGCCTGGGTGCAGGCACTCGGCGACGCCGGAATCAAGGCGCCCCTGTG GTCCGTCACCCAGGGCGCGGTCTCCGTCGGACGTCTCGACACCCCCGCCG ACCCCGACCGGGCCATGCTCTGGGGCCTCGGCCGCGTCGTCGCCCTTGAG CACCCCGAACGCTGGGCCGGCCTCGTCGACCTCCCCGCCCAGCCCGATGC CGCCGCCCTCGCCCACCTCGTCACCGCACTCTCCGGCGCCACCGGCGAGG ACCAGATCGCCATCCGCACCACCGGACTCCACGCCCGCCGCCTCGCCCGC GCACCCCTCCACGGACGTCGGCCCACCCGCGACTGGCAGCCCCACGGCAC CGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCACGCCGCACGCT GGATGGCCCACCACGGAGCCGAACACCTCCTCCTCGTCAGCCGCAGCGGC GAACAAGCCCCCGGAGCCACCCAACTCACCGCCGAACTCACCGCATCGGG CGCCCGCGTCACCATCGCCGCCTGCGACGTCGCCGACCCCCACGCCATGC GCACCCTCCTCGACGCCATCCCCGCCGAGACGCCCCTCACCGCCGTCGTC CACACCGCCGGCGCACCGGGCGGCGATCCGCTGGACGTCACCGGCCCGGA GGACATCGCCCGCATCCTGGGCGCGAAGACGAGCGGCGCCGAGGTCCTCG ACGACCTGCTCCGCGGCACTCCGCTGGACGCCTTCGTCCTCTACTCCTCG AACGCCGGGGTCTGGGGCAGCGGCAGCCAGGGCGTCTACGCGGCGGCCAA CGCCCACCTCGACGCGCTCGCCGCCCGGCGCCGCGCCCGGGGCGAGACGG CGACCTCGGTCGCCTGGGGCCTCTGGGCCGGCGACGGCATGGGCCGGGGC GCCGACGACGCGTACTGGCAGCGTCGCGGCATCCGTCCGATGAGCCCCGA CCGCGCCCTGGACGAACTGGCCAAGGCCCTGAGCCACGACGAGACCTTCG TCGCCGTGGCCGATGTCGACTGGGAGCGGTTCGCGCCCGCGTTCACGGTG TCCCGTCCCAGCCTTCTGCTCGACGGCGTCCCGGAGGCCCGGCAGGCGCT CGCCGCACCCGTCGGTGCCCCGGCTCCCGGCGACGCCGCCGTGGCGCCGA CCGGGCAGTCGTCGGCGCTGGCCGCGATCACCGCGCTCCCCGAGCCCGAG CGCCGGCCGGCGCTCCTCACCCTCGTCCGTACCCACGCGGCGGCCGTACT CGGCCATTCCTCCCCCGACCGGGTGGCCCCCGGCCGTGCCTTCACCGAGC TCGGCTTCGACTCGCTGACGGCCGTGCAGCTCCGCAACCAGCTCTCCACG GTGGTCGGCAACAGGCTCCCCGCCACCACGGTCTTCGACCACCCGACGCC CGCCGCACTCGCCGCGCACCTCCACGAGGCGTACCTCGCACCGGCCGAGC CGGCCCCGACGGACTGGGAGGGGCGGGTGCGCCGGGCCCTGGCCGAACTG CCCCTCGACCGGCTGCGGGACGCGGGGGTCCTCGACACCGTCCTGCGCCT CACCGGCATCGAGCCCGAGCCGGGTTCCGGCGGTTCGGACGGCGGCGCCG CCGACCCTGGTGCGGAGCCGGAGGCGTCGATCGACGACCTGGACGCCGAG GCCCTGATCCGGATGGCTCTCGGCCCCCGTAACACCTGA
PikAIV WT	ATGACGAGTTCCAACGAACAGTTGGTGGACGCTCTGCGCGCCTCTCTCAA GGAGAACGAAGAACTCCGGAAAGAGAGCCGTCGCCGGGCCGACCGTCGGC AGGAGCCCATGGCGATCGTCGGCATGAGCTGCCGGTTCGCGGGCGGAATC CGGTCCCCCGAGGACCTCTGGGACGCCGTCGCCGCGGGCAAGGACCTGGT CTCCGAGGTACCGGAGGAGCGCGGCTGGGACATCGACTCCCTCTACGACC CGGTGCCCGGGCGCAAGGGCACGACGTACGTCCGCAACGCCGCGTTCCTC GACGACGCCGCCGGATTCGACGCGGCCTTCTTCGGGATCTCGCCGCGCGA

GGCCCTCGCCATGGACCCGCAGCAGCGGCAGCTCCTCGAAGCCTCCTGGG
AGGTCTTCGAGCGGGCCGGCATCGACCCCGCGTCGGTCCGCGGCACCGAC
GTCGGCGTGTACGTGGGCTGTGGCTACCAGGACTACGCGCCGGACATCCG
GGTCGCCCCCGAAGGCACCGGCGGTTACGTCGTCACCGGCAACTCCTCCG
CCGTGGCCTCCGGGCGCATCGCGTACTCCCTCGGCCTGGAGGGACCCGCC
GTGACCGTGGACACGGCGTGCTCCTCTTCGCTCGTCGCCCTGCACCTCGC
CCTGAAGGGCCTGCGGAACGGCGACTGCTCGACGGCACTCGTGGGCGGCG
TGGCCGTCCTCGCGACGCCGGGCGCGTTCATCGAGTTCAGCAGCCAGCAG
GCCATGGCCGCCGACGGCCGGACCAAGGGCTTCGCCTCGGCGGCGGACGG
CCTCGCCTGGGGCGAGGGCGTCGCCGTACTCCTCCTCGAACGGCTCTCCG
ACGCGCGGCGCAAGGGCCACCGGGTCCTGGCCGTCGTGCGCGGCAGCGCC
ATCAACCAGGACGGCGCGAGCAACGGCCTCACGGCTCCGCACGGGCCCTC
CCAGCAGCGCCTGATCCGCCAGGCCCTGGCCGACGCGCGGCTCACGTCGA
GCGACGTGGACGTCGTGGAGGGCCACGGCACGGGGACCCGTCTCGGCGAC
CCGATCGAGGCGCAGGCGCTGCTCGCCACGTACGGGCAGGGGCGCGCCCC
GGGGCAGCCGCTGCGGCTGGGGACGCTGAAGTCGAACATCGGGCACACGC
AGGCCGCTTCGGGTGTCGCCGGTGTCATCAAGATGGTGCAGGCGCTGCGC
CACGGGGTGCTGCCGAAGACCCTGCACGTGGACGAGCCGACGGACCAGGT
CGACTGGTCGGCCGGTTCGGTCGAGCTGCTCACCGAGGCCGTGGACTGGC
CGGAGCGGCCGGGCCGGCTCCGCCGGGCGGGCGTCTCCGCGTTCGGCGTG
GGC**GGGACGAACGCGCACGTCGTCCTGGAGGAGGCCCCGGCGGTCGAGGA**
GTCCCCTGCCGTCGAGCCGCCGGCCGGTGGCGGCGTGGTGCCGTGGCCGG
TGTCCGCGAAGACCTCGGCCGCACTGGACGCCCAGATCGGGCAGCTCGCC
GCATACGCGGAAGACCGCACGGACGTGGATCCGGCGGTGGCCGCCCGCGC
CCTGGTCGACAGCCGTACGGCGATGGAGCACCGCGCGGTCGCGGTCGGCG
ACAGCCGGGAGGCACTGCGGGACGCCCTGCGGATGCCGGAAGGACTGGTA
CGGGGCACGGTCACCGATCCGGGCCGGGTGGCGTTCGTCTTCCCCGGCCA
GGGCACGCAGTGGGCCGGCATGGGCGCCGAACTCCTCGACAGCTCACCCG
AATTCGCCGCCGCCATGGCCGAATGCGAGACCGCACTCTCCCCGTACGTC
GACTGGTCTCTCGAAGCCGTCGTCCGACAGGCTCCCAGCGCACCGACACT
CGACCGCGTCGACGTCGTCCAGCCCGTCACCTTCGCCGTCATGGTCTCCC
TCGCCAAGGTCTGGCAGCACCACGGCATCACCCCCGAGGCCGTCATCGGC
CACTCCCAGGGCGAGATCGCCGCCGCGTACGTCGCCGGTGCCCTCACCCT
CGACGACGCCGCTCGTGTCGTGACCCTCCGCAGCAAGTCCATCGCCGCCC
ACCTCGCCGGCAAGGGCGGCATGATCTCCCTCGCCCTCAGCGAGGAAGCC
ACCCGGCAGCGCATCGAGAACCTCCACGGACTGTCGATCGCCGCCGTCAA
CGGGCCTACCGCCACCGTGGTTTCGGGCGACCCCACCCAGATCCAAGAAC
TTGCTCAGGCGTGTGAGGCCGACGGCATCCGCGCACGGATCATCCCCGTC
GACTACGCCTCCCACAGCGCCCACGTCGAGACCATCGAGAACGAACTCGC
CGACGTCCTGGCGGGGTTGTCCCCCCAGACACCCCAGGTCCCCTTCTTCT
CCACCCTCGAAGGCACCTGGATCACCGAACCCGCCCTCGACGGCGGCTAC
TGGTACCGCAACCTCCGCCATCGTGTGGGCTTCGCCCCGGCCGTCGAGAC
CCTCGCCACCGACGAAGGCTTCACCCACTTCATCGAGGTCAGCGCCCACC
CCGTCCTCACCATGACCCTCCCCGACAAGGTCACCGGCCTGGCCACCCTC
CGACGCGAGGACGGCGGACAGCACCGCCTCACCACCTCCCTTGCCGAGGC
CTGGGCCAACGGCCTCGCCCTCGACTGGGCCTCCCTCCTGCCCGCCACGG
GCGCCCTCAGCCCCGCCGTCCCCGACCTCCCGACGTACGCCTTCCAGCAC

	CGCTCGTACTGGATCAGCCCCGCGGGTCCCGGCGAGGCGCCCGCGCACAC CGCTTCCGGGCGCGAGGCCGTCGCCGAGACGGGGCTCGCGTGGGGCCCGG GTGCCGAGGACCTCGACGAGGAGGGCCGGCGCAGCGCCGTACTCGCGATG GTGATGCGGCAGGCGGCCTCCGTGCTCCGGTGCGACTCGCCCGAAGAGGT CCCCGTCGACCGCCCGCTGCGGGAGATCGGCTTCGACTCGCTGACCGCCG TCGACTTCCGCAACCGCGTCAACCGGCTGACCGGTCTCCAGCTGCCGCCC ACCGTCGTGTTCGAGCACCCGACGCCCGTCGCGCTCGCCGAGCGCATCAG CGACGAGCTGGCCGAGCGGAACTGGGCCGTCGCCGAGCCGTCGGATCACG AGCAGGCGGAGGAGGAGAAGGCCGCCGCTCCGGCGGGGGCCCGCTCCGGG GCCGACACCGGCGCCGGCGCCGGGATGTTCCGCGCCCTGTTCCGGCAGGC CGTGGAGGACGACCGGTACGGCGAGTTCCTCGACGTCCTCGCCGAAGCCT CCGCGTTCCGCCCGCAGTTCGCCTCGCCCGAGGCCTGCTCGGAGCGGCTC GACCCGGTGCTGCTCGCCGGCGGTCCGACGGACCGGGCGGAAGGCCGTGC CGTTCTCGTCGGCTGCACCGGCACCGCGGCGAACGGCGGCCCGCACGAGT TCCTGCGGCTCAGCACCTCCTTCCAGGAGGAGCGGGACTTCCTCGCCGTA CCTCTCCCCGGCTACGGCACGGGTACGGGCACCGGCACGGCCCTCCTCCC GGCCGATCTCGACACCGCGCTCGACGCCCAGGCCCGGGCGATCCTCCGGG CCGCCGGGGACGCCCCGGTCGTCCTGCTCGGGCACTCCGGCGGCGCCCTG CTCGCGCACGAGCTGGCCTTCCGCCTGGAGCGGGCGCACGGCGCGCCGCC GGCCGGGATCGTCCTGGTCGACCCCTATCCGCCGGGCCATCAGGAGCCCA TCGAGGTGTGGAGCAGGCAGCTGGGCGAGGGCCTGTTCGCGGGCGAGCTG GAGCCGATGTCCGATGCGCGGCTGCTGGCCATGGGCCGGTACGCGCGGTT CCTCGCCGGCCCGCGGCCGGGCCGCAGCAGCGCGCCCGTGCTTCTGGTCC GTGCCTCCGAACCGCTGGGCGACTGGCAGGAGGAGCGGGGCGACTGGCGT GCCCACTGGGACCTTCCGCACACCGTCGCGGACGTGCCGGGCGACCACTT CACGATGATGCGGGACCACGCGCCGGCCGTCGCCGAGGCCGTCCTCTCCT GGCTCGACGCCATCGAGGGCATCGAGGGGGCGGGCAAG
PDDEry6TE **WT**	*ATGACGAGTTCCAACGAACAGTTGGTGGACGCTCTGCGCGCCTCTCTCAA* *GGAGAACGAAGAACTCCGGAAAGAGAGCCGTCGCCGGGCCGACCGTCGGC* *AGGAGGAG*ATCGCGATCGTCGGCATGGCCTGCCGCTTCCCCGGCGGCGTG CACAACCCCGGTGAGCTGTGGGAGTTCATCGTCGGCGGCGGAGACGCCGT GACGGAGATGCCCACCGACCGCGGCTGGGACCTCGACGCGCTGTTCGACC CCGACCCGCAGCGCCACGGAACCAGCTACTCGCGACACGGCGCGTTCCTC GACGGGGCCGCCGACTTCGACGCGGCGTTCTTCGGGATCTCGCCGCGCGA GGCGCTGGCGATGGACCCGCAGCAGCGCCAGGTCCTGGAAACGACGTGGG AGCTGTTCGAGAACGCCGGCATCGACCCGCACTCGCTGCGGGGCAGCGAC ACCGGCGTCTTCCTCGGCGCCGCGTACCAGGGCTACGGCCAGGACGCGGT GGTGCCCGAGGACAGCGAGGGCTACCTGCTCACCGGCAACTCCTCCGCCG TGGTGTCCGGCCGGGTCGCCTACGTGCTGGGGCTGGAAGGCCCCGCGGTC ACGGTGGACACGGCGTGTTCGTCGTCGTTGGTGGCCTTGCATTCGGCGTG TGGGTCGTTGCGTGACGGTGACTGCGGTCTTGCGGTGGCCGGTGGTGTGT CGGTGATGGCGGGCCCGGAGGTGTTCACCGAGTTCTCCCGCCAGGGCGGC TTGGCCGTGGACGGGCGCTGCAAGGCGTTCTCCGCGGAGGCCGACGGCTT CGGTTTCGCCGAGGGCGTCGCGGTGGTCCTGCTCCAGCGGTTGTCCGACG CCCGCAGGGCGGGTCGCCAGGTGCTCGGCGTGGTCGCGGGCTCGGCGATC AACCAGGACGGCGCGAGCAACGGTCTCGCGGCGCCGAGCGGCGTCGCCCA GCAGCGCGTGATCCGCAAGGCGTGGGCGCGTGCGGGGATCACGGGCGCGG

