Copyright

by

Amanda Jane Hughes

2013

## The Dissertation Committee for Amanda Jane Hughes certifies that this is the approved version of the following dissertation:

# Towards Preparative *in vitro* Enzymatic Synthesis of New Polyketide Metabolites

Committee:				
Adrian Keatinge-Clay, Supervisor				
Dionicio Siegel				
Karen Browning				
Kalch Blowning				
William				
Walter Fast				
Ben Liu				

# Towards Preparative *in vitro* Enzymatic Synthesis of New Polyketide Metabolites

by

### Amanda Jane Hughes, B.S. Chem.

### **Dissertation**

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

### **Doctor of Philosophy**

The University of Texas at Austin
August 2013



Towards Preparative in vitro Enzymatic Synthesis of New Polyketide

**Metabolites** 

Amanda Jane Hughes, Ph. D

The University of Texas at Austin, 2013

Supervisor: Adrian Keatinge-Clay

Modular polyketide synthases (PKSs) are the largest enzymes known to man and

are responsible for synthesizing some of the most important human medicines. Their

ability to construct stereochemically-rich carbon chains containing diverse substituents

has inspired the biosynthetic community to engineer these factories for the in vitro

synthesis of a small library of polyketide compounds. New complex polyketides are

discovered every year, yet the lack of compound prohibits characterization and testing of

these new compounds for medicinal properties. Smaller polyketide compounds generated

in vitro could be organically manipulated to generate larger, more complex polyketide

natural products and natural product analogs. Chemoenzymatic approaches like this

would be extremely beneficial to the scientific community; however, there are still

obstacles that must be overcome before the use of PKS for the preparative synthesis of an

in vitro generated polyketide library would prove fruitful: purchasing substrates such as

methylmalonyl-CoA is cost-prohibitive, PKSs are often difficult to express and purify,

and the products generated are typically nonchromophoric. The use of a malonyl-CoA

ligase from Streptomyces coelicolor (MatB) was investigated for the enzymatic synthesis

of polyketide extender units such as methylmalonyl-CoA (Chapter 2). MatB synthesized

a total of 5 CoA-linked extender units in vitro: malonyl-, methylmalonyl-, ethylmalonyl-,

hydroxymalonyl- and methoxymalonyl-CoA. Two ternary complex structures of MatB with bound product and leaving group were also solved to sub-2Å resolution. MatB generated extender units were employed in the module-catalyzed synthesis of a triketide pyrone. The selectivity of a PKS module to incorporate a variety of side chains into triketide pyrones was also investigated (Chapter 3). A total of 10 triketide pyrone compounds were synthesized, 5 produced via modular "stuttering" and one possessing a terminal alkyne chemical handle. Lastly, nonchromphoric polyketide products were made visible upon copper(I)-catalyzed azide alkyne cycloaddition (CuAAC) with fluorescent sulforhodamine B azide revealing insights into *in vitro* reactivites of a PKS module (Chapter 4). The work described in this dissertation has helped advance the scientific community towards procuring an *in vitro* synthesized polyketide library for future synthetic applications.

## **Table of Contents**

List of Figuresxi
List of Schemesxiii
List of Tablesxiv
Chapter 1 Introduction
Chapter 2 Enzymatic Extender Unit Generation for <i>in vitro</i> Polyketide Synthase Reactions: Structural and Function Showcasing of <i>Streptomyces coelicolor</i> MatB
Abstract
Introduction
Results
MatB synthesizes diverse CoA-, PANT- and NAC-linked extender units
Incorporation of MatB-synthesized mmCoA, mmPANT and mm-S-NAG into a triketide pyrone
Overall MatB structure
MatB observed in the thioester-forming conformation20
Comparison to adenylate-forming enzyme core motifs20
Bound malonyl-CoA and methylmalonyl-CoA22
Bound adenosine monophosphate23
Discussion24
Materials and Methods
Synthesis of compounds utilized
N-acetylcysteamine (NAC)28
Methoxymalonic acid
D-pantetheine
Cloning
Protein expression and purification

	Protein crystallization	31
	Data collection, processing and refinement	31
	HPLC analysis of MatB reactions	32
	HPLC analysis of triketide pyrone reactions	32
	Mass spectrometry analysis	33
	LC/MS analysis	33
	Accession numbers	33
Chap	oter 3 Employing a Polyketide Synthase Module and Thioesterase in the Semipreparative Biocatalysis of Diverse Triketide Pyrones	43
	Abstract	
	Introduction	44
	Results	44
	Triketide pyrones from diketide substrates	44
	Triketide pyrones produced by "stuttering"	
	TE-catalyzed substrate hydrolysis	47
	Discussion	48
	Materials and Methods	49
	Synthesis of compounds utilized	49
	N-acetylcysteamine (NAC)	49
	β-ketoacyl-S-NAC diketides	49
	4-pentynoyl-S-NAC	50
	acyl-S-NAC monoketides	50
	Cloning	51
	Protein expression and purification	51
	Methylmalonyl-S-NAC MatB reaction	52
	Enzymatic synthesis of triketide pyrones by EryMod6TE	52
	Scaled-up synthesis of triketide pyrone 3.11	53
	Enzymatic synthesis of triketide pyrones via "stuttering"	54
	TE-catalyzed hydrolysis reactions	55
	Activity assay of the five EryMod6TE variants	. 56

	LC/MS analysis	56
Chap	oter 4 Investigating the Reactivities of a Polyketide Synthase Module through Fluorescent Click Chemistry	
	Abstract	66
	Introduction	66
	Results	68
	Bioorthogonal detection of polyketide products in cell lysate	68
	An ethanethiol-linked extender unit	68
	Ethanethiol-linked priming units	69
	Reactivity of EryMod6TE towards a β-ketoacyl ethanethioester	70
	Phosphate increases substrate turnover	71
	TE-catalyzed glycerolysis	71
	Discussion	72
	Materials and Methods	74
	Synthesis of compounds utilized	74
	Methylmalonyl-N-acetylcysteamine thioester	74
	Methylmalonyl-ethanethiol thioester	75
	NAC- and ethanethiol-linked monoketides	75
	β-ketoacyl-S-ethanethiol diketide	77
	Sulforhodamine B azide	77
	Engineering EryMod6TE mutant plasmids	78
	ΔKS, ΔAT, ΔTE Plasmids	78
	ΔKS+AT Plasmid	79
	ΔKS+TE Plasmid	79
	ΔAT+TE Plasmid	79
	Protein expression, dialysis, and purification	79
	EryMod6TE reaction conditions	80
	Preparative in vitro synthesis of reduced diketide 4.5	81
	Phosphate-containing reaction conditions	81
	Glycerol-containing reaction condition	81

CuAAC reaction conditions	82
HPLC and fluorescence detection	82
LC/MS analysis	82
Chapter Five: Conclusions	90
Appendix 1. Supporting Information for Chapter 3	93
HPLC data for TE-mediated hydrolysis of polyketide substrates	93
TE-mediated hydrolysis of 3.4	93
TE-mediated hydrolysis of 3.15	95
TE-mediated hydrolysis of 3.16	96
HPLC peak areas for activity assay of EryMod6TE variants	97
<sup>1</sup> H NMR characterization of triketide pyrone compounds	99
Compound 3.8	99
Compound 3.9	100
Compound 3.10	101
Compound 3.11	102
Compound 3.12	103
Compound 3.13	104
Compound 3.20	105
<sup>13</sup> C NMR characterization of triketide pyrone 3.11	106
Appendix 2. Supporting Information for Chapter 4	107
HPLC Peak areas for CuAAC reactions	107
<sup>1</sup> H NMR characterization of terminal alkyne-containing compound	ls131
Compound 4.5	131
Compound 4.21	132
Bibliography	133

## List of Figures

<b>Figure 1-1</b> Target polyketide compounds for chemoenzymatic synthesis7
<b>Figure 1-2</b> The erythromycin polyketide synthase (PKS)8
<b>Figure 1-3</b> Polyketide diversity through extender unit incorporation9
Figure 1-4 The six ketoreductase (KR) types
Figure 1-5 NADPH regeneration from <i>Bacillus subtilis</i> glucose dehydrogenase.11
<b>Figure 1-6</b> Triketide pyrone formation from EryMod6TE12
<b>Figure 1-7</b> Polyketide detection method using fluorescent click chemistry13
Figure 2-1 HPLC analysis of MatB reactions
<b>Figure 2-2</b> HPLC analysis of the formation of triketide pyrones
Figure 2-3 Ternary complex MatB structure reveals bound product
Figure 2-4 Ternary complex MatB structure reveals AMP leaving group bound.41
Figure 2-5 Engineering MatB substrate specificity and product stereochemistry.42
<b>Figure 3-1</b> Enzymatic synthesis of triketide pyrones from diketides
<b>Figure 3-2</b> Enzymatic synthesis of triketide pyrones from monoketides61
Figure 3-3 TE-mediated hydrolysis of priming units 3.4, 3.7, 3.15 and 3.16 62
Figure 3-4 Polyketides and TE cyclization
Figure 3-5 Domain swapping of EryTE leads to the formation of triketide pyrones
64
Figure 4-1 Lighting up polyketide products through fluorescent click chemistry.84
Figure 4-2 A comparison of <i>in vitro</i> extender units85
<b>Figure 4-3</b> EryKS6 promiscuity towards thiol handles and chain lengths 86
Figure 4-4 EryMod6TE reactivity towards an advanced priming unit
Figure 4-5 Phosphate enhances turnover

Figuro 1	6 Em/TE ma	diated alveer	alveie of	nrimina	unite	80
r igure 4-c	o Ery i E-ille	dialed grycer	orysis or	priming	units	09

## **List of Schemes**

Scheme 2-1	MatB reaction mechanism.	34
Scheme 2-2	Enzymatic synthesis of triketide pyrone by EryMod6TE	36
Scheme 3-1	Semipreparative synthesis of triketide pyrones from diketides and	
	methylmalonyl-S-NAC	57
Scheme 3-2	Semipreparative synthesis of triketide pyrones from monoketides ar	ıd
	methylmalonyl-S-NAC via stuttering	60
Scheme 3-3	Engineering EryMod6TE plasmid constructs for domain swapping	
	experiments	65

## **List of Tables**

<b>Table 2-1</b> Data collection and refinement of <i>S. coelicolor</i> MatB structures	38
Table 2-2 Core motifs conserved in adenylate-forming enzymes.	40
Table 3-1 <sup>1</sup> H NMR and LC/MS characterization of triketide pyrones	59

### **Chapter 1 Introduction**

Due to the ability of polyketide synthases (PKSs) to incorporate a variety of functional groups and stereochemistries into a growing polyketide chain (Khosla, 2007; Keatinge-Clay, 2013), a longstanding vision of the PKS community has been to harness this enzymatic power and produce libraries of stereochemically diverse molecules *in vitro*. Every year, new complex natural products are discovered, yet a lack of synthetic compound hinders testing for potential medicinal efficacy. A library of stereoisomeric polyketides could serve as chiral synthons to generate stereochemically complex natural products and natural product analogs such as bryostatin, concanomycin, soraphen and discodermolide (Figure 1-1). A PKS-generated library of complex molecules would prove extremely beneficial to the scientific community; however, there are still many obstacles that must be overcome before PKSs may be employed in the preparative synthesis of diverse molecules *in vitro*.

Modular type I PKSs are enzymatic assembly lines responsible for the condensation and processing of small carbon building blocks, termed extender units. The erythromycin synthase is the most thoroughly studied of all PKSs and may be used as a model system when describing the enzymatic synthesis of polyketides (Figure 1-2). A loading didomain commences polyketide growth by priming the enzymatic assembly line with a propionyl group. Within the following modules, an acyltransferase (AT) selects an extender unit, a ketosynthase (KS) catalyzes claisen-like condensation reactions, and optional  $\beta$ -carbon processing enzymes, such as a ketoreductase (KR), a dehydratase (DH), and an enoylreductase (ER), set the functional groups and stereochemistries at the  $\alpha$ - and  $\beta$ -carbons. The growing polyketide chain is shuttled through the PKS in a handover-hand fashion via an  $\sim$ 18Å phosphopantetheinyl arm attached to the acyl carrier

protein (ACP) of each module. The final polyketide is both hydrolyzed and cyclized from the PKS by the thioesterase (TE). After cyclization, tailoring enzymes such as P450 oxygenases and glycosyl transferase decorate the polyketide with hydroxyl groups and sugar moieties to produce the final, biologically-active polyketide.

The complexity of polyketide natural products largely results from (1) the number of modules found within the PKS system, (2) extender unit selection by ATs and (3) KR activity. The most common extender units selected for by ATs are malonyl-CoA, (2S)-methylmalonyl-CoA and (2S)-ethylmalonyl-CoA. More exotic extender units incorporated into a growing polyketide chain include (2R)-methoxymalonyl-ACP, (2R)-hydroxymalonyl-ACP and (2S)-aminomalonyl-ACP (Figure 1-3). These exotic extender units are synthesized on the ACP arm by tailoring enzymes found within the PKS gene cluster.

The KR type within each module also has a great influence on polyketide complexity. There are six KR types found within modular PKSs: A1, A2, B1, B2, C1 and C2 (Figure 1-4). A-type KRs reduce the  $\beta$ -carbonyl to a L-hydroxy group, and B-type KRs reduce the  $\beta$ -carbonyl to a D-hydroxy group. C-type KRs are reductase incompetent, for they lack the essential active site tyrosine residue. Numerical values associated with the KR type is indicative of  $\alpha$ -carbon epimerase functionality; type 1 KRs are epimerase inactive and leave  $\alpha$ -substituents in D-configuration whereas type 2 KRs epimerize  $\alpha$ -substituents to L-configuration. The KRs of modules 1 and 4 in the erythromycin PKS (EryKR1 and EryKR4, respectively) are B2 type KRs. EryKR2, EryKR5 and EryKR6 are A1 type KRs, while EryKR3 is a C2 type KR (Figure 1-2).

In addition to the AT extender unit selection and KR type, their modular nature makes PKSs desirable for the *in vitro* generation of complex molecules. Domain swapping experiments to generate hybrid modules have been successful in their ability to

produce polyketide compounds possessing new functionalities. For example, Oliynyk and coworkers demonstrated the formation of a new triketide lactone compound by swapping the native (2S)-methylmalonyl-CoA selecting AT for a malony-CoA selecting AT (Oliynyk, 1996). Kao and coworkers employed this technique to produce two triketide lactones with new stereochemical configurations after swapping the native erythromycin PKS A1-type KR for two different non-native B-type KRs from the rapamycin PKS (Kao, 1998). In both experiments, the new functionality of the swapped domain was incorporated into the polyketide product; however, an overall loss of activity was reported for these hybrid enzymes (Oliynyk, 1996; Kao, 1998). This loss was hypothesized to result from inter-modular structural interactions. A second, more recent approach to generate polyketides possessing new functionalities has been through directed evolution. This approach does not alter inter-modular structural interactions, but focuses on altering substrate specificity of the native domain within that module. Sundermann and coworkers demonstrated the power of this technique by successfully evolving EryAT6 to accept a terminal alkyne extender unit (Sundermann, 2012).

The extraordinary promise of modular PKSs to produce a small library of compounds *in vitro* is apparent; however, there are many challenges associated with increasing *in vitro* polyketide production to a preparative scale: substrates such as NADPH and methylmalonyl-CoA are costly, modules are often difficult to express and purify, and the products generated are typically nonchromophoric. NADPH is an essential cofactor required for KR activity. Fortunately, a glucose-fueled NADPH-regeneration system has already been developed to provide a constant supply of reduced nicotinamide coenzyme (Figure 1-5) (Wong, 1985). In order to reduce financial burden of purchased extender units for *in vitro* production of polyketides, the use of a malonyl-CoA ligase from *Streptomyces coelicolor* (MatB) was investigated (Chapter 2).