ATGTGGCCGTGGTGGAGGCGCATGGGACCGGTACGCGGCTGGGCGATCCG
GTGGAGGCGTCGGCGTTGCTGGCTACTTACGGCAAGTCGCGCGGGTCGTC
GGGCCCGGTGCTGCTGGGTTCGGTGAAGTCGAACATCGGTCACGCGCAGG
CGGCCGCGGGTGTCGCGGGCGTGATCAAGGTGGTCCTGGGGTTGAACCGC
GGCCTGGTGCCGCCGATGCTCTGCCGCGGCGAGCGGTCGCCGCTGATCGA
ATGGTCCTCGGGTGGTGTGGAACTTGCCGAGGCCGTGAGCCCGTGGCCTC
CGGCCGCGGACGGGGTGCGCCGGGCCGGTGTGTCGGCGTTCGGGGTGAGC
GGGACGAACGCGCACGTGATCATCGCCGAGCCCCCGGAGCCCGAGCCGCT
GCCGGAACCCGGACCGGTGGGCGTGCTGGCCGCTGCGAACTCGGTGCCCG
TACTGCTGTCGGCCAGGACCGAGACCGCGTTGGCAGCGCAGGCGCGGCTC
CTGGAGTCCGCAGTGGACGACTCGGTTCCGTTGACGGCATTGGCTTCCGC
GCTGGCCACCGGACGCGCCCACCTGCCGCGTCGTGCGGCGTTGCTGGCAG
GCGACCACGAACAGCTCCGCGGGCAGTTGCGAGCGGTCGCCGAGGGCGTT
GCGGCTCCCGGTGCCACCACCGGAACCGCCTCCGCCGGCGGCGTGGTTTT
CGTCTTCCCAGGTCAGGGTGCTCAGTGGGAGGGCATGGCCCGGGGCTTGC
TCTCGGTCCCCGTCTTCGCCGAGTCGATCGCCGAGTGCGATGCGGTGTTG
TCGGAGGTGGCCGGGTTCTCGGCCTCCGAAGTGCTGGAGCAGCGTCCGGA
CGCGCCGTCGCTGGAGCGGGTCGACGTCGTACAGCCGGTGTTGTTCTCCG
TGATGGTGTCGCTGGCGCGGCTGTGGGGCGCTTGCGGAGTCAGCCCCTCG
GCCGTCATCGGCCATTCGCAGGGCGAGATCGCCGCCGCGGTGGTGGCCGG
GGTGTTGTCGCTGGAGGACGGCGTGCGCGTCGTGGCCCTGCGCGCGAAGG
CGTTGCGTGCGCTGGCGGGCAAGGGCGGCATGGTCTCGTTGGCGGCTCCC
GGTGAACGCGCCCGCGCGCTGATCGCACCGTGGGAGGACCGGATCTCCGT
CGCGGCGGTCAACTCCCCGTCCTCGGTCGTGGTCTCCGGCGATCCGGAGG
CGCTGGCCGAACTCGTCGCACGTTGCGAGGACGAGGGCGTGCGCGCCAAG
ACGCTCCCGGTGGACTACGCCTCGCACTCCCGCCACGTCGAGGAGATCCG
CGAGACGATCCTCGCCGACCTCGACGGCATCTCCGCGCGGCGTGCCGCCA
TCCCGCTCTACTCCACGCTGCACGGCGAACGGCGCGACGGCGCCGACATG
GGTCCGCGGTACTGGTACGACAACCTGCGCTCCCAGGTGCGCTTCGACGA
GGCGGTCTCGGCCGCCGTCGCCGACGGTCACGCCACCTTCGTCGAGATGA
GCCCGCACCCGGTGCTCACCGCGGCGGTGCAGGAGATCGCCGCGGACGCC
GTGGCCATCGGGTCGCTGCACCGCGACACCGCGGAGGAGCACCTGATCGC
CGAGCTCGCCCGGGCGCACGTGCACGGCGTGGCCGTGGACTGGCGGAACG
TCTTCCCGGCGGCACCTCCGGTGGCGCTGCCCAACTACCCGTTCGAGCCC
CAGCGGTACTGGCTCGCGCCGGAGGTGTCCGACCAGCTCGCCGACAGCCG
CTACCGCGTCGACTGGCGACCGCTGGCCACCACGCCGGTGGACCTGGAAG
GCGGCTTCCTGGTCCACGGGTCCGCACCGGAGTCGCTGACCAGCGCAGTC
GAGAAGGCCGGAGGCCGCGTCGTGCCGGTCGCCTCGGCCGACCGCGAAGC
GCTCGCGGCGGCCCTGCGGGAGGTGCCGGGCGAGGTCGCCGGCGTGCTCT
CGGTCCACACCGGCGCCGCAACGCACCTCGCCCTGCACCAGTCGCTGGGT
GAGGCCGGCGTGCGGGCCCCGCTCTGGCTGGTCACCAGCCGAGCGGTCGC
GCTCGGGGAGTCCGAGCCGGTCGATCCCGAGCAGGCGATGGTGTGGGGTC
TCGGGCGCGTCATGGGCCTGGAGACCCCGGAACGGTGGGGCGGTCTGGTG
GACCTGCCCGCCGAACCCGCGCCGGGGGACGGCGAGGCGTTCGTCGCCTG
CCTCGGCGCGGACGGCCACGAGGACCAGGTCGCGATCCGTGACCACGCCC
GCTACGGCCGCCGCCTCGTCCGCGCCCCGCTGGGCACCCGCGAGTCGAGC
TGGGAGCCGGCGGGCACGGCGCTGGTCACCGGCGGCACCGGTGCGCTCGG

	CGGCCACGTCGCCCGCCACCTCGCCAGGTGCGGGGTGGAGGACCTGGTGC TGGTCAGCAGGCGCGGCGTCGACGCTCCCGGCGCGGCCGAGCTGGAAGCC GAACTGGTCGCCCTCGGCGCGAAGACGACCATCACCGCCTGCGACGTGGC CGACCGCGAGCAGCTCTCCAAGCTGCTGGAAGAACTGCGCGGGCAGGGAC GTCCGGTGCGGACCGTCGTGCACACCGCCGGGGTGCCCGAATCGAGGCCG CTGCACGAGATCGGCGAGCTGGAGTCGGTCTGCGCGGCGAAGGTGACCGG GGCCCGGCTGCTCGACGAGCTGTGCCCGGACGCCGAGACCTTCGTCCTGT TCTCGTCCGGAGCGGGGGTGTGGGGCAGTGCGAACCTCGGCGCCTACTCC GCGGCCAACGCCTACCTCGACGCGCTGGCCCACCGCCGCCGTGCGGAAGG CCGTGCGGCGACGTCCGTCGCGTGGGGCGCCTGGGCGGGCGAGGGCATGG CCACCGGCGACCTCGAGGGGCTCACCCGGCGCGGCCTGCGCCCGATGGCG CCCGAGCGCGCGATCCGCGCGCTGCACCAGGCGCTGGACAACGGCGACAC GTGCGTTTCGATCGCCGACGTCGACTGGGAGCGCTTCGCGGTCGGCTTCA CCGCCGCCCGGCCGCGTCCGCTGCTGGACGAGCTCGTCACGCCGGCGGTG GGGGCCGTCCCCGCGGTGCAGGCGGCCCCGGCGCGGGAGATGACGTCGCA GGAGTTGCTGGAGTTCACGCACTCGCACGTCGCGGCGATCCTCGGGCATT CCAGCCCGGACGCGGTCGGGCAGGACCAGCCGTTCACCGAGCTCGGCTTC GACTCGCTGACCGCGGTCGGGCTGCGCAACCAGCTCCAGCAGGCCACCGG GCTCGCGCTGCCCGCGACCCTGGTGTTCGAGCACCCCACGGTCCGCAGGT TGGCCGACCACATAGGACAGCAGCTCGACAGCGGGACTCCCGCCCGGGAA GCGAGCAGCGCTCTTCGCGACGGCTACCGGCAGGCGGGCGTGTCGGGCAG GGTCCGGTCCTACCTCGACCTGCTGGCGGGGCTGTCGGACTTCCGCGAGC ACTTCGACGGCTCCGACGGGTTCTCCCTCGATCTCGTGGACATGGCCGAC GGTCCCGGAGAGGTCACGGTGATCTGCTGCGCGGGAACGGCGGCGATCTC CGGTCCGCACGAGTTCACCCGGCTCGCCGGGGCGCTGCGCGGAATCGCTC CGGTTCGGGCCGTGCCCCAGCCCGGCTACGAGGAGGGCGAACCTCTGCCG TCGTCGATGGCGGCGGTGGCGGCGGTGCAGGCCGATGCGGTCATCAGGAC ACAGGGGGACAAGCCGTTCGTGGTGGCCGGTCACTCCGCGGGGGCACTGA TGGCCTACGCGCTGGCGACCGAACTGCTCGATCGCGGGCACCCGCCACGC GGTGTCGTCCTGATCGACGTCTACCCGCCCGGTCACCAGGACGCGATGAA CGCCTGGCTGGAGGAGCTGACCGCCACGCTGTTCGACCGCGAGACGGTGC GGATGGACGACACCAGGCTCACCGCCCTGGGCGCCTACGACCGCCTCACC GGTCAGTGGCGACCCCGGGAAACCGGGCTGCCGACGCTGCTGGTCAGCGC CGGCGAGCCGATGGGTCCGTGGCCCGACGACAGCTGGAAGCCGACGTGGC CCTTCGAGCACGACACCGTCGCCGTCCCCGGCGACCACTTCACGATGGTG CAGGAACACGCCGACGCGATCGCGCGGCACATCGACGCCTGGCTGGGCGG AGGGAATTCAAGA

Table C2. Primers for Pik and DEBS PKS Construction and Mutagenesis.

Primer Name	5'→3' Primer Sequence	Primer Function
PikAT5_PikAIV.Gib1	CGTGGGCGGGACGAACGCG	

PikAT5_PikAIV.Gib2	GAGCGGTGCTGGAAGGCGTAGGTGGGG	Amplifies PikAT5 for GA with PikAIV
PikAIV_PikAT5.Gib1	CTTCCAGCACCGCTCGTAC	Amplifies PikAIV for
PikAIV_PikAT5.Gib2	TTCGTCCCGCCCACGCCGAA	GA with PikAT5
PikAT6_PikAIII.Gib1	CATCAGCGGGACGAACGCG	Amplifies PikAT6 for
PikAT6_PikAIII.Gib2	CGCTCGGTCTGGAAGGCGTACGTCGGGAG	GA with PikAIII
PikAIII_PikAT6.Gib1	CTTCCAGACCGAGCGCTTC	Amplifies PikAIII for
PikAIII_PikAT6.Gib2	TTCGTCCCGCTGATGCCGAA	GA with PikAT6
PikAIII_V753X.FOR	ATCATCCCGDNTGACTACGCCTCCCACAGCCGGCAGGTCGAGATCA	Partial saturation of PikAIII Val753
PikAIII_V753X.REV	GGCGTAGTCANHCGGGATGATCCGCGCACGGACGCCGTCGGCCTCG	
PikAIII_Y755V.FOR	GTCGCCTCCCACAGCCGG	Introduces Y755V
PikAIII_Y755V.REV	GTCGACCGGGATGATCCGC	mutation in PikAIII
PikAIV_Y753X.FOR	NNKGCCTCCCACAGCGCCC	Full saturation of
PikAIV_Y753X.REV	GTCGACGGGGATGATCCGTGC	PikAIV Tyr753
Ery6-AT.Gib3	CTGCCCAACTACCCGTTCG	Amplifies Ery6TE for
Ery6-AT.Gib2	GCTCACCCCGAACGCC	GA with Pik ATs
PikAT6_Ery6.Gib1	GCGTTCGGGGTGAGCGGGACGAACGCG	Amplifies PikAT6 for GA with Ery6TE
PikAT6_Ery6.Gib3	GGGTAGTTGGGCAGGTCGGGGACGGCGG	

PikAT5_Ery6.Gib1	GCGTTCGGGGTGAGCGGGACGAACGC	Amplifies PiKAT5 for GA with Ery6TE
PikAT5_Ery6.Gib3	GGGTAGTTGGGCAGCTCGGGGTGGTGGCC	

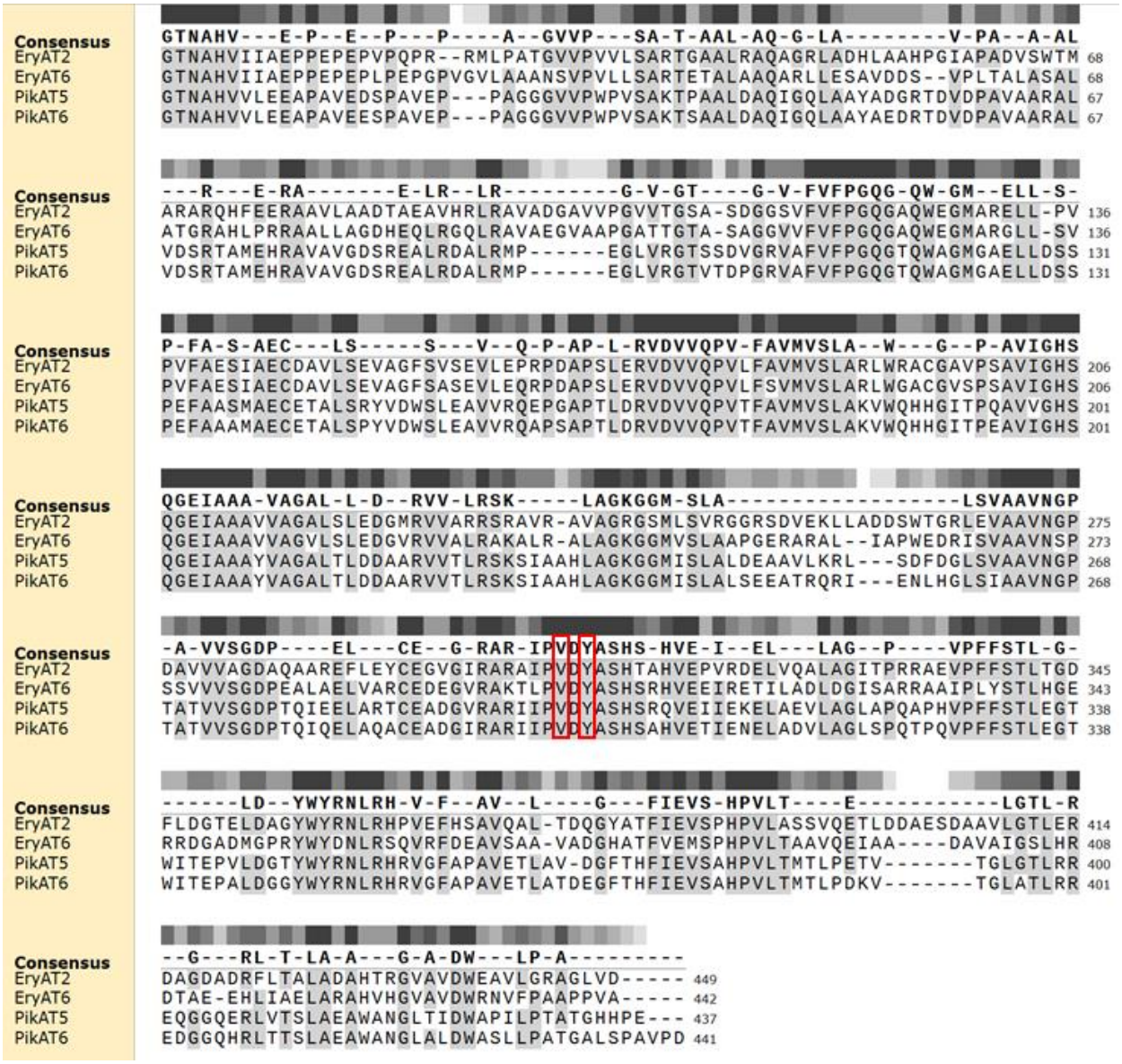


Figure C1. Amino Acid Alignment of EryAT2, EryAT6, PikAT5, and PikAT6. Residues targeted for mutagenesis are boxed in red.

Table C3. High-Resolution LC-MS Parameters and Gradient.

HESI Source Parameters	
Spray Voltage	3.5 kV
Capillary Temperature	350 °C
Heater Temperature	300 °C
S Lens RF Level	70 V
Sheath Gas Flow Rate	60 au
Resolution	70,000 FWHM
Scan Range	100-1000 m/z

LC Gradient

Time (min)	% B
0.0	25.0
1.0	25.0
10.0	59.0
11.0	85.0
12.0	85.0
12.5	25.0
16.5	25.0

Table C4. High-Resolution LC-MS Retention Times, Calculated Masses, and Observed Masses for DEBS/Pik PKS-Catalyzed Reaction Products. N.D. = Not Detected. Products with more than one peak are likely due to different ring conformations of narbonolide products.

Compound & Retention Time (min)		Calculated Mass ($[M-H_2O+H]^+$)	Observed Mass ($[M-H_2O+H]^+$)	Calculated Mass ($[M+H]^+$)	Observed Mass ($[M+H]^+$)	Calculated Mass ($[M+Na]^+$)	Observed Mass ($[M+Na]^+$)
1	7.68	279.1955	279.1949	297.2060	297.2055	319.1880	319.1873
2	9.21/9.92	335.2217	335.2210	353.2323	353.2316	375.2142	375.2129
3a	8.65	303.1955	303.1949	321.2060	321.2053	343.1880	343.1877
3b	9.41	293.2111	293.2106	311.2217	311.2211	333.2036	333.2037
3c	8.13	305.2111	305.2104	323.2217	323.2210	345.2036	345.2058
3d	N.D.	321.2424	N.D.	339.2530	N.D.	361.2349	N.D.
4a	11.79	--	--	319.1904	319.1889	341.1723	341.1718
4b	12.42	--	--	309.2060	309.2054	331.1880	331.1872
4c	12.87	--	--	321.2060	321.2054	343.1880	343.1873
4d	14.55	--	--	337.2373	337.2363	359.2193	359.2182
5a **6a**	10.31/10.82	359.2217	359.2208	377.2323	377.2314	399.2142	399.2133
5b	10.40/10.71	349.2373	349.2364	367.2479	367.2471	389.2298	389.2288

6b	11.46/11.80						
5c	11.34/12.01	361.2373	361.2364	379.2479	379.2469	401.2298	401.2288
6c							
5d	13.19/13.98	377.2686	377.2677	395.2792	395.2777	417.2611	417.2618
6d	14.42/15.48						
7a	11.15/11.45	383.2217	383.2209	401.2323	401.2315	423.2142	423.2133
7b	11.42/12.49	363.2530	363.2519	381.2636	381.2626	403.2455	403.2447
7c	12.92/14.00	387.2530	387.2520	405.2636	405.2618	427.2455	427.2449
7d	N.D.	419.3156	N.D.	437.3262	N.D.	459.3081	N.D.

Table C5. High-Resolution LC-MS Peak Areas for PKS-Catalyzed Reaction Products with Non-Natural Extenders. Extracted ion count peak areas from one replicate are shown for each reaction condition. N.D. = Not Detected. For products with multiple peaks, peak areas are summed.