MatB is an adenylate-forming enzyme that synthesizes malonyl-CoA from ATP, Mg<sup>2+</sup>, malonic acid and CoA. Experiments showed MatB possessed a high substrate tolerance toward the diacid incorporated; a total of 5 CoA-linked extender units were synthesized: malonyl-, methylmalonyl-, ethylmalonyl-, hydroxymalonylmethoxymalonyl-CoA. Promiscuity towards the thiol handle was also investigated; resulting in the first report of enzymatically synthesized D-pantetheine (PANT-) and Nacetylcysteamine (NAC-) linked extender units. The terminal module – thioesterase complex of the erythromycin PKS (EryMod6TE) was employed for the chemoenzymatic synthesis of a triketide pyrone from a diketide-S-NAC priming substrate and MatB synthesized CoA-, PANT- and NAC-linked methylmalonyl extender units. In addition to enzymatic assays, two ternary-complex structures of MatB were solved using single wavelength anomalous dispersion - molecular replacement (SAD-MR), revealing density in the active site for the bound product and leaving group (malonyl-CoA or (2R)methylmalonyl-CoA and AMP, accession codes 3NYQ and 3NYR, respectively).

With the ultimate goal of expanding the diversity of an *in vitro* synthesized polyketide library, and having developed an established route to triketide compounds from MatB synthesized extender units, the substrate tolerance of EryKS6 from EryMod6TE was investigated (Chapter 3). EryMod6TE was incubated with MatB-generated methylmalonyl-*S*-NAC and a variety of β–ketoacyl-*S*-NAC diketide and acyl-*S*-NAC monoketide priming substrates (Figure 1-6). EryKS6 demonstrated a high substrate tolerance; a total of 10 triketide pyrone compounds were synthesized, 5 produced via modular "stuttering" and one possessing a terminal alkyne chemical handle. "Stuttering" refers to the phenomenon of a single module catalyzing more than one condensation reaction. This was the first reported observation of EryMod6 "stuttering." Pyrone product (20 mg) was isolated from a reaction where 200 mg of diketide-*S*-NAC

EryMod6TE lysate resulting in an overall 12% yield. It was hypothesized that TE-catalyzed hydrolysis of priming substrate was the major source for the low percent yield. Hoping that a non-native TE would be less active towards substrate hydrolysis, pyrone formation was examined for three EryMod6TE hybrid modules where the native EryTE was swapped for non-native TEs, two PKS TEs and one non-ribosomal peptide synthase (NRPS) TE. A decrease in pyrone production was reported for the domain swapping experiments, and the source(s) of the low polyketide product yields remained unidentified. Many polyketide products and intermediates do not contain chromophoric moieties, making it very difficult to pinpoint where loss of product occurs during enzymatic synthesis. With only overall 4-12% yields for triketide product formation with our established EryMod6TE system (Harper, 2012; Hughes, 2012), a method utilizing copper catalyzed azide-alkyne cycloaddition (CuAAC, click chemistry) was developed to detect module-catalyzed polyketide formation (Chapter 4).

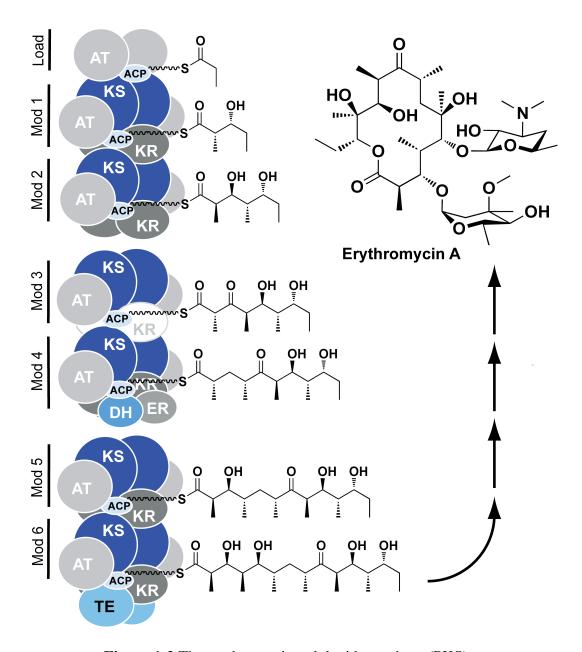
Prasad and coworkers employed sulforhodamine B azide modified ACPs for the detection of KS specificity on a small library of monoketide-S-NAC substrates by CuAAC (Prasad, 2012). Inspired by this work, experiments were designed to monitor polyketide formation using terminal alkyne priming substrates, dialyzed EryMod6TE lysate and synthetic extender units. Polyketides were visualized upon CuAAC with fluorescent sulforhodamine B azide by HPLC (Figure 1-7). The utility of this method was demonstrated through experiments designed to investigate the reactivity of EryMod6TE to various substrates and reaction conditions. Results demonstrated that ethanethiol was found to be as effective as N-acetylcysteamine in delivering both priming units and extender units to PKS domains. Insights into the competing reactivities within EryMod6TE were revealed through triketide lactone products generated from a β-

ketoacyl ethanethioester substrate. The addition of phosphate increased enzymatic turnover, while the addition of glycerol resulted in thioesterase-mediated glycerolysis of priming units.

Overall, our efforts focused on improving module-catalyzed polyketide yields and reducing costs for the preparative *in vitro* chemoenzymatic synthesis of polyketide compounds. Employing a promiscuous malonyl-CoA ligase enzyme allowed for the generation of diverse polyketide extender units by *S. coelicolor* MatB for the price of ATP. We demonstrated a high substrate tolerance of polyketide priming units by a PKS module, which lead to the module-catalyzed formation of 10 polyketide compounds. Lastly, a new method to follow terminal alkyne priming substrates through PKS module-catalyzed reactions was developed in order to optimize reaction conditions and identify polyketide products that went previously undetected. The work described in the following chapters has helped scientific community advance towards procuring a PKS-generated library of stereochemically diverse molecules *in vitro*.

Figure 1-1 Target polyketide compounds for chemoenzymatic synthesis.

Target complex natural products for preparative-scale chemoenzymatic synthesis include a) bryostatin, b) concanomycin, c) discodermolide and d) soraphen.



**Figure 1-2** The erythromycin polyketide synthase (PKS).

The erythromycin PKS is the most thoroughly studied of all PKSs. It is compromised of a loading didomain, which begins polyketide growth from a propionyl-CoA priming unit. The following six modules incorporate and process a (2S)-methylmalonyl-CoA extender unit. The polyketide is released from the PKS by the thioesterase (TE). It is then enzymatically decorated by tailoring enzymes to afford the biologically-active polyketide, Erythromycin A. KS – ketosynthase, AT – acyltransferase, KR – ketoreductase, DH – dehydratase, ER – enoylreductase, TE – thioesterase

**Figure 1-3** Polyketide diversity through extender unit incorporation.

The most common CoA-linked polyketide extender units are malonyl-CoA, (2S)-methylmalonyl-CoA, and (2S)-ethylmalonyl-CoA; the most common ACP-linked polyketide extender units are (2R)-methoxymalonyl-ACP, (2R)-hydroxymalonyl-ACP, and (2S)-aminomalonyl-ACP. These extender units are incorporated into polyketides a) 6-deoxyerythronolide B, b) amphotericin B aglycone, c) concanamycin A aglycone and d) a polyketide-nonribosomal peptide hybrid zwittermicin A.