Enzymes and Substrates		Representative EIC Peak Areas					
		1	**3x**	**4x**	**2**	**5x/6x**	**7x**
PikAIII WT PikAIV WT	**8/9a**	604644918	7,436,664	19,463,238	6,426,814,005	2,130,190,170	34,365,789
	8/9b	558,290,483	2,784,286	7,078,095	6,829,997,016	2,948,022,986	72,153,536
	8/9c	514,002,858	N.D.	N.D.	7,055,041,197	320,457,815	N.D.
	8/9d	353,277,555	N.D.	N.D.	628,694,156	N.D.	N.D.
PikAIII Y755V PikAIV Y753V	**8/9a**	306,251,903	207,133,468	16,446,420	231,423,866	502,212,154	109,769,795
	8/9b	763,629,093	69,743,628	15,336,705	976,106,367	1,528,345,552	564,559,230
	8/9c	275,694,862	8,959,070	11,751,681	611,588,065	153,840,570	30,925,719
	8/9d	311,277,819	N.D.	41,383,819	608,122,645	35,004,271	N.D.

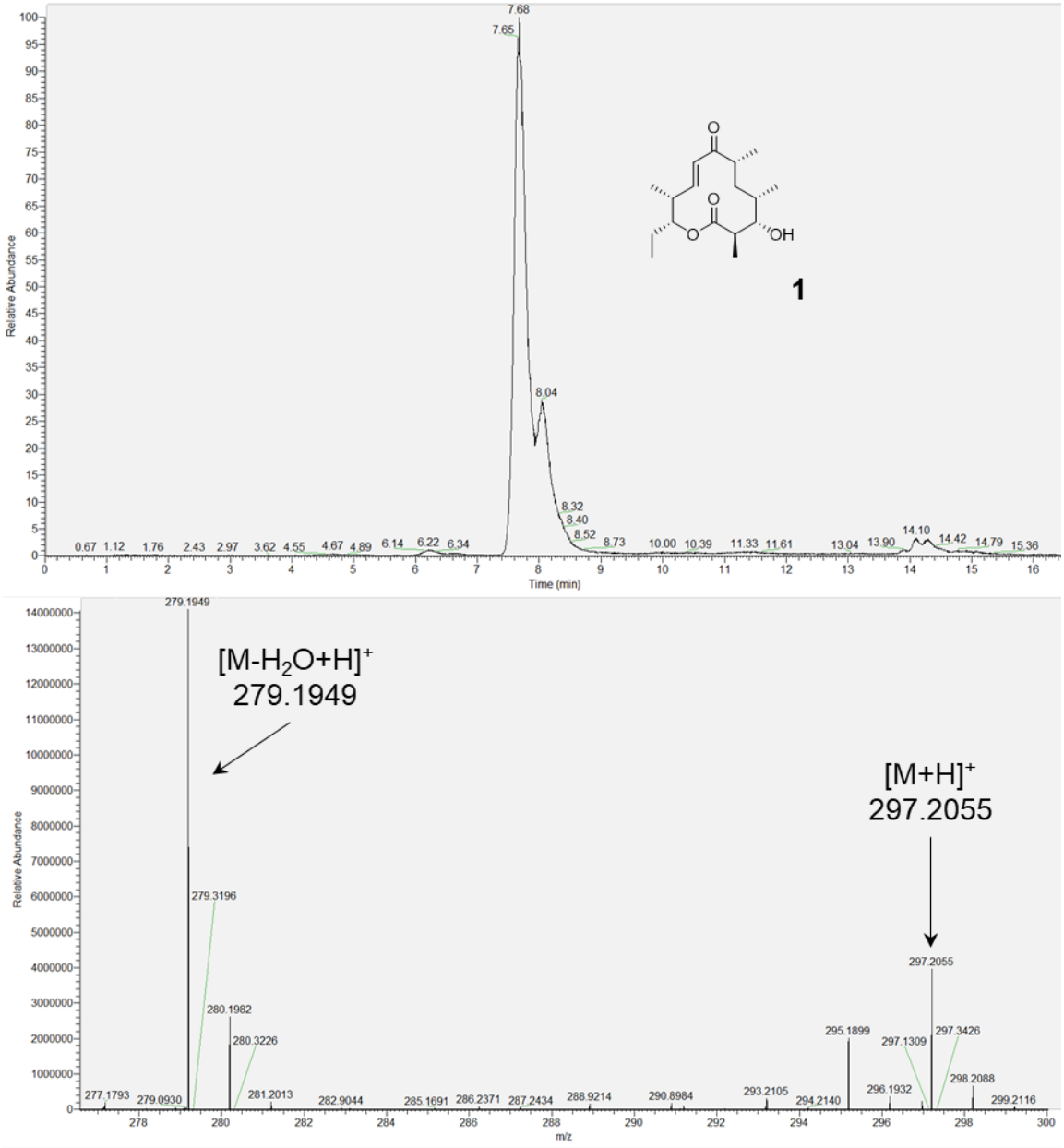
Table C6. High-Resolution LC-MS Peak Areas for PKS-Catalyzed Reaction Products with Substrates 7 and 8a. Extracted ion count peak areas from one replicate are shown for each reaction condition. N.D. = Not Detected. For products with multiple peaks, peak areas are summed.

	Representative EIC Peak Areas					
	1	**3a**	**4a**	**2**	**5a/6a**	**7a**
R1	99,267,500	1,283,686	2,951,168	1,418,188,126	409,872,173	5,118,193
R2	21,455,393	977,191	2,313,400	226,409,954	49,123,149	349,108
R3	79,186,380	1,335,707	160,599,190	933,317,508	570,558,286	10,970,635
R4	360,898,980	1,422,361	1,753,026	--	--	--
R5	N.D.	N.D.	N.D.	--	--	--
R6	660,712,285	29,785,869	28,612,559	--	--	--
R7	200,092,464	16,349,384	4,315,620	--	--	--
R8	N.D.	N.D.	N.D.	--	--	--
R9	247,579,476	1,042,640	1,144,330	--	--	--
R10	57,054,504	881,084	14,630,215	--	--	--
R11	N.D.	N.D.	N.D.	--	--	--

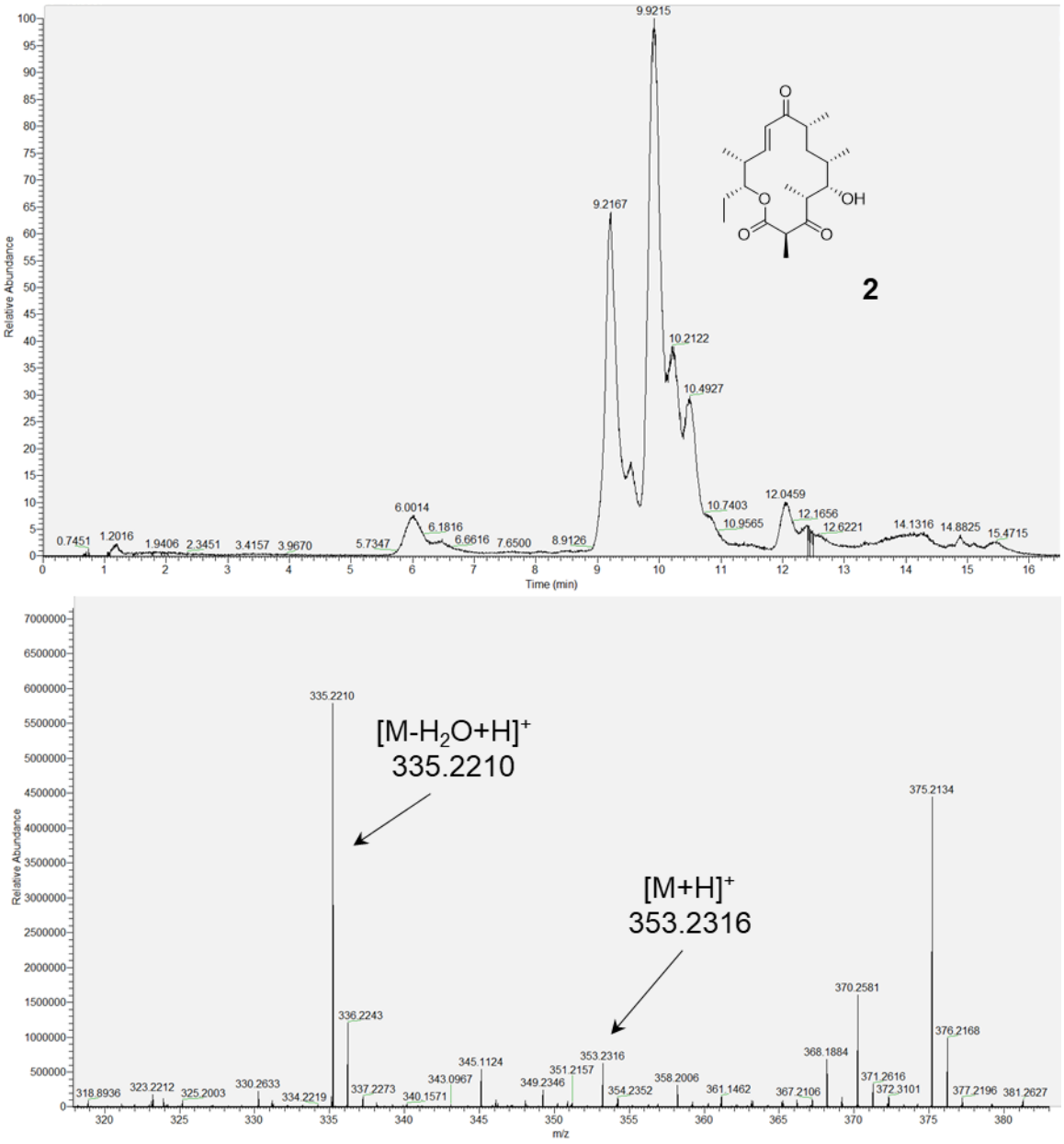
R12	290,068,634	125,867,179	93,320,119	--	--	--
R13	N.D.	N.D.	N.D.	--	--	--
R14	99,267,500	1,283,686	2,951,168	1,418,188,126	409,872,173	5,118,193
R15	799,983,835	46,534,105	21,451,594	10,281,410,822	2,836,675,028	107,026,112
R16	88,167,565	100,155,924	11,338,699	342,849,625	990,486,379	266,850,718
R17	72,284,698	2,134,500	1,235,595	62,733,768	223,361,437	3,860,401
R18	102,893,998	82,829,384	7,769,051	108,989,992	269,734,723	74,090,230
R19	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.

Figure C2. Representative LC-MS Chromatograms of Products from Lysate Module-Catalyzed Reactions. Top panel for each compound shows the extracted ion chromatogram. Bottom panel shows the total ion spectra.

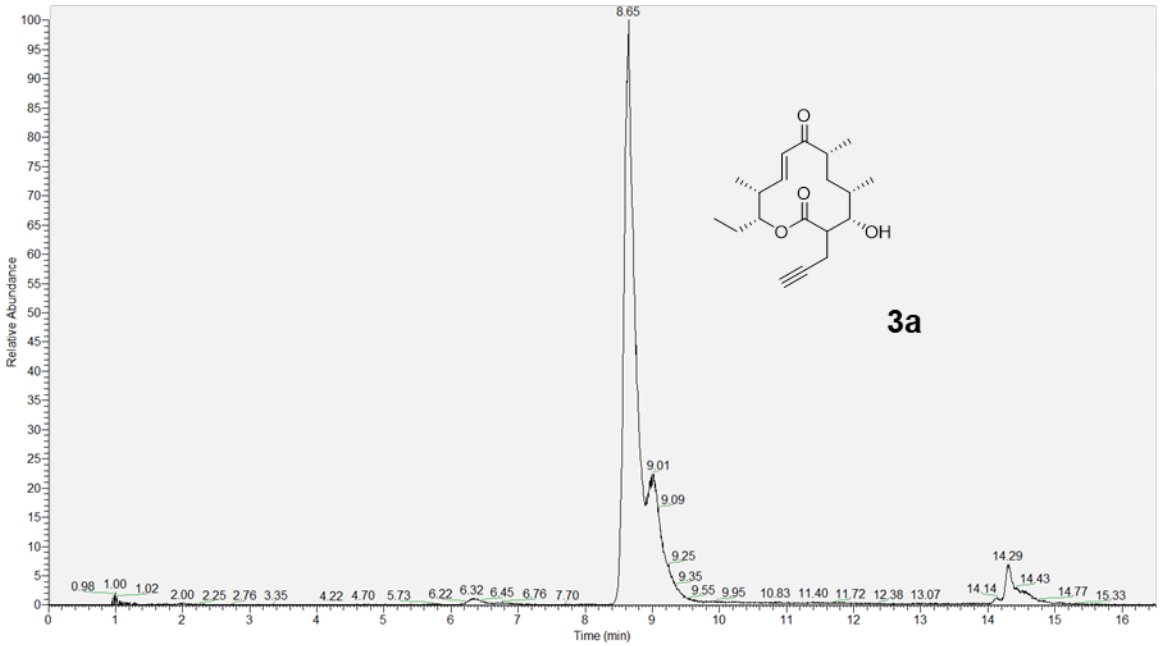
A

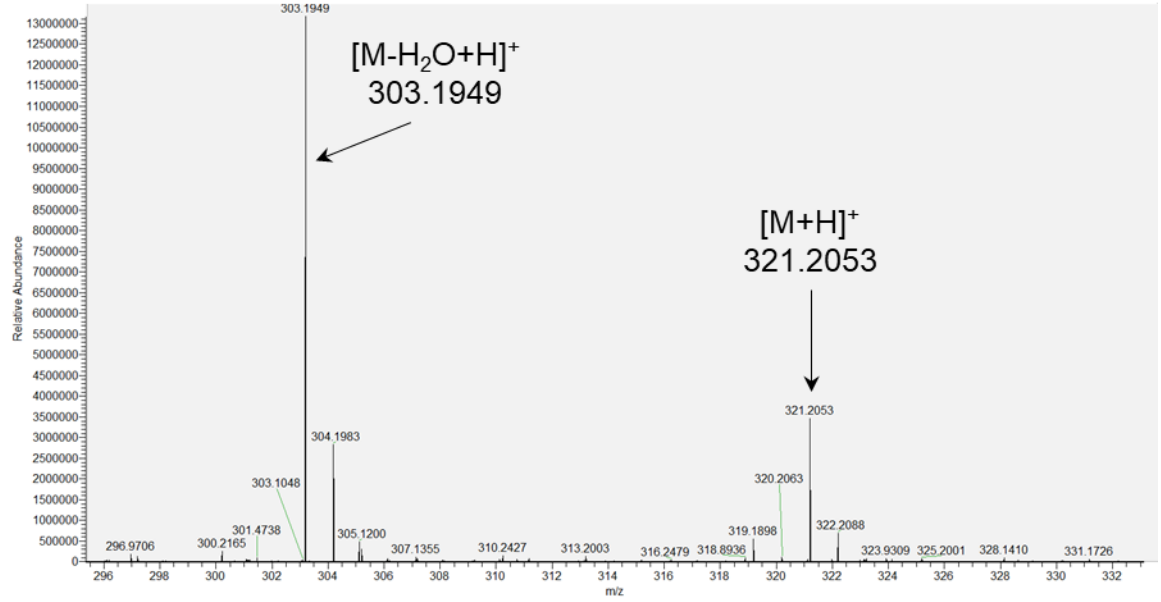


B

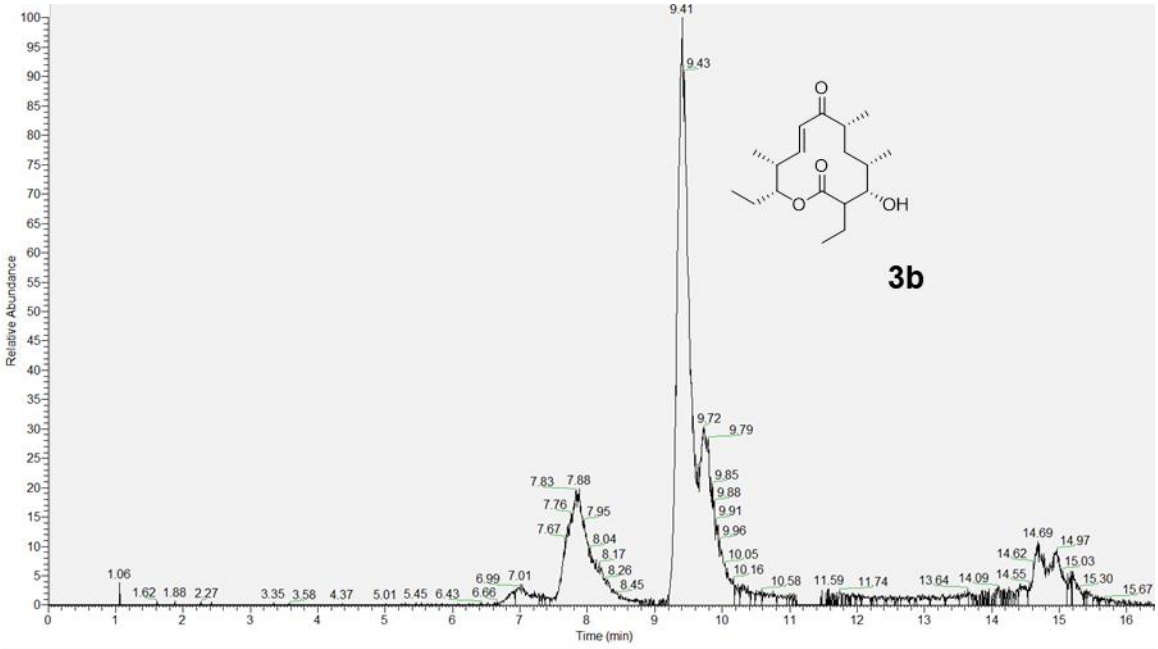

Relative Abundance
Time (min)
9.9215
9.2167
10.2122
10.4927
6.0014
6.1816
6.6616
7.6500
8.9126
5.7347
0.7451
1.2016
1.9406
2.3451
3.4157
3.9670
10.7403
10.9565
12.0459
12.1656
12.6221
14.1316
14.8825
15.4715
O
OH
O
O
O
2
[M-H2O+H]+
335.2210
[M+H]+
353.2316
m/z
335.2210
336.2243
353.2316
375.2134
370.2581
376.2168
345.1124
368.1884
371.2616
372.3101
351.2157
343.0967
349.2346
358.2006
354.2352
361.1462
367.2106
377.2196
381.2627
318.8936
323.2212
325.2003
330.2633
334.2219
337.2273
340.1571