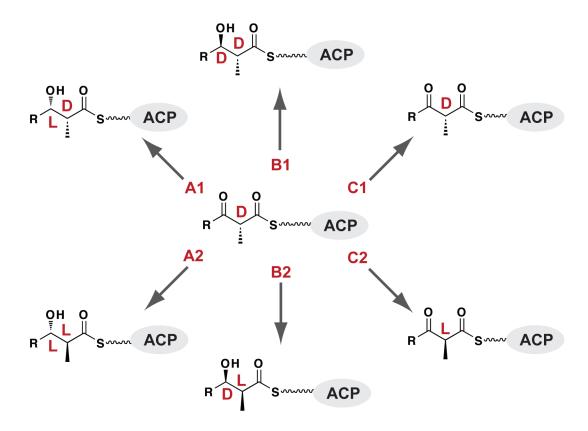
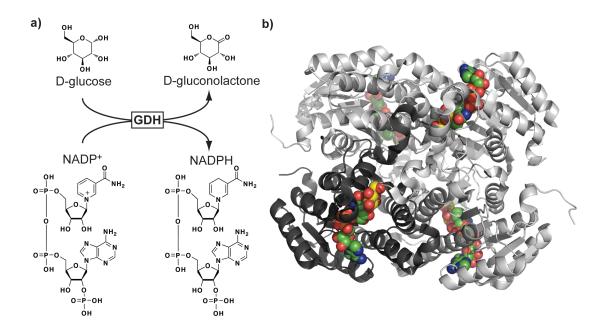


Figure 1-4 The six ketoreductase (KR) types.

There are six KR types found within modular PKSs: A1, A2, B1, B2, C1 and C2. A-type KRs reduce the  $\beta$ -carbonyl to a L-hydroxy group, and B-type KRs reduce the  $\beta$ -carbonyl to a D-hydroxy group. C-type KRs lack the active site tyrosine residue and are reductase incompetent. Type 1 KRs are epimerase inactive and leave  $\alpha$ -substituents in D-configuration whereas type 2 KRs epimerize  $\alpha$ -substituents to L-configuration.



**Figure 1-5** NADPH regeneration from *Bacillus subtilis* glucose dehydrogenase.

a) Schematic representation of the NADPH-regeneration system in which the reduction of NADP<sup>+</sup> to NADPH is catalyzed by the oxidation of glucose to gluconolactone by *Bacillus subtilis* glucose dehydrogenase (GDH). b) Pymol image of GDH tetramer depicts GDH monomer (black cartoon) with bound glucose (yellow spheres) and NADP<sup>+</sup> (green spheres) (PDB ID: 3ay6).

**Figure 1-6** Triketide pyrone formation from EryMod6TE.

a) Triketide pyrone compounds were enzymatically synthesized by EryMod6TE from MatB-generated methylmalonyl-S-NAC extender units and synthetic β-ketoacyl-S-NAC diketide priming units. b) Triketide pyrone compounds produced by modular "stuttering." EryMod6TE was incubated with MatB-generated methylmalonyl-S-NAC extender units and synthetic acyl-S-NAC monoketide priming units. "Stuttering" refers to the phenomenon of a single module catalyzing more than one condensation reaction.

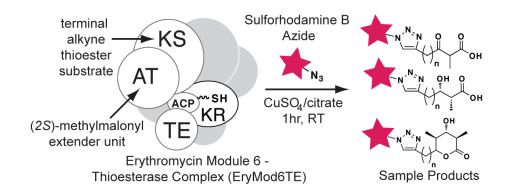


Figure 1-7 Polyketide detection method using fluorescent click chemistry.

Terminal alkyne thioester substrates were incubated with EryMod6TE, synthetic methylmalonyl extender units and the NADPH regeneration system. After 2 or 16 hours, polyketide products were visualized upon copper(I)-catalyzed azide alkyne cycloaddition (CuAAC) with sulforhodamine B azide and HPLC analysis.

# Chapter 2 Enzymatic Extender Unit Generation for *in vitro* Polyketide Synthase Reactions: Structural and Function Showcasing of Streptomyces coelicolor MatB

#### ABSTRACT

In vitro experiments with modular polyketide synthases (PKSs) are often limited by the availability of polyketide extender units. To determine the polyketide extender units that can be biocatalytically accessed via promiscuous malonyl-CoA ligases, structural and functional studies were conducted on Streptomyces coelicolor MatB. We demonstrate that this adenylate-forming enzyme is capable of producing most CoAlinked polyketide extender units as well as pantetheine- and N-acetylcysteamine-linked analogs useful for in vitro PKS studies. Two ternary product complex structures, one containing malonyl-CoA and AMP and the other containing (2R)-methylmalonyl-CoA and AMP, were solved to 1.45 Å and 1.43 Å resolution, respectively. MatB crystallized in the thioester-forming conformation, making extensive interactions with the bound extender unit products. This first structural characterization of an adenylate-forming enzyme that activates diacids reveals the molecular details for how malonate and its derivatives are accepted. The orientation of the  $\alpha$ -methyl group of bound (2R)methylmalonyl-CoA, indicates that it is necessary to epimerize α-substituted extender units formed by MatB before they can be accepted by PKS acyltransferase domains. We demonstrate the *in vitro* incorporation of methylmalonyl groups ligated by MatB to CoA, pantetheine, or N-acetylcysteamine into a triketide pyrone by the terminal module of the 6-deoxyerythronolide B synthase. Additionally, a means for quantitatively monitoring certain *in vitro* PKS reactions using MatB is presented.

### Introduction

To study the activities of PKS modules in vitro, a supply of polyketide extender units is necessary; however, the cost and availability of these molecules is prohibitive. If the extender units are obtained from a commercial source, experiments designed to produce over a milligram of polyketide product in vitro become impractical. An alternative is to use N-acetylcysteamine (NAC-)-linked extender units, which have been employed in feeding studies and *in vitro* reactions (Carroll, 2002; Pohl, 1998). Pohl et al. showed that on incubation of synthetic (2RS)-methylmalonyl-S-NAC with DEBS3, the third protein of the 6-deoxyerythronolide B synthase, in the presence of (2S, 3R)-3hydroxy-2-methylpentanoate-S-NAC and NADPH, the anticipated triketide lactone was generated (Pohl, 1998); this result indicates that the terminal DEBS3 AT accepts methylmalonyl units from (2S)-methylmalonyl-S-NAC. To access CoA-linked extender units enzymatically, Pohl and coworkers utilized the malonyl-CoA synthetase MatB from Rhizobium trifolii, which has a broad substrate tolerance, and demonstrated that it can produce malonyl-CoA, methylmalonyl-CoA, ethylmalonyl-CoA, as well as unnatural units such cyclopropylmalonyl-CoA, cyclobutylmalonyl-CoA, extender as benzylmalonyl-CoA, and cyclopropylmethylenemalonyl-CoA (Pohl, 2001). We have employed the previously uncharacterized Streptomyces coelicolor MatB to synthesize CoA-linked extender units and derivatives thereof.

The *S. coelicolor* malonyl-CoA synthetase MatB is an adenylate-forming enzyme that is best classified with acyl-CoA synthetases. MatB is anticipated to be enzymatically and structurally homologous to such enzymes as acyl-CoA synthetases (Protein Data Bank [PDB] codes: 3EQ6, 1V26, 1PG3, 1RY2) (Kochan, 2009; Hisanaga, 2004; Gulick, 2003; Jogl, 2004), 4-chlorobenzoyl-CoA ligase (PDB code: 1T5D) (Gulick, 2004), the adenylation domains (A-domains) of nonribosomal peptide synthetases (NRPSs; PDB

codes: 3ITE, 3E7W, 1AMU) (Lee, 2010; Yonus, 2008; Conti, 1997; Strieker, 2010; Mootz, 2002; Marahiel, 2009), and firefly luciferase (PDB code: 2D1R) (Nakatsu, 2006). A structure of a diacid-activating adenylate-forming enzyme has not yet been reported. Reactions catalyzed by adenylate-forming enzymes can be represented by a medium-chain acyl-CoA synthetase: in the adenylate-forming conformation a medium-sized fatty acid is selected to attack the α-phosphoryl moiety of adenosine triphosphate (ATP) via its carboxylate to yield an acyl-adenylate and pyrophosphate; in the thioester-forming conformation the thiol acceptor CoA is selected to perform the second half-reaction, generating acyl-CoA and AMP (Scheme 2-1) (Kochan, 2009).