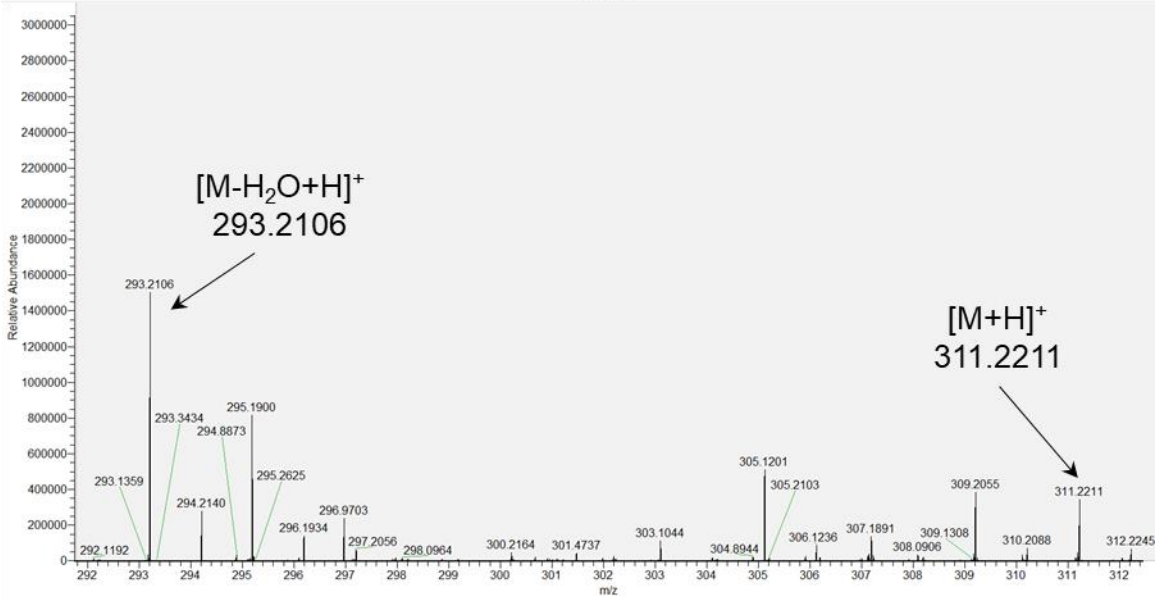
C


8.65
9.01
9.09
9.25
9.35
9.55
9.95
10.83
11.40
11.72
12.38
13.07
14.14
14.29
14.43
14.77
15.33
0.98
1.00
1.02
2.00
2.25
2.76
3.35
4.22
4.70
5.73
6.22
6.32
6.45
6.76
7.70
Relative Abundance
Time (min)
O
O
O
OH
3a

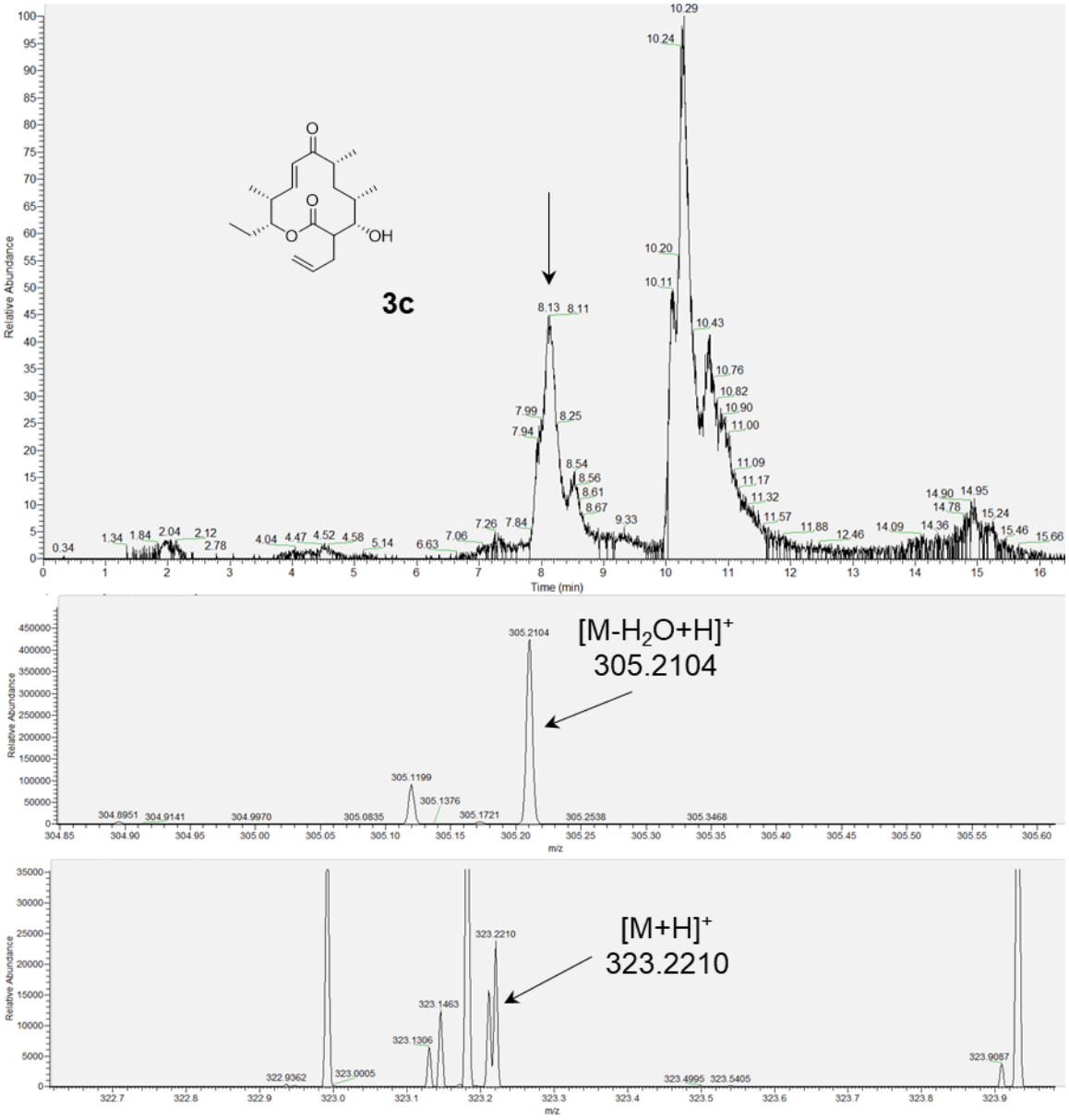

303.1949
[M-H2O+H]+
303.1949
[M+H]+
321.2053
321.2053
304.1983
303.1048
320.2063
301.4738
305.1200
319.1898
322.2088
296.9706
300.2165
307.1355
310.2427
313.2003
316.2479
318.8936
323.9309
325.2001
328.1410
331.1726
Relative Abundance
m/z

D

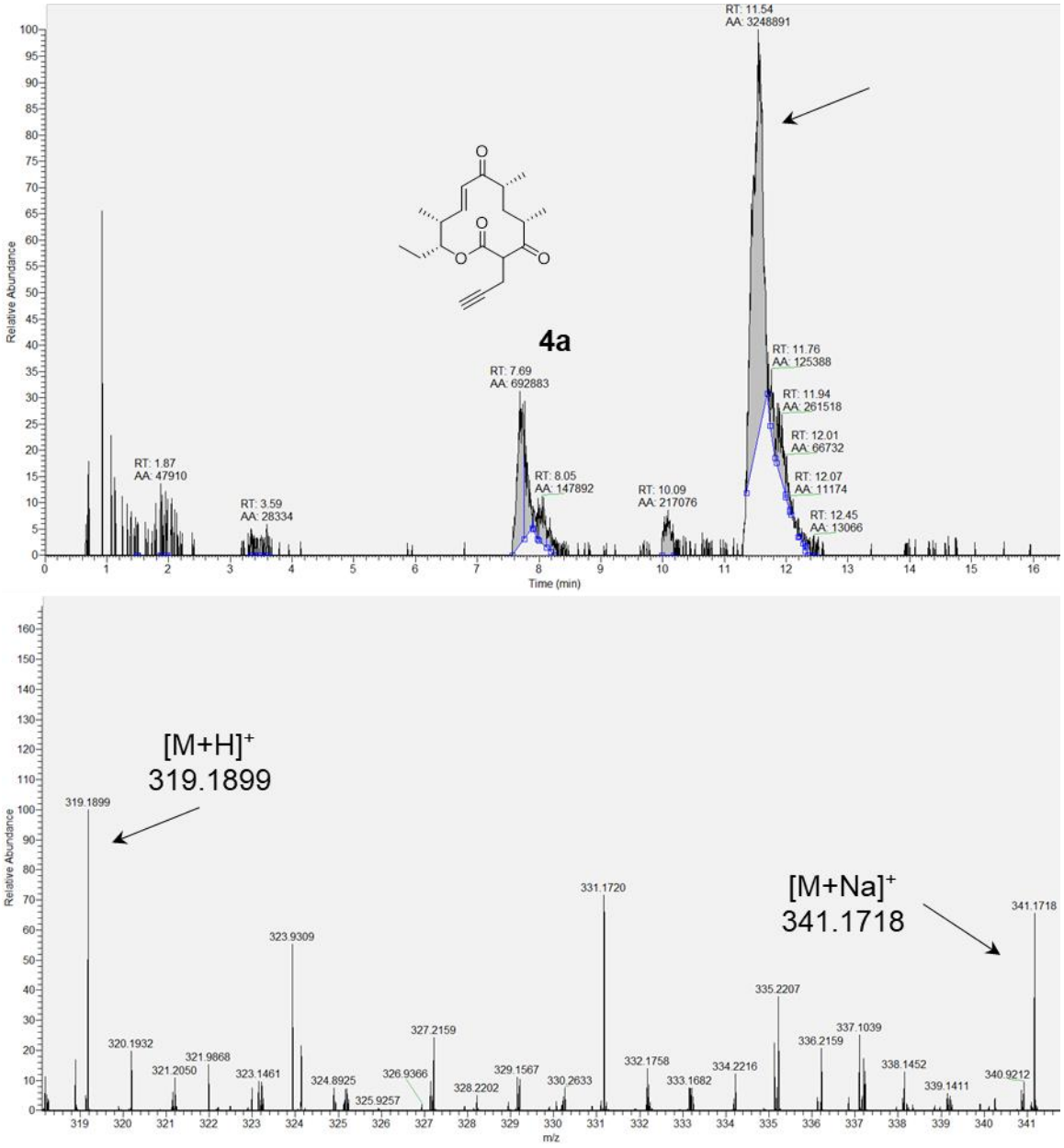

Relative Abundance
Time (min)
3b
9.41
9.43
9.72
9.79
7.83
7.88
7.76
7.95
7.67
8.04
8.17
8.26
8.45
14.69
14.97
14.62
15.03
14.55
15.30
15.67


[M-H2O+H]+
293.2106
[M+H]+
311.2211
Relative Abundance
m/z
293.2106
293.3434
295.1900
294.8873
293.1359
295.2625
294.2140
296.9703
296.1934
297.2056
292.1192
298.0964
300.2164
301.4737
303.1044
304.8944
305.1201
305.2103
306.1236
307.1891
309.2055
309.1308
308.0906
310.2088
311.2211
312.2245