Here, we demonstrate the S. coelicolor MatB-catalyzed synthesis of polyketide extender units malonyl-CoA (mCoA), methylmalonyl-CoA (mmCoA), ethylmalonyl-CoA (emCoA), methoxymalonyl-CoA (oxCoA), and hydroxymalonyl-CoA (hmCoA), marking the first enzymatic syntheses of oxCoA and hmCoA. Equivalent pantetheine-(PANT-) and NAC-linked extender units were also generated, demonstrating that MatB is not only promiscuous toward malonyl derivatives, but also thiol acceptors. Crystal structures of two MatB ternary product complexes were obtained with malonyl-CoA and AMP (1.45 Å resolution) and with (2R)-methylmalonyl-CoA and AMP (1.43 Å resolution). These are the first structures of a diacid-activating adenylate-forming enzyme. The strong electron density surrounding the protein, thioester product, and AMP provides some of the most complete snapshots of product formation in the enzyme superfamily. The orientation of the (2R)-methylmalonyl-CoA  $\alpha$ -methyl group reveals the stereoselectivity of MatB toward α-substituted malonate derivatives and also reveals the pocket that enables MatB to be promiscuous toward α-substituents. *In vitro* reactions in which the methylmalonyl groups of MatB synthesized extender units mmCoA, mmPANT, and mm-S-NAC are incorporated into a triketide pyrone by the final module of the 6-deoxyerythronolide B synthase exemplify the use of MatB. A protocol for monitoring such reactions through a coupled reaction, also catalyzed by MatB, is described.

### RESULTS

### MatB synthesizes diverse CoA-, PANT- and NAC-linked extender units

Products from incubations of CoA and a diacid (malonic acid, methylmalonic acid, ethylmalonic acid, methoxymalonic acid, and tartronic acid) with MatB were collected during high-performance liquid chromatography (HPLC) runs and analyzed by mass spectrometry. The observed masses (calculated masses) from the hmCoA, oxCoA, mCoA, mmCoA, and emCoA reactions were 867.8(868.1), 881.9(882.1), 851.7(852.1), 865.9(866.1), and 879.9(880.1), respectively, indicating the production of each of the five CoA-linked extender units (Figure 2-1a). Aminomalonyl-CoA could not be produced. MatB was also found capable of synthesizing PANT- and NAC-linked extender units; employing identical HPLC methanol gradients, the retention times of the PANT-linked extender units were increased relative to the CoA-linked extender units, whereas the retention times of the NAC-linked extender units were decreased (Figures 2-1b-c). Two product peaks were observed in each of the hmCoA, hmPANT, and hm-S-NAC reactions. Both peaks from the hm-S-NAC reaction were collected and subjected to LC/MS analysis. The observed masses (calculated masses) of 222.2(222.0) and 206.2(206.1) correspond to hm-S-NAC and m-S-NAC, respectively. The m-S-NAC peak may have resulted from a malonic acid impurity in the tartronic acid stock. With the addition of sufficient ATP, thiol, and diacid, MatB reactions can be scaled up to yield multimilligram quantities that can be readily analyzed by nuclear magnetic resonance (NMR). In

summary, *S. coelicolor* MatB produces each of the CoA-, PANT-, and NAC-linked extender units mCoA, mmCoA, emCoA, oxCoA, hmCoA, mPANT, mmPANT, emPANT, oxPANT, hmPANT, m-*S*-NAC, mm-*S*-NAC, em-*S*-NAC, ox-*S*-NAC, and hm-*S*-NAC.

## Incorporation of MatB-synthesized mmCoA, mmPANT and mm-S-NAC into a triketide pyrone

The final module and thioesterase from the 6-deoxyerythronolide B synthase (EryMod6TE) has been used in previous experiments to produce triketide lactones (Pohl, 1998; Gokhale, 1999; Menzella, 2005). We were able to incorporate methylmalonate from MatB-synthesized mmCoA, mmPANT, and mm-S-NAC into a triketide pyrone by (1) incubating MatB ligation reactions overnight at 22 °C; (2) adding EryMod6TE, S. coelicolor methylmalonyl-CoA epimerase (Dayem, 2002), and (2S, 3R)-2-methyl-3oxopentanoate-S-NAC (Keatinge-Clay, 2007); and (3) incubating at least 16 hours at 22 °C and analyzing for polyketide production via HPLC (Scheme 2-2, Figures 2-2a-b). To confirm the identity of the triketide pyrone, the fraction corresponding to the largest peak (at 290 nm) was collected and subjected to LC/MS; the observed mass (calculated mass) of the triketide pyrone was 169.2(169.1). In all reactions, mm-S-NAC is formed as a byproduct generated by MatB using methylmalonate and NAC liberated from (2S, 3R)-2methyl-3-oxopentanoate-S-NAC, although thioesterase-catalyzed hydrolysis of (2S, 3R)-2-methyl-3-oxopentanoate-S-NAC may also contribute to mm-S-NAC formation (Figure 2-2a) (Wang, 2009). Monitoring mm-S-NAC production represents a powerful, new method of following in vitro PKS reactions, especially those producing polyketides lacking chromophores.

### **Overall MatB structure**

The conditions that yielded protein crystals contained all of the substrates required for the synthesis of mCoA or mmCoA as described in the Methods section. Briefly, purified protein was incubated with the substrates of the MatB reaction: ATP, magnesium, CoA, and malonate or methylmalonate. Clusters of elongated, orthorhombic crystals, often hollowed along their long axes, grew overnight. Diffraction data was collected using crystals flash-frozen in their mother liquor. Molecular replacement proved unsuccessful, as is often the case with adenylate-forming enzymes (Shah, 2009). The mCoA/AMP-bound MatB structure was solved via single-wavelength anomalous dispersion using selenomethionine-labeled protein. The native malonyl-CoA/AMP-bound and methylmalonyl-CoA/AMP-bound complexes were then refined at resolutions of 1.45 Å and 1.43 Å, respectively (Table 2-1). A comparison of the ternary complexes yields root-mean-square deviation of 0.06 Å.

MatB is comprised of 485 residues, possesses a molecular weight of 50.6 kDa, and is monomeric as determined through gel filtration chromatography and through an analysis of the packing within the crystals. Like other adenylate-forming enzymes, MatB has a large N-terminal body (residues 1–387) and a small C-terminal lid (residues 388–485), yet is the smallest of the adenylate-forming enzymes solved to date (Figure 2-3a). No density was observed for the N-terminal histidine tag, residues 1–3, or residues 472–485 of the C-terminal lid. No density for the 3',5'-diphosphoadenosine of CoA was observed either, permitting only the phosphopantetheine moiety of CoA to be modeled. The overall structure for mCoA/AMP-bound MatB contains 467 residues, 348 waters, 1 mCoA, 1 AMP molecule, and 1 solvated Mg<sup>2+</sup> in the asymmetric unit. The overall structure of mmCoA/AMP-bound MatB contains 467 residues, 407 water molecules, 1 mmCoA, 1 AMP molecule, and 1 solvated Mg<sup>2+</sup> in the asymmetric unit. The electron

density clearly reveals an R stereochemistry at the  $\alpha$ -carbon indicating that (2R)-mmCoA is bound. Other than the  $\alpha$ -methyl group, the positions of the bound mCoA and (2R)-mmCoA extender units are superimposable between the two structures.