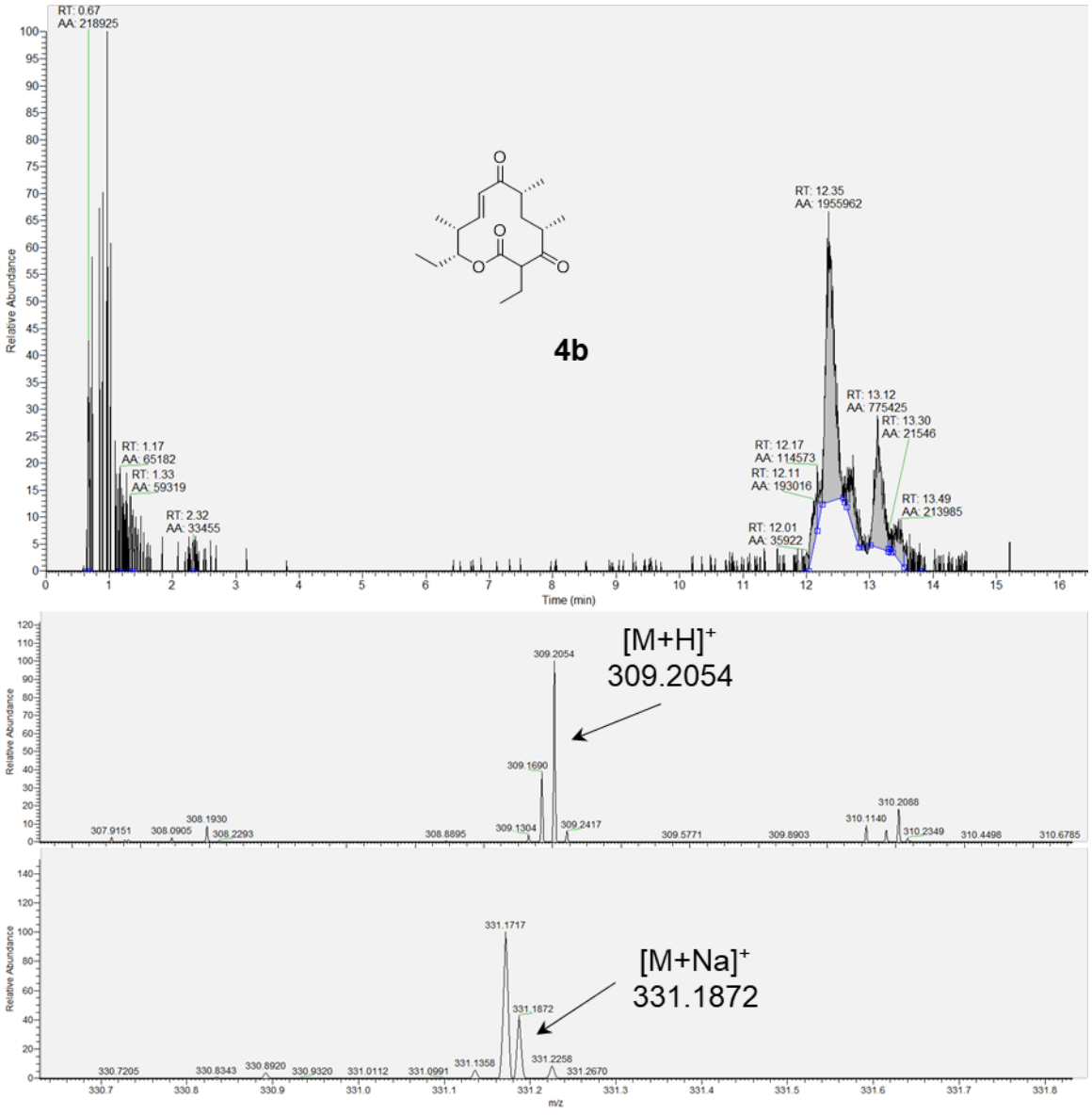
E



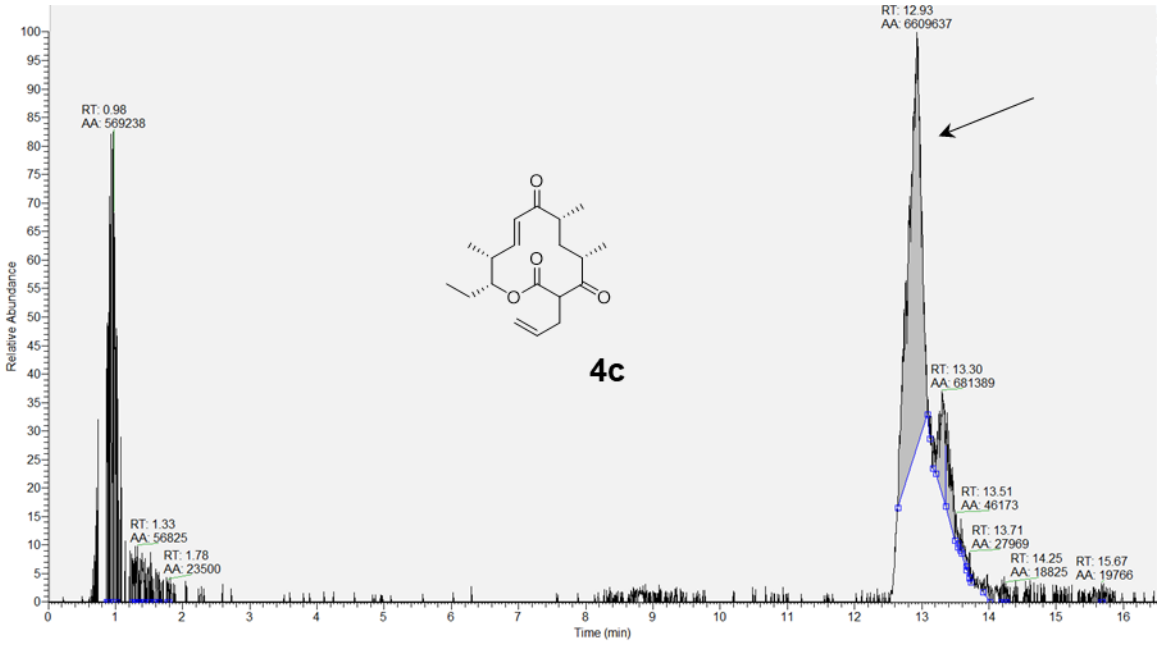
F

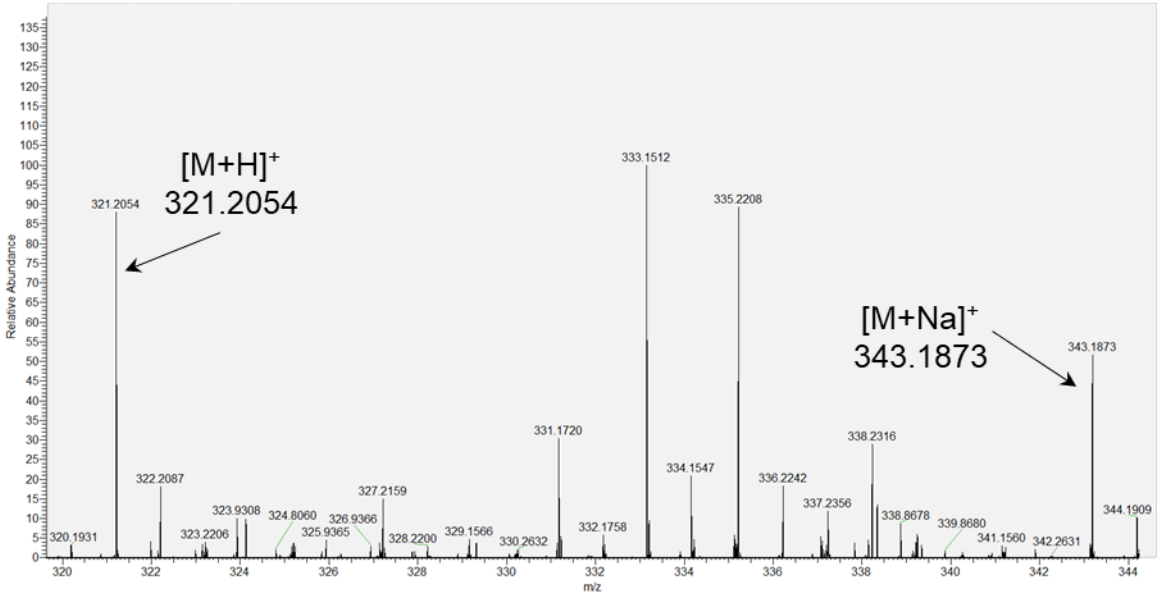

RT: 11.54
AA: 3248891
4a
RT: 7.69
AA: 692883
RT: 11.76
AA: 125388
RT: 11.94
AA: 261518
RT: 12.01
AA: 66732
RT: 12.07
AA: 11174
RT: 12.45
AA: 13066
RT: 1.87
AA: 47910
RT: 3.59
AA: 28334
RT: 8.05
AA: 147892
RT: 10.09
AA: 217076
Relative Abundance
Time (min)
[M+H]+
319.1899
[M+Na]+
341.1718
319.1899
331.1720
323.9309
341.1718
335.2207
327.2159
337.1039
320.1932
336.2159
321.9868
321.2050
323.1461
332.1758
334.2216
338.1452
324.8925
326.9366
333.1682
339.1411
329.1567
340.9212
328.2202
330.2633
325.9257
m/z

G

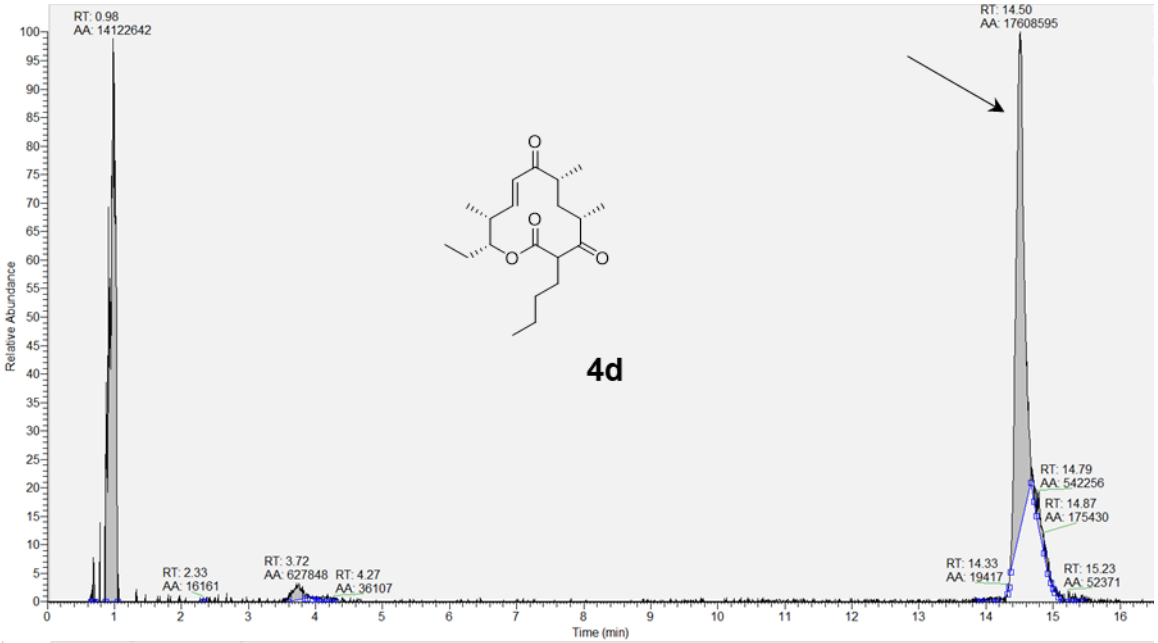

RT: 0.67
AA: 218925
RT: 1.17
AA: 65182
RT: 1.33
AA: 59319
RT: 2.32
AA: 33455
RT: 12.35
AA: 1955962
RT: 13.12
AA: 775425
RT: 13.30
AA: 21546
RT: 12.17
AA: 114573
RT: 12.11
AA: 193016
RT: 13.49
AA: 213985
RT: 12.01
AA: 35922
Relative Abundance
Time (min)
4b
[M+H]+
309.2054
309.2054
309.1690
310.2088
[M+Na]+
331.1872
331.1717
331.1872
m/z

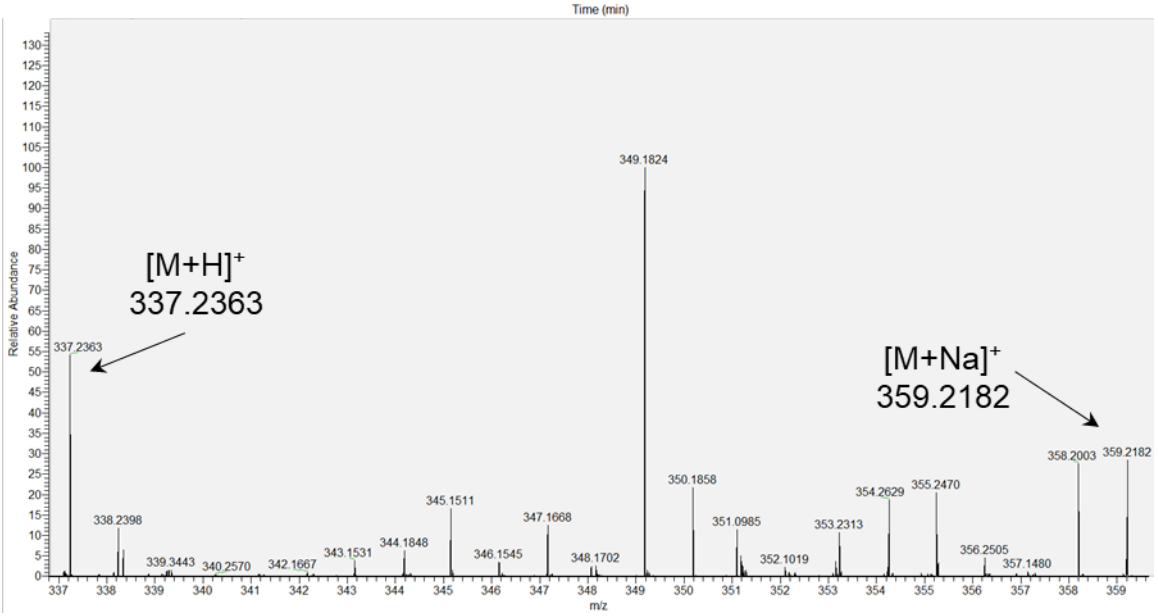
H


RT: 12.93
AA: 6609637
RT: 0.98
AA: 569238
4c
RT: 13.30
AA: 681389
RT: 13.51
AA: 46173
RT: 13.71
AA: 27969
RT: 14.25
AA: 18825
RT: 15.67
AA: 19766
RT: 1.33
AA: 56825
RT: 1.78
AA: 23500
Relative Abundance
Time (min)

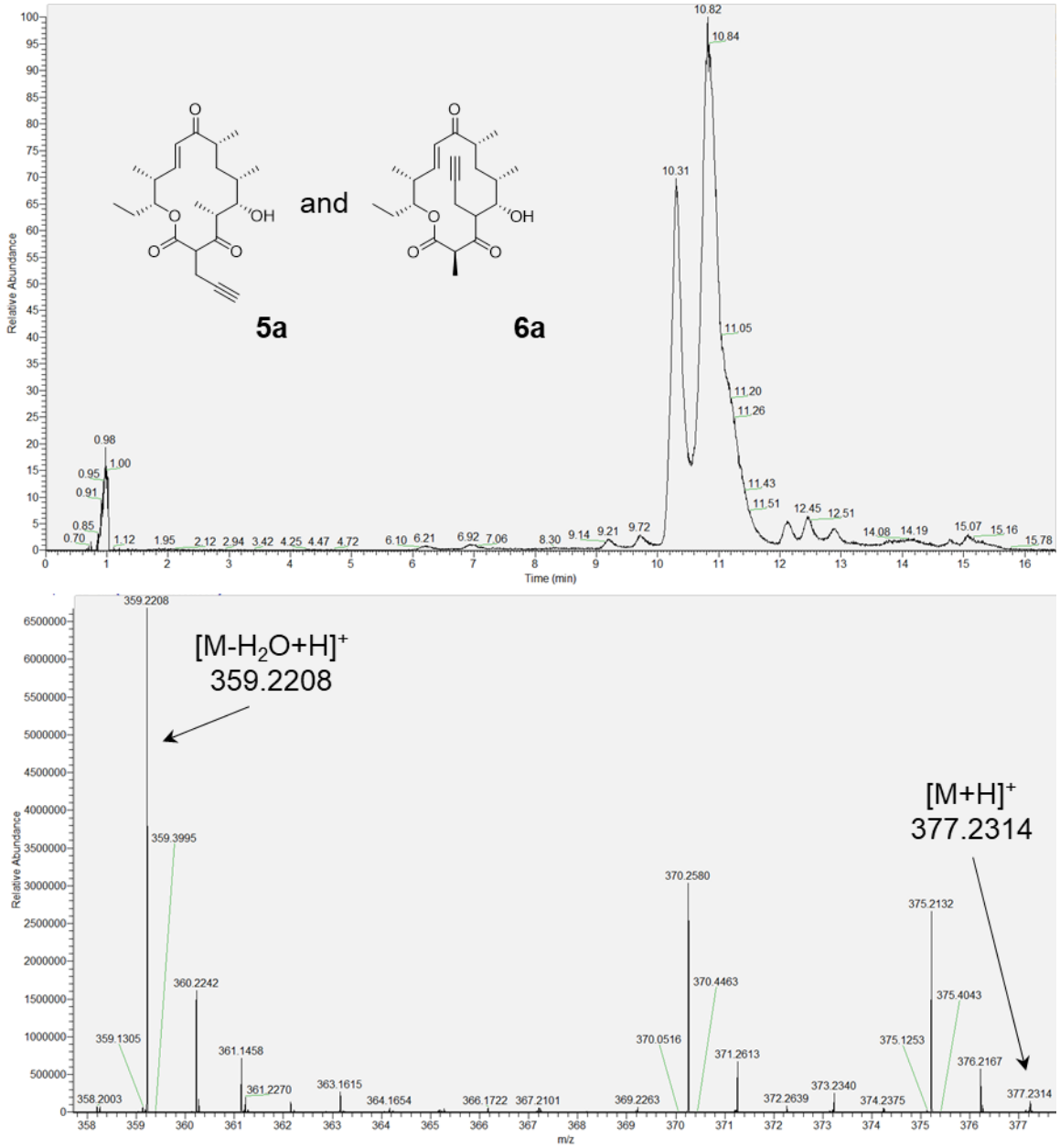

[M+H]+
321.2054
[M+Na]+
343.1873
Relative Abundance
m/z

I

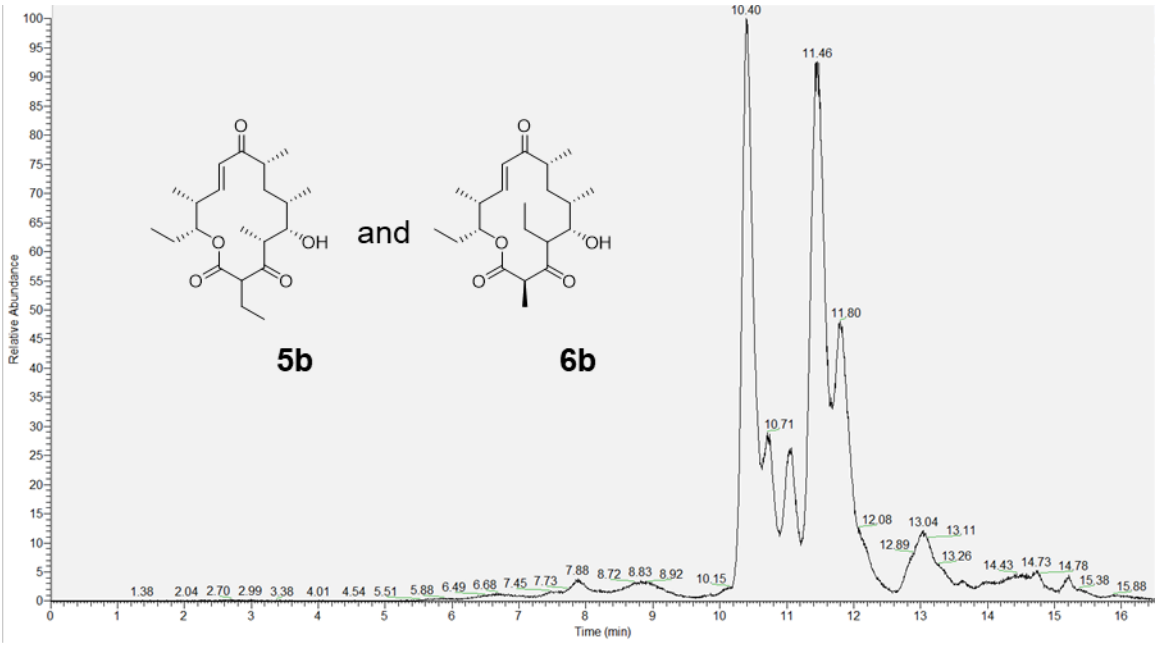

RT: 0.98
AA: 14122642
RT: 14.50
AA: 17608595
RT: 14.79
AA: 542256
RT: 14.87
AA: 175430
RT: 2.33
AA: 16161
RT: 3.72
AA: 627848
RT: 4.27
AA: 36107
RT: 14.33
AA: 19417
RT: 15.23
AA: 52371
Relative Abundance
Time (min)
4d

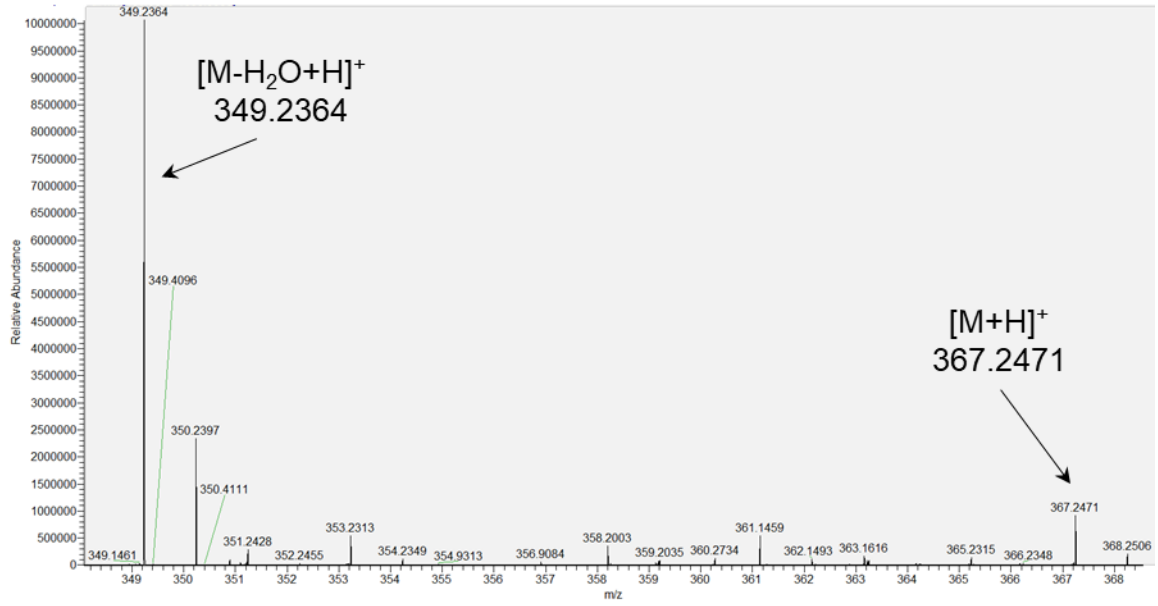

[M+H]+
337.2363
[M+Na]+
359.2182
349.1824
337.2363
359.2182
Relative Abundance
m/z

J

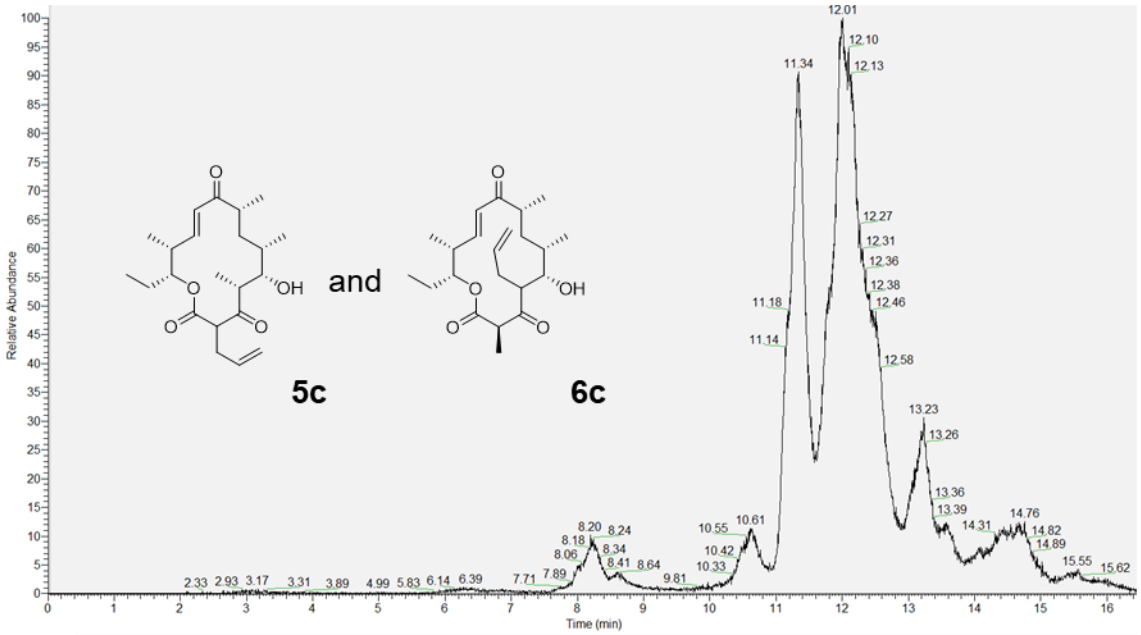

and
5a
6a
Relative Abundance
Time (min)
10.82
10.84
10.31
11.05
11.20
11.26
11.43
11.51
12.45
12.51
[M-H2O+H]+
359.2208
[M+H]+
377.2314
359.3995
360.2242
359.1305
361.1458
361.2270
363.1615
364.1654
366.1722
367.2101
369.2263
370.2580
370.4463
370.0516
371.2613
372.2639
373.2340
374.2375
375.2132
375.4043
375.1253
376.2167
377.2314
358.2003
m/z

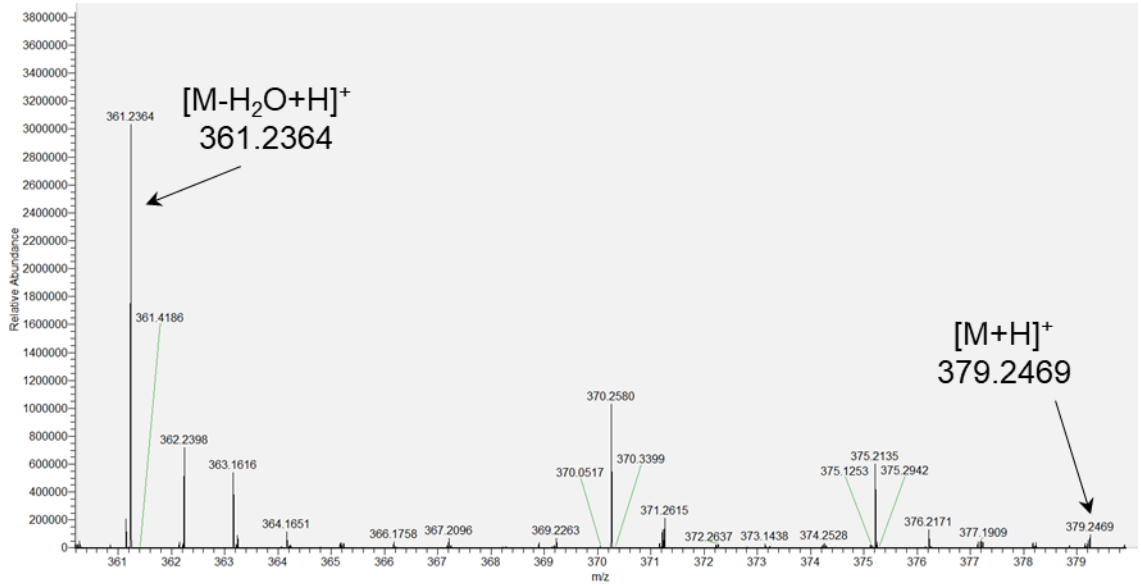
K


5b
and
6b
Relative Abundance
Time (min)
10.40
11.46
11.80
10.71
12.08
13.04
13.11
12.89
13.26
14.43
14.73
14.78
15.38
15.88
1.38
2.04
2.70
2.99
3.38
4.01
4.54
5.51
5.88
6.49
6.68
7.45
7.73
7.88
8.72
8.83
8.92
10.15

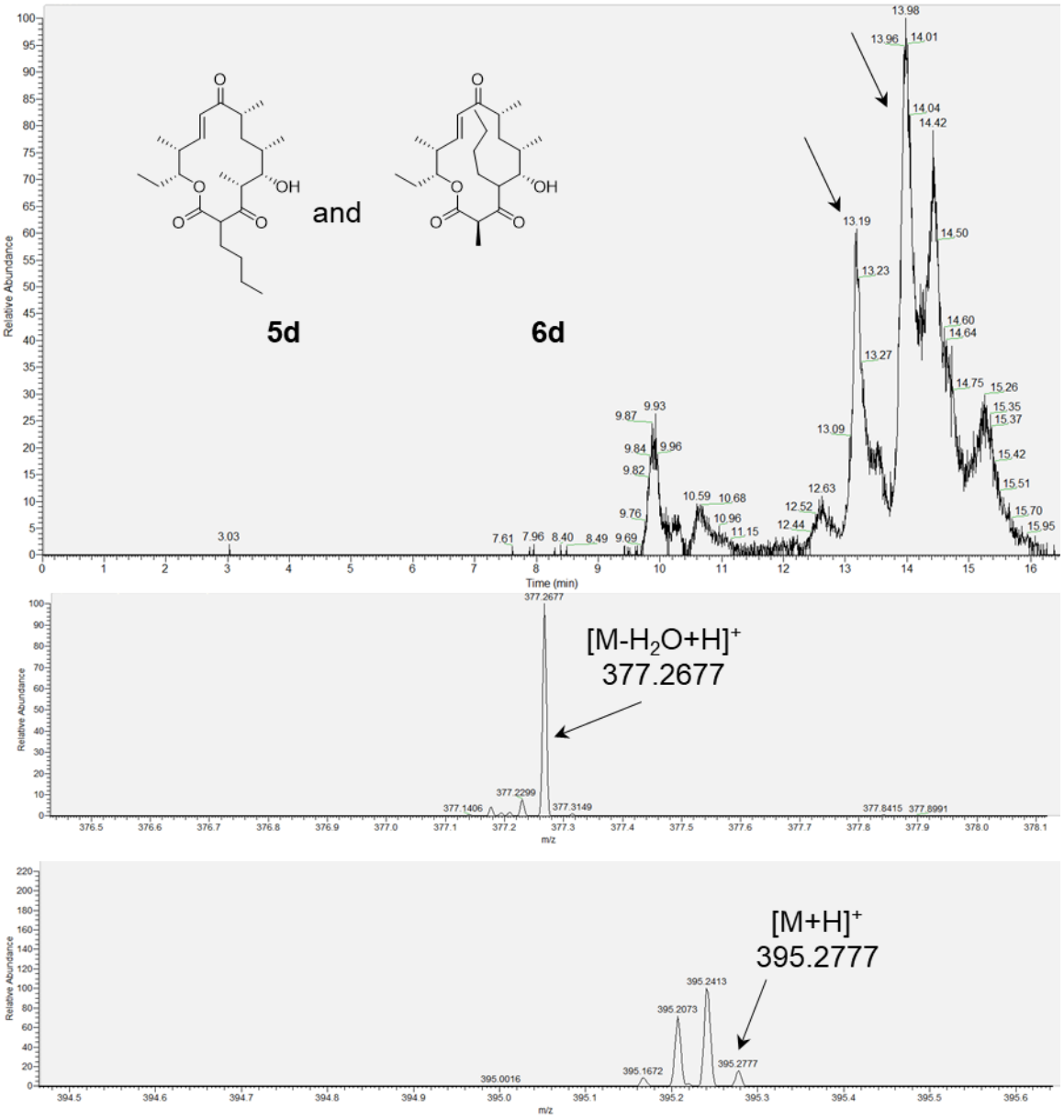

[M-H2O+H]+
349.2364
[M+H]+
367.2471
Relative Abundance
m/z
349.2364
349.4096
350.2397
350.4111
351.2428
352.2455
353.2313
354.2349
354.9313
356.9084
358.2003
359.2035
360.2734
361.1459
362.1493
363.1616
365.2315
366.2348
367.2471
368.2506
349.1461

L

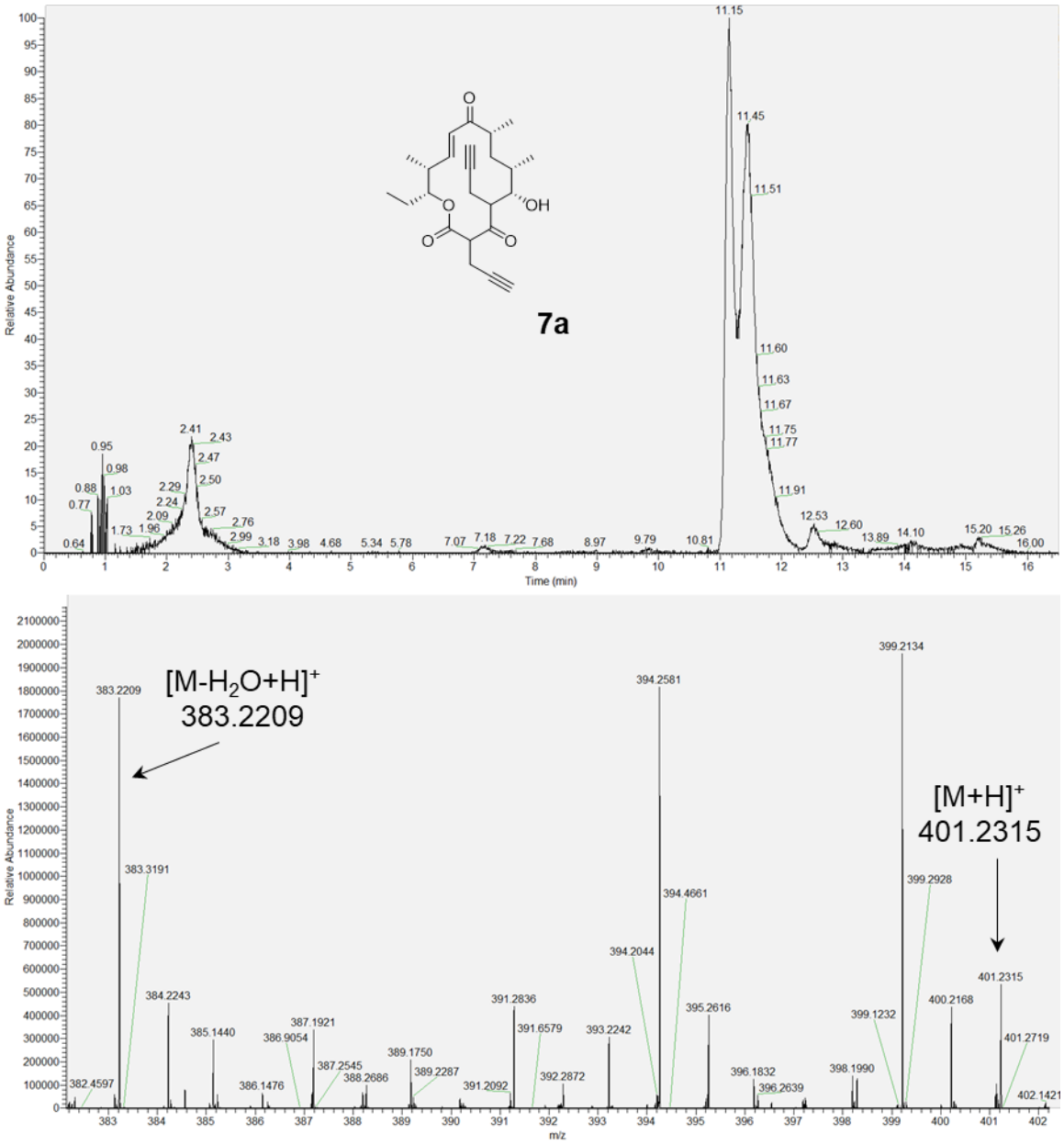

and
5c
6c
Relative Abundance
Time (min)