### MatB observed in the thioester-forming conformation

Adenylate-forming enzymes can adopt at least two catalytically distinct conformations (Reger, 2008; Kochan, 2009; Lee, 2010). The C-terminal lid pivots 140° between the two catalytically-active conformations (the adenylate-forming conformation and the thioester-forming conformation) in a process termed "domain alternation" (Reger, 2008; Shah, 2009) (Figure 2-3a). An aspartate located in a linker between the Nterminal body and the C-terminal lid of adenylate-forming enzymes, is usually considered to be the principle pivot point about which the C-terminal lid rotates during domain alternation and is hypothesized to interact with the arginine that is the final residue of the N-terminal body in the thiolation conformation (Wu, 2008). The thioester-forming conformation of 4-chlorobenzoate-CoA ligase shows an ionic interaction between hinge residue D402 and the final residue of the N-terminal body (R400); the D402P and D402A mutants showed decreased catalytic activity (Reger, 2008; Wu, 2008). The putative hinge residue in MatB is D387, which is four residues after the final residue of the N-terminal body (R383) and highly conserved among malonyl-CoA synthetases. The side-chain electron density for the MatB linker (residues 384–388, KATDL) is too weak to observe interactions with R383.

### Comparison to adenylate-forming enzyme core motifs

Highly conserved regions of adenylate-forming enzymes have been categorized into ten major core motifs (A1–A10, Table 2-2) (Marahiel, 1997). Motif A3 contains a serine/threonine-rich region hypothesized to interact with the  $\beta$ - and  $\gamma$ -phosphates of

ATP during the adenylation reaction, stabilizing the release of pyrophosphate; S. coelicolor MatB contains an equivalent region (Marahiel, 1997). The role of motifs A4 and A5 are to bind the carboxylate substrate and stabilize the formation of the adenylate intermediate; ten residues, mostly located between these motifs, have been identified as "specificity-conferring residues" that form the binding site for a carboxylate substrate (Stachelhaus, 1999). A conserved aspartate in motif A4 of A-domains (FDXS) ionically interacts with the α-amino group of bound amino acids; in S. coelicolor MatB, the corresponding residue is V188 (HVHG) (PDB codes: 1AMU and 3E7W) (Conti, 1997; Yonus, 2008). H189, also within S. coelicolor MatB motif A4, coordinates with the specificity-conferring residues S261 and R238 to help form the malonate binding pocket. Within the S. coelicolor MatB motif A5 (ERYGMTE), R238 forms a salt bridge with the extender unit carboxylate in each of the ternary product complex structures; the corresponding residue in the phenylalanine-selective A-domain of the gramicidin synthetase is A322 (NAYGPTE), which helps create a large hydrophobic pocket for the phenylalanine side chain (PDB code: 1AMU) (Conti, 1997). Interestingly, the loop immediately following motif A5, which contains two specificity-conferring residues, is one residue shorter in S. coelicolor MatB than in most adenylate-forming enzymes. Thus, the position of the specificity-determining residue M291 in S. coelicolor MatB is slightly different than the most equivalent residue, C331, in the phenylalanine-activating Adomain of the gramicidin synthetase. In terms of length and orientation, the loop following motif A5 that is most similar to that of S. coelicolor MatB is from a mediumchain adenylate-forming enzyme (PDB code: 3EQ6) (Kochan, 2009).

### Bound malonyl-CoA and methylmalonyl-CoA

Both mCoA and (2R)-mmCoA products are bound between the N-terminal body and C-terminal lid of MatB. Although a few adenylate-forming enzymes have been solved with CoA or thioester products bound (PDB codes: 2P2F, 1PG4, 3EQ6) (Reger, 2007; Gulick, 2003; Kochan, 2009), the MatB complex structures presented here provide the strongest density for a phosphopantetheinyl arm to date. In both structures, the bound extender unit displays electron density that continues past the CoA  $\beta$ -phosphoryl moiety and possesses an average B-factor lower than the average B-factor of MatB. The interactions with the extender unit products can be divided into three categories:

- 1. Malonyl groups are bound primarily by residues S261 and R283 (Figure 2-3d). The presence of the arginine in core motif A5 is a hallmark of malonyl-CoA synthetases. Indeed, R283 serves an important functional role: the ternary complexes reveal that it forms strong ionic interactions with the carboxylates of both mCoA and mmCoA (another active site arginine forms strong ionic interactions with this carboxylate in ATs when the extender unit is selected for transfer to an ACP [Tang 2006]). The observed ψ and Φ angles for S261 are -17.9 ° and 88.5 °, respectively; that these values are outside of the expected Ramachandran regions suggests the importance of S261 in the MatB reaction (Herzberg, 1991). In the ternary complexes presented in this work, S261 participates in a hydrogen-bonding network that includes R283 and the extender unit carboxylate (Figure 2-3d).
- 2. The pantetheine arm is contacted by hydrogen bonds from the backbone carbonyls of residues G392 and G393 as well as hydrophobic interactions

principally mediated by Y384 (as observed in a medium-chain acyl-CoA synthetase ternary complex, PDB code: 3EQ6) (Figure 2-3e) (Kochan, 2009). The A8 core motif, which contains these residues, is hypothesized to possess multiple functions: in the adenylate-forming conformation, the first arginine may help stabilize the pyrophosphate leaving group; in the thioester-forming conformation, it may interact with the pantetheinyl arm, as observed in the MatB structure.

3. The CoA  $\beta$ -phosphate is ionically bound by R236 and R461 (Figure 2-3f). A third arginine, R459, may form ionically interact with the CoA  $\alpha$ -phosphate as it is within 6 Å.

# **Bound adenosine monophosphate**

AMP is bound in the MatB ternary complex structures at a site equivalent to that observed in other adenylate-forming enzymes (e.g., PDB codes: 3EQ6, 1MD9, 3E7W, 1AMU, 2P2F, 3FCC). The  $2F_o$ - $F_c$  electron density maps (contoured at  $\sigma$  = 1.6) do not completely surround AMP; however, the overall B-factors for AMP in the mCoA- and mmCoA-bound MatB structures are 22 Ų and 23 Ų, respectively, which is only slightly higher than the average B-factors for the MatB residues (16.6 Ų and 17.3 Ų, respectively) (Figure 2-4b). The distance from the thioester carbon to the nearest oxygen atom of the AMP phosphate, which had been covalently bonded to one another in the adenylate intermediate, is ~3.6 Å. Positively-charged K390 and K395 form ionic interactions with the AMP  $\alpha$ -phosphate, whereas T287 also forms a hydrogen bond with it (Figure 2-4e). The backbone carbonyl of R283 forms a hydrogen bond to the N7-amine of adenine, while a conserved aspartic acid, D368, interacts with both AMP hydroxyl groups (Figure 2-4c).

### **DISCUSSION**

Malonyl-CoA synthetases are capable of generating many extender units for *in vitro* PKS reactions. *R. trifolii* MatB is frequently used to generate malonyl-CoA for PKSs that utilize α-unsubstituted extender units (Gao, 2010). Here, we have demonstrated that *S. coelicolor* MatB, like *R. trifol*ii MatB, ligates diverse malonate derivatives to CoA to generate a diversity of α-substituted polyketide extender units that are utilized by modular PKSs. We have shown that *S. coelicolor* MatB is also promiscuous toward thiol acceptors, as established by the synthesis of PANT- and NAC-linked extender units. These extender units are nearly as efficient as CoA-linked extender units at transferring malonate derivatives to PKS modules, as demonstrated by our experiments using EryMod6TE in the production of a triketide pyrone (Figure 2-2b).