[M-H2O+H]+
361.2364
[M+H]+
379.2469
Relative Abundance
m/z

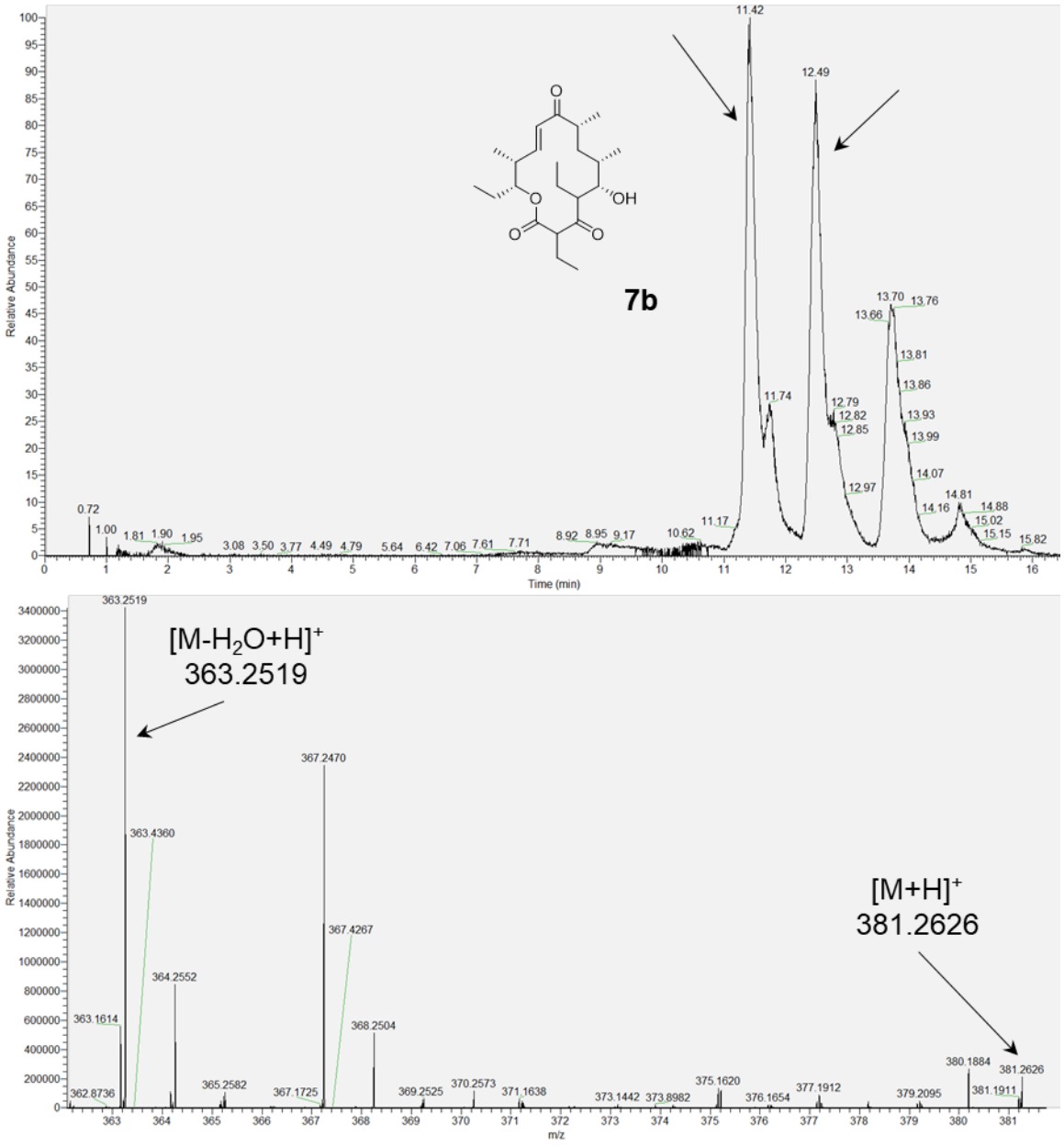
M


and
5d
6d
Relative Abundance
Time (min)
[M-H2O+H]+
377.2677
m/z
[M+H]+
395.2777

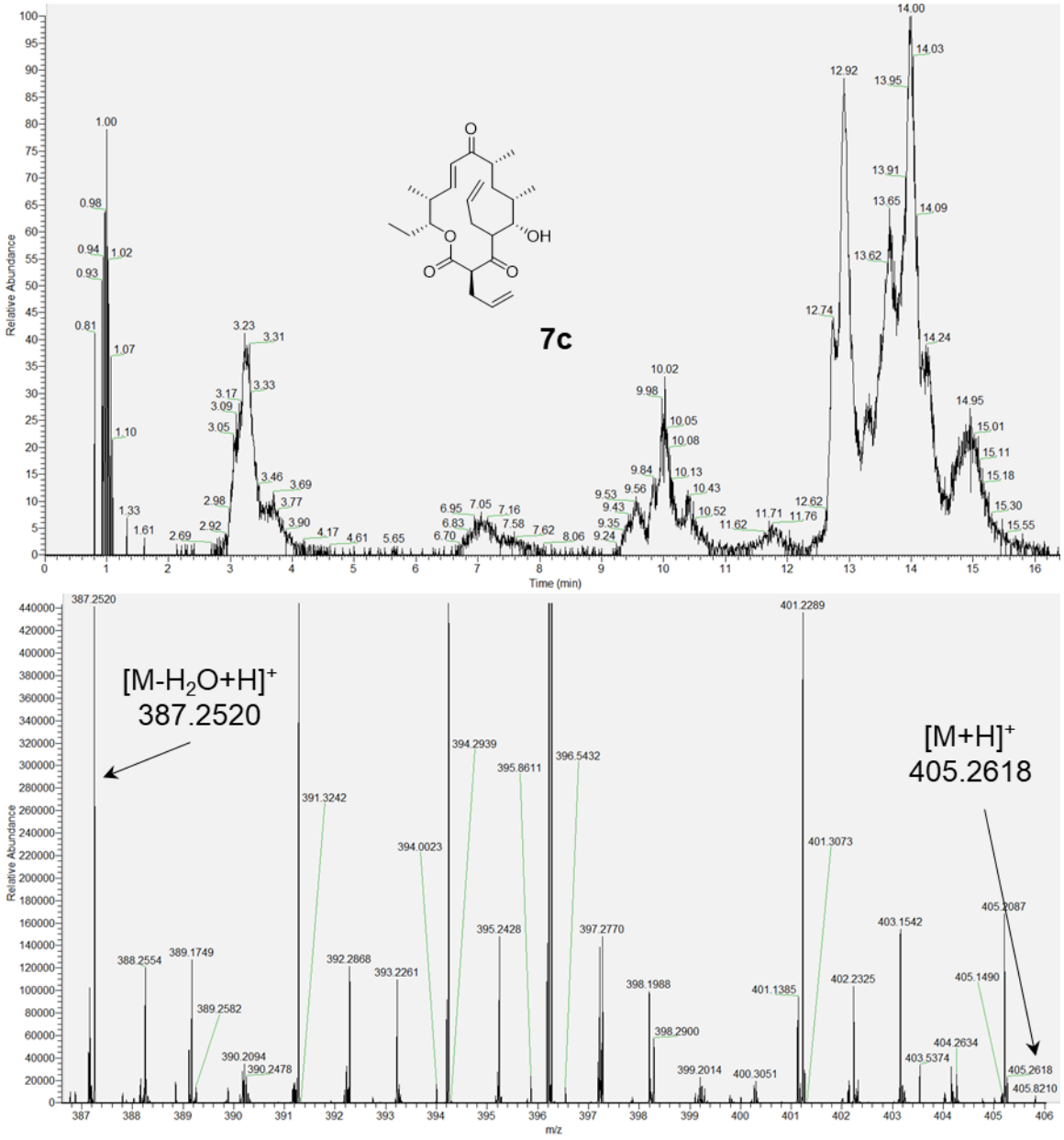
N



O


7b
Relative Abundance
Time (min)
11.42
12.49
11.74
12.79
12.82
12.85
12.97
13.70
13.76
13.66
13.81
13.86
13.93
13.99
14.07
14.81
14.16
14.88
15.02
15.15
15.82
0.72
1.00
1.81
1.90
1.95
3.08
3.50
3.77
4.49
4.79
5.64
6.42
7.06
7.61
7.71
8.92
8.95
9.17
10.62
11.17
[M-H2O+H]+
363.2519
[M+H]+
381.2626
m/z
363.2519
367.2470
363.4360
367.4267
364.2552
363.1614
368.2504
362.8736
365.2582
367.1725
369.2525
370.2573
371.1638
373.1442
373.8982
375.1620
376.1654
377.1912
379.2095
380.1884
381.2626
381.1911

P


7c
OH
Relative Abundance
Time (min)
[M-H2O+H]+
387.2520
[M+H]+
405.2618
m/z

Appendix D

Table D1. *E. coli* Strains and Genotypes.

Bacterial Strain	Genotype
DH5α	F$^-$ φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r_K^-, m_K^+) *pho*A *sup*E44 λ$^-$ *thi*-1 *gyr*A96 *rel*A1
***E. cloni*® 10G**	F$^-$ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) *end*A1 *rec*A1 φ80d*lac*ZΔM15 Δ*lac*X74 *ara*D139 Δ(*ara,leu*)7697 *gal*U *gal*K *rps*L(Str^R) *nup*G λ-*ton*A
TOP10	F$^-$ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ$^-$ *rps*L(Str^R) *end*A1 *nup*G

Table D2. Sequences of pSENSE2FF and pSENSE2FFM.

Construct	Nucleotide Sequences
FapR_pET28a	TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGG TGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCT CCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCG TCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTAC GGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGG CCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTT CTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCT CGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGG TTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAAT ATTAACGTTTACAATTTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAA CCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATG AATTAATTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTAT TCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAAT GAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTA TCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCC CCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACT GAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTC AACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAAC CGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTG TTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACAC TGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATA CCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCA TCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGT CAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTAC CTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAAT CGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATA CCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAG ACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATG TAAGCAGACAGTTTTATTGTTCATGACCAAAATCCCTTAACGTGAGTTTT

	CGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGA GATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACC GCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTC CGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTA GTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTAC ATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATA AGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCG CAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCG AACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCG CCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGG GTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTA TCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTT TGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCG GCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTT TCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGT GAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTG AGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCT GTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTC TGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTG GGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGAC GGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCC GGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAG GCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGT CTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAAT GTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTT GGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGA TACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAA CATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGAT GCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTA ATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAG ATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTA CGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACG TTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTC TGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAG GAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCT GCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCG AGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGC GCTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCA CCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACG ATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCT CAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACAT TAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGC CAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTAT TGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGA TTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCT GGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGA TATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCC

	GCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGC CATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCA GCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCC CGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCC AGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCG CGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTA CCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGAC ATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGG CATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGC GCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTC TACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAA TCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCA ACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTT GGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTT TCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAA GAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATT CACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAA AGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATG CGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCA CCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCC CCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAG CCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGG CGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCG GCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATA GGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAA CTTTAAGAAGGAGATATACCATGGGGCGCCGCAACAAACGCGAACGTCAA GAGCTGTTACAGCAAACCATCCAGGCGACGCCGTTTATTACCGACGAAGA ATTGGCAGGTAAGTTCGGCGTCTCGATTCAGACAATCCGTCTGGATCGTC TGGAGTTGTCCATTCCGGAATTGCGTGAACGTATTAAAAACGTCGCCGAA AAGACGTTAGAAGACGAGGTGAAATCTTTATCATTGGACGAGGTAATTGG CGAAATTATTGACCTTGAATTAGACGACCAGGCTATTTCAATTTTGGAGA TTAAACAGGAACACGTGTTCAGTCGTAACCAGATCGCCCGCGGTCATCAT CTGTTCGCGCAGGCCAACAGCCTGGCTGTGGCTGTTATTGATGACGAACT TGCGCTGACCGCTTCGGCAGACATCCGCTTTACCCGTCAGGTGAAACAAG GCGAGCGCGTCGTAGCGAAAGCTAAAGTTACCGCGGTTGAGAAGGAAAAA GGTCGTACGGTTGTGGAAGTCAACTCATACGTCGGCGAAGAAATCGTGTT TTCAGGACGTTTTGATATGTACCGCAGCAAACACAGCCTCGAGCACCACC ACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAG TTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTC TAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT
pSENSE2FF	GGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGT GAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTG GCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACG CTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGT TTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTT ACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCA TAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGC

	TGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCC GGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACT GGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTG CTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACA GTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGT TGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTT TTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGAT CCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACG TTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCC TTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATCCATGGATAT GAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTAT CTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCG TGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCA ATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAA CCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCG CCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCG CCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGT GTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGAT CAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCC TTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACT CATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAA GATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAG TGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAACACGGGATAATAC CGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTT CGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATG TAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAG CGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAA TAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATAT TATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGA ATGTATTTAGAAAAATAAACAAAAAGAGCATGCGTTTGTAGAAACGCAAA AAGGCCATCCGTCAGGATGGCCTTCTGCTTAATTTGATGCCTGGCAGTTT ATGGCGGGCGTCCTGCCCGCCACCCTCCGGGCCGTTGCTTCGCAACGTTC AAATCCGCTCCCGGCGGATTTGTCCTACTCAGGAGAGCGTTCACCGACAA ACAACAGATAAAACGAAAGGCCCAGTCTTTCGACTGAGCCTTTCGTTTTA TTTGATGCCTGGCAGTTCCCTACTCTCGCATGGGGAGACCCCACACTACC ATCGGCGCTACGGCGTTTCACTTCTGAGTTCGGCATGGGGTCAGGTGGGA CCACCGCGCTACTGCCGCCAGGCAAATTCTGTTTTATCAGACCGCTTCTG CGTTCTGATTTAATCTGTATCAGGCTGAAAATCTTCTCTCATCCGCCAAA ACAGCCAAGCTGGAGACCGTTTAAACGGGCCCAAGCTTTTTGTAGAGCTC ATCCATGCCATGTGTAATCCCAGCAGCAGTTACAAACTCAAGAAGGACCA TGTGGTCACGCTTTTCGTTGGGATCTTTCGAAAGGACAGATTGTGTCGAC AGGTAATGGTTGTCTGGTAAAAGGACAGGGCCATCGCCAATTGGAGTATT TTGTTGATAATGGTCTGCTAGTTGAACGGAACCATCTTCAACGTTGTGGC GAATTTTGAAGTTAGCTTTGATTCCATTCTTTTGTTTGTCTGCCGTGATG TATACATTGTGTGAGTTAAAGTTGTACTCGAGTTTGTGTCCAAGAATGTT TCCATCTTCTTTAAAATCAATACCCTTTAACTCGATACGATTAACAAGGG TATCACCTTCAAACTTGACTTCAGCACGCGTCTTGTAGGTCCCGTCATCT

	TTGAAAGATATAGTGCGTTCCTGTACATAACCTTCGGGCATGGCACTCTT GAAAAAGTCATGCCGTTTCATGTGATCCGGATAACGGGAAAAGCATTGAA CACCATAGGTCAGAGTAGTGACAAGTGTTGGCCACGGAACAGGTAGTTTT CCAGTAGTGCAAATAAATTTAAGGGTGAGTTTTCCGTTTGTAGCATCACC TTCACCCTCTCCACGGACAGAAAATTTGTGCCCATTAACATCACCATCTA ATTCAACAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTC ATACTAGTAACTCCTTAGCTGTTTCGGATGCCGGACAATTAAGACTAGGT ACTAATAGTCCTAGGCAACATACGAGCCGGAAGCATAAAGTGTAAAGCCT GGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCTAGA TGGTGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCAACGATCCTCCA CTCCGCGGCAAGGAGGTTGCCATATGCGCCGCAACAAACGCGAACGTCAA GAGCTGTTACAGCAAACCATCCAGGCGACGCCGTTTATTACCGACGAAGA ATTGGCAGGTAAGTTCGGCGTCTCGATTCAGACAATCCGTCTGGATCGTC TGGAGTTGTCCATTCCGGAATTGCGTGAACGTATTAAAAACGTCGCCGAA AAGACGTTAGAAGACGAGGTGAAATCTTTATCATTGGACGAGGTAATTGG CGAAATTATTGACCTTGAATTAGACGACCAGGCTATTTCAATTTTGGAGA TTAAACAGGAACACGTGTTCAGTCGTAACCAGATCGCCCGCGGTCATCAT CTGTTCGCGCAGGCCAACAGCCTGGCTGTGGCTGTTATTGATGACGAACT TGCGCTGACCGCTTCGGCAGACATCCGCTTTACCCGTCAGGTGAAACAAG GCGAGCGCGTCGTAGCGAAAGCTAAAGTTACCGCGGTTGAGAAGGAAAAA GGTCGTACGGTTGTGGAAGTCAACTCATACGTCGGCGAAGAAATCGTGTT TTCAGGACGTTTTGATATGTACCGCAGCAAACACAGCTAAGGTACCCATA GTATCCAAAATAAGGCTAAGAATTCATGGTCAGCAAGGGCGAAGAGGACA ATATGGCGATCATCAAGGAGTTTATGAGATTCAAGGTACACATGGAGGGT AGCGTTAACGGCCACGAGTTTGAAATCGAGGGTGAGGGCGAGGGTCGCCC GTACGAGGGGACTCAGACGGCAAAGTTAAAAGTTACTAAAGGTGGTCCGC TTCCTTTTGCGTGGGACATTCTTTCTCCGCAGTTTATGTATGGGAGTAAA GCGTACGTCAAGCATCCAGCCGACATACCAGACTACCTTAAATTATCGTT CCCCGAGGGGTTCAAGTGGGAGCGCGTGATGAACTTCGAAGATGGCGGAG TGGTCACTGTGACCCAGGACTCCTCCCTTCAGGATGGGGAGTTCATATAT AAAGTGAAACTTCGGGGGACCAATTTCCCGTCAGACGGCCCTGTCATGCA GAAAAAGACTATGGGATGGGAAGCATCGAGCGAACGTATGTACCCAGAAG ATGGGGCGTTAAAGGGGGAGATCAAACAGCGTCTGAAATTAAAAGACGGT GGACACTACGATGCTGAGGTCAAAACTACTTACAAGGCTAAAAAACCCGT GCAGCTTCCCGGTGCCTATAATGTTAATATTAAGTTGGACATCACGTCCC ATAACGAAGACTATACAATTGTAGAACAGTATGAGCGGGCCGAAGGACGC CACTCAACTGGAGGGATGGATGAGTTGTACAAATAGCTTAAGGTGTGCAG AGTCCCTGCGGCAGGCGACGAACACGACCGTCGTCGATTAGTACCGGTAC GGTCGGTGGTATCGAAGTCTTGATCACTGTACACTAGA
pSENSE2FFM	ACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAA AAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTT TTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAG TCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCC CTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGA TACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTC ACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCT GTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAAC

	TATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGC AGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAG AGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTT GGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAG CTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTT GCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTG ATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGG GATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAA ATTAAAAATGAAGTTTTAAATCAATCTAAAGTATCCATGGATATGAGTAA ACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGC GATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGA TAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATA CCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCC AGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCA TCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTT AATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACG CTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGC GAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGT CCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGT TATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCT TTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATG CGGCGACCGAGTTGCTCTTGCCCGGCGTCAACACGGGATAATACCGCGCC ACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGC GAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCC ACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTC TGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGG CGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGA AGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTAT TTAGAAAAATAAACAAAAAGAGCATGCGTTTGTAGAAACGCAAAAAGGCC ATCCGTCAGGATGGCCTTCTGCTTAATTTGATGCCTGGCAGTTTATGGCG GGCGTCCTGCCCGCCACCCTCCGGGCCGTTGCTTCGCAACGTTCAAATCC GCTCCCGGCGGATTTGTCCTACTCAGGAGAGCGTTCACCGACAAACAACA GATAAAACGAAAGGCCCAGTCTTTCGACTGAGCCTTTCGTTTTATTTGAT GCCTGGCAGTTCCCTACTCTCGCATGGGGAGACCCCACACTACCATCGGC GCTACGGCGTTTCACTTCTGAGTTCGGCATGGGGTCAGGTGGGACCACCG CGCTACTGCCGCCAGGCAAATTCTGTTTTATCAGACCGCTTCTGCGTTCT GATTTAATCTGTATCAGGCTGAAAATCTTCTCTCATCCGCCAAAACAGCC AAGCTGGAGACCGTTTAAACGGGCCCAAGCTTTTTGTAGAGCTCATCCAT GCCATGTGTAATCCCAGCAGCAGTTACAAACTCAAGAAGGACCATGTGGT CACGCTTTTCGTTGGGATCTTTCGAAAGGACAGATTGTGTCGACAGGTAA TGGTTGTCTGGTAAAAGGACAGGGCCATCGCCAATTGGAGTATTTTGTTG ATAATGGTCTGCTAGTTGAACGGAACCATCTTCAACGTTGTGGCGAATTT TGAAGTTAGCTTTGATTCCATTCTTTTGTTTGTCTGCCGTGATGTATACA TTGTGTGAGTTAAAGTTGTACTCGAGTTTGTGTCCAAGAATGTTTCCATC TTCTTTAAAATCAATACCCTTTAACTCGATACGATTAACAAGGGTATCAC CTTCAAACTTGACTTCAGCACGCGTCTTGTAGGTCCCGTCATCTTTGAAA GATATAGTGCGTTCCTGTACATAACCTTCGGGCATGGCACTCTTGAAAAA

	GTCATGCCGTTTCATGTGATCCGGATAACGGGAAAAGCATTGAACACCAT AGGTCAGAGTAGTGACAAGTGTTGGCCACGGAACAGGTAGTTTTCCAGTA GTGCAAATAAATTTAAGGGTGAGTTTTCCGTTTGTAGCATCACCTTCACC CTCTCCACGGACAGAAAATTTGTGCCCATTAACATCACCATCTAATTCAA CAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTCATACTA GTAACTCCTTAGCTGTTTCGGATGCCGGACAATTAAGACTAGGTACTAAT AGTCCTAGGCAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTG CCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCTAGATGGTGC AAAACCTTTCGCGGTATGGCATGATAGCGCCCAACGATCCTCCACTCCGC GGCAAGGAGGTTGCCATATGCGCCGCAACAAACGCGAACGTCAAGAGCTG TTACAGCAAACCATCCAGGCGACGCCGTTTATTACCGACGAAGAATTGGC AGGTAAGTTCGGCGTCTCGATTCAGACAATCCGTCTGGATCGTCTGGAGT TGTCCATTCCGGAATTGCGTGAACGTATTAAAAACGTCGCCGAAAAGACG TTAGAAGACGAGGTGAAATCTTTATCATTGGACGAGGTAATTGGCGAAAT TATTGACCTTGAATTAGACGACCAGGCTATTTCAATTTTGGAGATTAAAC AGGAACACGTGTTCAGTCGTAACCAGATCGCCCGCGGTCATCATCTGTTC GCGCAGGCCAACAGCCTGGCTGTGGCTGTTATTGATGACGAACTTGCGCT GACCGCTTCGGCAGACATCCGCTTTACCCGTCAGGTGAAACAAGGCGAGC GCGTCGTAGCGAAAGCTAAAGTTACCGCGGTTGAGAAGGAAAAAGGTCGT ACGGTTGTGGAAGTCAACTCATACGTCGGCGAAGAAATCGTGTTTTCAGG ACGTTTTGATATGTACCGCAGCAAACACAGCTAAGGTACCATTTTGTTTA ACTTTAAGAAGGAGATATACCATGAGCAACCACCTGTTTGACGCTATGCG TGCGGCGGCTCCGGGTGATGCCCCGTTTATCCGTATCGACAATGCTCGTA CCTGGACCTACGATGACGCAATTGCGCTGAGCGGTCGTATTGCCGGTGCA ATGGATGCACTGGGCATTCGTCCGGGTGACCGCGTTGCTGTCCAGGTGGA AAAATCTGCAGAAGCTCTGATCCTGTATCTGGCGTGCCTGCGTACCGGTG CCGTGTATCTGCCGCTGAACACCGCGTACACGCTGGCCGAACTGGATTAT TTTATTGGCGACGCAGAACCGCGTCTGGTGGTTGTTGCTCCGGCCGCACG CGGCGGTGTCGAAACGATTGCGAAACGTCATGGTGCCATCGTGGAAACCC TGGATGCGGACGGCCGTGGTAGTCTGCTGGATCTGGCACGCGACGAACCG GCTGATTTCGTGGACGCAAGCCGCTCTGCTGATGACCTGGCAGCTATTCT GTACACGAGCGGCACCACGGGTCGTTCTAAAGGCGCGATGCTGACCCATG GTAACCTGCTGTCCAATGCCCTGACGCTGCGTGATTATTGGCGCGTTACC GCGGATGACCGCCTGATTCACGCCCTGCCGATCTTTCATACCCACGGTCT GTTCGTTGCCACCAATGTGACCCTGCTGGCCGGTGCCTCAATGTTTCTGC TGTCGAAATTCGATGCCGACGAAGTTGTTAGTCTGATGCCGCAGGCAACG ATGCTGATGGGCGTTCCGACCTTTTACGTCCGTCTGCTGCAAAGTCCGCG CCTGGAAAAAGGTGCAGTGGCTTCCATTCGTCTGTTTATCAGTGGTTCCG CCCCGCTGCTGGCGGAAACCCATGCAGAATTCCACGCTCGTACCGGTCAC GCGATTCTGGAACGCTACGGCATGACCGAAACGAACATGAATACGTCTAA CCCGTATGAAGGTAAACGCATCGCAGGCACCGTTGGTTTTCCGCTGCCGG ATGTTACCGTCCGTGTGACCGACCCGGCAACCGGTCTGGTGCTGCCGCCG GAAGAAACCGGCATGATCGAAATCAAAGGTCCGAACGTTTTCAAAGGCTA CTGGCGCATGCCGGAAAAAACGGCGGCCGAATTCACCGCGGATGGCTTTT TCATTAGCGGCGATCTGGGTAAAATCGACCGCGAAGGCTATGTTCATATT GTCGGCCGTGGTAAAGATCTGGTTATTTCAGGCGGTTATAACATCTACCC GAAAGAAGTCGAAGGTGAAATTGATCAAATCGAAGGCGTCGTGGAATCGG

	CGGTGATTGGTGTTCCGCACCCGGATTTTGGCGAAGGTGTGACCGCGGTT GTCGTGTGTAAACCGGGCGCCGTGCTGGATGAAAAAACCATCGTTAGCGC ACTGCAGGATCGTCTGGCTCGCTATAAACAACCGAAACGCATTATCTTCG CCGATGACCTGCCGCGTAATACCATGGGCAAAGTCCAGAAAAACATCCTG CGTCAGCAATATGCGGACCTGTACACCCGTCGCTAAGCGGCCGCGTGTGC AGAGTCCCTGCGGCAGGCGACGAACACGACCGTCGTCGATTAGTACCGGT ACGGTCGGTGGTATCGAAGTCTTGATCACTGTACACTAGAGGTAAT

Appendix E

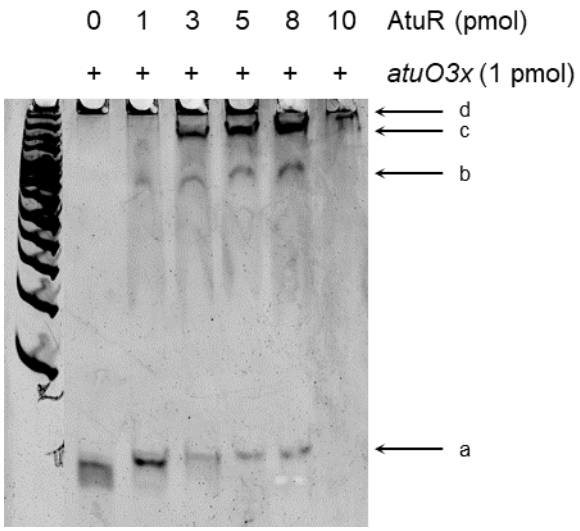


Figure E1. AtuR Electrophoretic Mobility Shift Assay. Purified AtuR_Opt (0-10 pmol) was used in an EMSA reaction with the DNA fragment *atuO3x* with 3 consectutive *atuO* sites. (**a**) is the unbound DNA, (**b**) has one AtuR dimer bound, (**c**) has two AtuR dimers bound, and (**d**) has three AtuR dimers bound.

Table E1. AtuR and pSENSE2AA DNA FASTA Sequences.

Construct	Nucleotide Sequences
AtuR_Old	ATGCTGGAGCTGGTGGCTACCGGACAGCTCACCGATCCGGAAAGCGCCCG CGGCAAGCTGCTGCAGACCGCCGCCCACCTGTTCCGCAGCAAAGGCTACG AACGCACCACGGTGCGCGACCTGGCCAGCGCGGTGGGCATCCAGTCGGGC AGCATCTTCCATCACTTCAAGAGCAAGGACGAGATCCTGCGCTCGGTGAT GGAAGAGACCATCCTCTACAACACCGCCCTGATGCGCGCCGCCCTGGCCG ACGCCGAGGACCTGCGCGAGCGGGTGCTGGGACTGATCCGCTGCGAGCTG CAATCGATCATGGGCGGCACCGGCGAGGCGATGGCGGTGCTGGTCTACGA GTGGCGCTCGCTGTCCGCCGAAGGCCAGGCGTACATCCTCGGCCTGCGCG ACATCTACGAGCAGATGTGGCTCGACGTGCTGGGGGAGGCGCGGCTGGCC GGCTACTGCCAGGGCGATCCGTTCATCCTGCGGCGCTTCCTCACCGGCGC GCTGTCCTGGACCACCACCTGGTTCCGTCCGGAAGGACCGATGAGTCTCG ACCAGCTCGCCGAGGAGGCCCTGGCGCTGGTGATCAAGAACGCCTGA
AtuR_Opt	ATGGACGATCAGAAAGCCCGGGAAGTGATGCTGGAGCTGGTGGCTACCGG ACAGCTCACCGATCCGGAAAGCGCCCGCGGCAAGCTGCTGCAGACCGCCG CCCACCTGTTCCGCAGCAAAGGCTACGAACGCACCACGGTGCGCGACCTG GCCAGCGCGGTGGGCATCCAGTCGGGCAGCATCTTCCATCACTTCAAGAG CAAGGACGAGATCCTGCGCTCGGTGATGGAAGAGACCATCCTCTACAACA CCGCCCTGATGCGCGCCGCCCTGGCCGACGCCGAGGACCTGCGCGAGCGG

	GTGCTGGGACTGATCCGCTGCGAGCTGCAATCGATCATGGGCGGCACCGG CGAGGCGATGGCGGTGCTGGTCTACGAGTGGCGCTCGCTGTCCGCCGAAG GCCAGGCGTACATCCTCGGCCTGCGCGACATCTACGAGCAGATGTGGCTC GACGTGCTGGGGGAGGCGCGGCTGGCCGGCTACTGCCAGGGCGATCCGTT CATCCTGCGGCGCTTCCTCACCGGCGCGCTGTCCTGGACCACCACCTGGT TCCGTCCGGAAGGACCGATGAGTCTCGACCAGCTCGCCGAGGAGGCCCTG GCGCTGGTGATCAAGAACGCCTGA
pSENSE2AA	TCTAGTGTACAGTGATCAAGACTTCGATACCACCGACCGTACCG GTACTA ATCGACGACGGTCGTGTTCGTCGCCTGCCGCAGGGACTCTGCAC ACCTTA AGCTATTTGTACAACTCATCCATCCCTCCAGTTGAGTGGCGTCC TTCGGC CCGCTCATACTGTTCTACAATTGTATAGTCTTCGTTATGGGACG TGATGT CCAACTTAATATTAACATTATAGGCACCGGGAAGCTGCACGGGT TTTTTA GCCTTGTAAGTAGTTTTGACCTCAGCATCGTAGTGTCCACCGTC TTTTAA TTTCAGACGCTGTTTGATCTCCCCCTTTAACGCCCCATCTTCTG GGTACA TACGTTCGCTCGATGCTTCCCATCCCATAGTCTTTTTCTGCATG ACAGGG CCGTCTGACGGGAAATTGGTCCCCCGAAGTTTCACTTTATATAT GAACTC CCCATCCTGAAGGGAGGAGTCCTGGGTCACAGTGACCACTCCGC CATCTT CGAAGTTCATCACGCGCTCCCACTTGAACCCCTCGGGGAACGAT AATTTA AGGTAGTCTGGTATGTCGGCTGGATGCTTGACGTACGCTTTACT CCCATA CATAAACTGCGGAGAAAGAATGTCCCACGCAAAAGGAAGCGGAC CACCTT TAGTAACTTTTAACTTTGCCGTCTGAGTCCCCTCGTACGGGCGA CCCTCG CCCTCACCCTCGATTTCAAACTCGTGGCCGTTAACGCTACCCTC CATGTG TACCTTGAATCTCATAAACTCCTTGATGATCGCCATATTGTCCT CTTCGC CCTTGCTGACCATGAATTCTTAGCCTTATTTTGGATACTATGGG TACCTC AGGCGTTCTTGATCACCAGCGCCAGGGCCTCCTCGGCGAGCTGG TCGAGA CTCATCGGTCCTTCCGGACGGAACCAGGTGGTGGTCCAGGACAG CGCGCC GGTGAGGAAGCGCCGCAGGATGAACGGATCGCCCTGGCAGTAGC CGGCCA

	GCCGCGCCTCCCCCAGCACGTCGAGCCACATCTGCTCGTAGATG TCGCGC AGGCCGAGGATGTACGCCTGGCCTTCGGCGGACAGCGAGCGCCA CTCGTA GACCAGCACCGCCATCGCCTCGCCGGTGCCGCCCATGATCGATT GCAGCT CGCAGCGGATCAGTCCCAGCACCCGCTCGCGCAGGTCCTCGGCG TCGGCC AGGGCGGCGCGCATCAGGGCGGTGTTGTAGAGGATGGTCTCTTC CATCAC CGAGCGCAGGATCTCGTCCTTGCTCTTGAAGTGATGGAAGATGC TGCCCG ACTGGATGCCCACCGCGCTGGCCAGGTCGCGCACCGTGGTGCGT TCGTAG CCTTTGCTGCGGAACAGGTGGGCGGCGGTCTGCAGCAGCTTGCC GCGGGC GCTTTCCGGATCGGTGAGCTGTCCGGTAGCCACCAGCTCCAGCA TCACTT CCCGGGCTTTCTGATCGTCCATATGACGTCCTCCTTGCTCTGTG GAGCTA GGCGCTATCATGCCATACCGCGAAAGGTTTTGCACCATCTAGAG CGCAAC GCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTT ACACTT TATGCTTCCGGCTCGTATGTTGTGTGGCCAAGCGCTTGCTCCAA GCGCTT GCTCCAAGCGCTTGCTAAGGAAAGGGGGTACTAGTATGAGCAAA GGAGAA GAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGG TGATGT TAATGGGCACAAATTTTCTGTCCGTGGAGAGGGTGAAGGTGATG CTACAA ACGGAAAACTCACCCTTAAATTTATTTGCACTACTGGAAAACTA CCTGTT CCGTGGCCAACACTTGTCACTACTCTGACCTATGGTGTTCAATG CTTTTC CCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGTG CCATGC CCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGAC GGGACC TACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGT TAATCG TATCGAGTTAAAGGGTATTGATTTTAAAGAAGATGGAAACATTC TTGGAC ACAAACTCGAGTACAACTTTAACTCACACAATGTATACATCACG GCAGAC

	AAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAA CGTTGA AGATGGTTCCGTTCAACTAGCAGACCATTATCAACAAAATACTC CAATTG GCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACA CAATCT GTCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCT TCTTGA GTTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCT ACAAAA AGCTTGGGCCCGTTTAAACGGTCTCCAGCTTGGCTGTTTTGGCG GATGAG AGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGC GGTCTG ATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACC TGACCC CATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTG TGGGGT CTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACG AAAGGC TCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGG TGAACG CTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTT GCGAAG CAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGC CAGGCA TCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGT TTCTAC AAACGCATGCTCTTTTTGTTTATTTTTCTAAATACATTCAAATA TGTATC CGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGA AAAAGG AAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTT TTTTGC GGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGA AAGTAA AAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAA CTGGAT CTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACG TTTTCC AATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTAT CCCGTG TTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCT CAGAAT GACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGA TGGCAT

	GACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATA ACACTG CGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTA ACCGCT TTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTG GGAACC GGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGA TGCCTG TAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTA CTTACT CTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAA AGTTGC AGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTG CTGATA AATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCA CTGGGG CCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGG GAGTCA GGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTG CCTCAC TGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATCCA TGGATA CTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTA GGTGAA GATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGT TTTCGT TCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCT TGAGAT CCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACC ACCGCT ACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTT TTCCGA AGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTT CTAGTG TAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCC TACATA CCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCG ATAAGT CGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAG GCGCAG CGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGA GCGAAC GACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAA GCGCCA CGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGC AGGGTC

	GGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTG GTATCT TTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGAT TTTTGT GATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAAC GCGGCC TTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTT CTTTCC TGCGTTATCCCCTGATTCTGTGGATAACCGTATTACC