In an effort to understand the nature of the promiscuity of MatB toward malonate derivatives and thiol acceptors, as well as the stereoselectivity imposed by MatB toward α-substituted malonate derivatives, the structures of two MatB ternary product complexes were obtained. These structures, combined with information from analyses of more thoroughly studied adenylate-forming enzymes, allow the following catalytic cycle to be proposed for MatB-catalyzed synthesis of malonyl-CoA: malonate and ATP bind to MatB, with the binding of ATP potentially causing the C-terminal lid to dock to the N-terminal body in the adenylate-forming conformation (Kochan, 2009). Malonate binds largely via a salt-bridge between one of its carboxylates and R283 and S261 (Figure 2-3d). By analogy with other adenylate-forming enzymes, MatB interacts with ATP by bonding to its phosphates with the serine/threonine-rich A3 core motif, its ribose with the conserved aspartate D368, and its adenine through van der Waals interactions with loop

residues S261-A263 (Figures 2-4c-e). The free carboxylate of the malonate is positioned to attack the ATP α-phosphate, releasing pyrophosphate, which may be stabilized as a leaving group by K476 (Kochan, 2009). After pyrophosphate diffuses from the active site, the C-terminal lid pivots about D387 to adopt the thioester-forming conformation in which the "pantetheine tunnel" is created. The malonyl-adenylate remains bound during this domain alternation process. CoA then binds through hydrogen bonds with the G392 and G393 carbonyls, hydrophobic interactions with Y394, and ionic interactions with R236 and R461 (Figures 2-3e-f). Inspection of the MatB structures does not reveal a site that recognizes the 3',5'- diphosphoadenosine of CoA as is observed other adenylateforming enzymes (e.g., PDB codes: 1PG3 and 2P2F). The entry of CoA into the MatB active site is halted when the diphosphate moiety forms salt-bridges with positively charged residues at the tunnel opening (Figure 2-3f), although these ionic interactions are not required for catalysis because PANT and NAC can serve as thiol acceptors. Because the pantetheinyl arm makes as many contacts to the C-terminal lid as it does with the Nterminal body, CoA likely enters the pantetheine tunnel after tunnel formation. After an attack of the CoA thiolate on the malonyl-adenylate carbonyl, the leaving group AMP, stabilized by K395, is generated. Finally, AMP and the malonyl-CoA extender unit diffuse out of the active site, aided by the pivoting motion of the C-terminal lid during domain alternation.

The 2F<sub>o</sub>-F<sub>c</sub> electron density maps from the mmCoA/AMP-bound MatB ternary complex suggest that MatB forms (2*R*)-mmCoA and not (2*S*)-mmCoA. In related structures, the position of a bound AMP is similar to its position within the adenylate intermediate, thus this intermediate can be approximately reconstructed from the positions of the malonyl group and AMP (Reger, 2007; Reger, 2008; Conti, 1997; Yonus, 2008; Hisanaga, 2004; Kochan, 2009). Due to the proximity of the α-substituent and the

AMP phosphoryl moiety (~4 Å in the ternary product complex and likely closer within the adenylate intermediate), the stereochemistry of the α-carbon is limited to the *R*-configuration (Figure 2-5c). The formation of (2*R*)-mmCoA is consistent with observations from previous experiments that failed to synthesize polyketides from mmCoA produced by MatB in the absence of an epimerase (Murli, 2003). Thus, we hypothesize that MatB also synthesizes (2*R*)-emCoA, (2*R*)-hmCoA, and (2*R*)-oxCoA in a stereocontrolled manner. As the AT domains of PKSs are very stereoselective, only accepting (2*S*)-mmCoA and (2*S*)-emCoA, epimerization of mmCoA and emCoA is necessary for the incorporation of these MatB-synthesized extender units into PKSs and can be accomplished via spontaneous epimerization under acidic conditions or enzymatically-catalyzed epimerization through the addition of an epimerase (e.g., mmCoA epimerase) (Murli, 2003).

Organic syntheses of mSNAC, mmSNAC, oxSNAC, and hmSNAC have been reported (Pohl, 1998; Carroll, 2002); however, the syntheses of hmSNAC and oxSNAC involve multistep syntheses with 13% and 29% yields, respectively (Carroll, 2002). Using MatB, super-stoichiometric quantities of extender units can be enzymatically produced *in situ* to drive the synthesis of milligram quantities of polyketides from *in vitro* reactions. MatB also serves as an extender unit regeneration system, as supported by the formation of mm-S-NAC in the Mod6TE reactions. With the addition of excess diacid and ATP, MatB will continually regenerate product by recycling the appropriate thiol acceptor (e.g., CoA, PANT, or SNAC). The MatB-catalyzed formation of α-substituted extender units enables the generation of otherwise unavailable isotopically- or radioactively-labeled extender units for more detailed *in vitro* PKS studies. The generation of hmCoA and oxCoA also suggests the potential of genetically engineering PKSs to produce sugar derivatives.

The hydrophobic residues V188 and M291 adjacent to the α-methyl group of (2R)-mmCoA, as observed in the mmCoA/AMP/MatB ternary complex, likely influence the promiscuity of MatB for α-substituted malonate derivatives and may explain why S. coelicolor MatB was unable to synthesize aminomalonyl-CoA. Within A-domains, an aspartate is usually positioned in the A4 core motif to ionically interact with the amino group of a selected amino acid; within MatB, this position is occupied by V188. The chemical incompatibility of a valine side chain with a charged amino group may preclude the formation of aminomalonyl-CoA. A structural alignment of MatB with a D-alanine:D-alanyl carrier protein ligase (PDB code: 3E7W) suggests that the MatB mutant V144D would bind aminomalonate and catalyze the formation of aminomalonyl-CoA (Figures 2-5a-b). We have not tested for the formation of the less commonly utilized extender units propylmalonyl-CoA or chloroethylmalonyl-CoA; however, if MatB is not capable of selecting the corresponding diacids, appropriate mutagenesis of V188 and M291 may enable the formation of these extender units.

Thus, an enzymatic means to produce CoA-, PANT-, and NAC-linked polyketide extender units is provided. This reaction may be utilized by many PKS laboratories to produce significant quantities of extender units, especially when they are economically unviable or commercially unavailable. The utility of MatB for *in vitro* PKS reactions was demonstrated not only through the incorporation of the methylmalonyl groups of mmCoA, mmPANT, and mm-S-NAC by a PKS module to form a polyketide product but also through the ability to observe reaction progress by monitoring the production of the coupled product, mm-S-NAC. The MatB ternary product complexes have provided chemical insights into the malonyl-CoA synthetase reaction as well as structural insights into substrate promiscuity. Efforts to alter the stereochemistry of α-substituted extender unit products will most likely not be successful due to the hypothesized geometry of the

adenylate intermediate; however, the range of the sizes and chemistries of  $\alpha$ -substituents of malonate derivatives ligated by MatB is quite large, and apparently only limited by the nearby residues V188 and M291.

# MATERIALS AND METHODS

# Synthesis of compounds utilized

# *N-acetylcysteamine (NAC)*

Cysteamine hydrochloride (1.14 g, 10.0 mmol, 1.0 eq.), KOH (0.56 g, 10.0 mmol, 1.0 eq.), and NaHCO<sub>3</sub> (2.52 g, 30.0 mmol, 3.0 eq) was dissolved in a round-bottom flask containing 50 ml water. After the dropwise addition of acetic anhydride (0.95 mL, 10.0 mmol, 1.0 eq.), the solution was stirred at 22°C for 10 minutes. The pH was then adjusted to 7.3 using 12.1 N HCl. The resulting mixture was extracted with ethyl acetate (50 mL x 3). The organic layer was dried using MgSO<sub>4</sub>, and after filtration a quantitative yield of N-acetylcysteamine was isolated *in vacuo*.

<u>N-acetylcysteamine (NAC)</u> <sup>1</sup>H-NMR (CDCl<sub>3</sub>) 400 MHz,  $\delta$  (ppm): 5.97 (1H, bs), 3.40 (2H, q, J = 8.0), 2.65 (2H, m, J = 8.4), 1.98 (3H, s), 1.33 (1H, t, J = 8.8).

# Methoxymalonic acid

Dimethyl methoxymalonate (0.85 mL, 6.2 mmol, 1.0 eq.) and NaOH (2.4 g, 60 mmol, 9.7 eq.) was dissolved in a round-bottom flask containing 10 mL water. The resulting solution was stirred at 65°C overnight. The solution was then acidified with 10 mL 12.1 N HCl and extracted with ethyl acetate (30 mL x 3). The organic layer was dried using MgSO<sub>4</sub>, and after filtration a quantitative yield of methoxymalonic acid was isolated *in vacuo*.

# **D-pantetheine**

D-pantethine (0.021 mmol, 1.0 eq.) and DTT (0.022 mmol, 1.0 eq.) were dissolved in 100  $\mu$ l 20% (w/v) glycerol (aq.). The solution was incubated at 60°C for 15 minutes, and the resulting D-pantetheine solution was stored at 22°C.

## Cloning

The matB gene was amplified from S. coelicolor genomic DNA using primers 5'-TCGATTGCACATATGTCCTCTCTCTCTCCCGGCCCTCT-3' 5'and ATCGGATAGCTCGAGTCACGGTTCAGCGCCCGCTT-3'. The gene encoding methylmalonyl-CoA epimerase was amplified from S. coelicolor genomic DNA using primers 5'-ATCCCGAATCATATGCTGACGCGAATCGACCA-3' and 5'-TTAGTCTGGCTCGAGTCAGTGCTCAGGTGACTCAA-3'. Fragments were digested with NdeI and XhoI (italicized) and ligated into pET28b plasmid between the corresponding restriction sites. Plasmid design allowed for the incorporation of a stop codon (underlined) at the 3' terminus of the DNA encoding both MatB and methylmalonyl-CoA epimerase. Synthetic DNA encoding the N-terminal docking domain from DEBS3 fused to the sixth module and thioesterase from DEBS3 was a gift from Kosan Biosciences (Menzella, 2005). The DNA was digested at the flanking NdeI and *Eco*RI restriction sites and ligated into pET28b.

# **Protein expression and purification**

MatB and methylmalonyl-CoA epimerase expression plasmids were transformed into *Escherichia coli* BL21(DE3), whereas the EryMod6TE expression vector was transformed into *E. coli* K207-3 cells (Murli, 2003). Starter cultures (50 ml) were grown overnight and used to inoculate prewarmed (37°C) Luria broth, supplemented with 50 mg/L kanamycin. When  $OD_{600} = 0.4$ , the media was cooled (15°C) and protein

expression was induced with 0.5 mM IPTG. After 16 hr, cells were harvested by centrifugation, resuspended in lysis buffer (10% glycerol, 0.5 M NaCl, 100 mM HEPES pH 7.5), sonicated, and centrifuged (30,000 relative centrifugal force for 45 min) to remove cellular debris. Cell-free lysate was passed over a nickel-NTA column equilibrated with lysis buffer. The column was washed with lysis buffer containing 15 mM imidazole and protein was eluted using lysis buffer containing 150 mM imidazole. Final protein concentrations were determined using a Thermo Scientific Nanodrop 1000.

MatB was further purified for crystallization via gel filtration using a Superdex 200 column equilibrated in gel filtration buffer (10% glycerol, 150 mM NaCl, and 10 mM HEPES pH 7.5). By comparison to a gel filtration standard (Bio-Rad), MatB eluted as a ~50 kDa monomer. Fractions were collected and concentrated to 15 mg/ml in gel filtration buffer (Millipore Centrifugal Filter Unit, 30 kDa MWCO). The final protein concentration was determined by using a Thermo Scientific Nanodrop 1000.

The same purification protocol was followed for selenomethionine-labeled MatB (Se-MatB); however, growth and expression conditions differed. Each liter of M9 growth medium contained 6 g Na<sub>2</sub>HPO<sub>4</sub>, 3 g KH<sub>2</sub>PO<sub>4</sub>, 500 mg NaCl, 1 g NH<sub>4</sub>Cl, 0.1 mM CaCl<sub>2</sub>, 1 mM MgSO<sub>4</sub>, 0.4% glucose, and 20 mg kanamycin. Prewarmed (37°C) media was inoculated with saturated starter culture (3 ml). When the OD<sub>600</sub> = 0.4, each liter of media was supplemented with lysine, phenylalanine, and threonine (each 100 mg), isoleucine, leucine, and valine (each 50 mg), and selenomethionine (60 mg). After 15 min protein expression was induced with 0.5 mM IPTG. The cultures were then cooled (15°C) and after 16 hr cells were harvested. Protein purification, gel filtration, and concentration of Se-MatB protein followed the methods described above.

# **Protein crystallization**

The following conditions were used for the protein solution of mmCoA/AMPbound selenomethionine-labeled MatB: 15.8 mg/ml Se-MatB, 1 mM CoA, 5 mM methylmalonate pH 7.5, 1 mM ATP, 2 mM DTT, and 5 mM MgCl<sub>2</sub>. Optimal mmCoA/AMP-bound selenomethionine-labeled MatB crystals grew overnight in 31% PEG 4000, 100 mM MgCl<sub>2</sub>, and 100 mM Tris-HCl pH 7.5 via sitting drop vapor diffusion at 22°C with a protein solution to crystallization buffer ratio of 1:1. The following conditions were used for the protein solution of mCoA/AMP-bound MatB: 10 mg/ml MatB, 20 mM CoA, 24 mM sodium malonate pH 7.7, 24 mM ATP, and 48 mM MgCl<sub>2</sub>. Optimal mCoA/AMP-bound MatB crystals grew overnight in 35% PEG 4000, 100 mM MgCl<sub>2</sub>, and 100 mM Tris-HCl pH 8.2 via sitting drop vapor diffusion at 22°C with a protein solution to crystallization buffer ratio of 1:1. The following conditions were used for the protein solution of mmCoA/AMP-bound MatB: 9.6 mg/ml MatB, 20 mM CoA, 24 mM sodium methylmalonate pH 7.5, 24 mM ATP, and 48 mM MgCl<sub>2</sub>. Optimal mmCoA/AMP-bound MatB crystals were grown overnight in 33% PEG 4000, 100 mM MgCl<sub>2</sub>, and 100 mM Tris-HCl pH 9.5 via sitting drop vapor diffusion at 22°C with a protein solution to crystallization buffer ratio of 1:1.

## Data collection, processing and refinement

Diffraction data were collected at ALS beamline 5.0.2, then integrated and scaled using HKL2000 (Otwinowski, 1997). The structure of mmCoA/AMP-bound selenomethionine-labeled MatB was solved by single-wavelength anomalous dispersion using Phenix (Adams, 2002). This structure was used as a search model for molecular replacement using Phaser (CCP4, 1994), yielding the solutions for both mCoA/AMP-bound MatB and mmCoA/AMP-bound MatB. Water molecules were identified using Coot ( $F_{\circ}$ - $F_{c}$  map contoured at  $\sigma$  = 3.0) (Emsley, 2004). Refinement cycles using Coot

(Emsley, 2004) and Refmac5 (CCP4, 1994) were performed until the R-factors could no longer be improved (Table 2-1). Figures were prepared using PyMol (DeLano, 2002) and LIGPLOT (Wallace, 1995).

# **HPLC** analysis of MatB reactions

Enzymatic reactions to generate extender units were set up using the following conditions: 6 nM MatB, 0.55 mM CoA or 1.0 mM D-pantetheine or 4.2 mM NAC, 1.3 mM ATP, 22 mM malonate derivative, 9 mM MgCl<sub>2</sub>, 15% (w/v) glycerol, and 100 mM HEPES pH 7.5. All reactions were incubated at 22°C overnight. Samples were analyzed by HPLC (Waters) using a Varian Microsorb-MV 300-5 C<sub>18</sub> column and a Waters 2998 photodiode array detector. The mobile phases consisted of water containing 0.1% TFA (solvent A) and methanol containing 0.1% TFA (solvent B). A linear gradient (flow rate = 1 ml/min) of 0%–67% B over 20 minutes was used for sample analysis. Reactions were monitored at 254 nm (adenine, CoA) or 235 nm (thioester bond, NAC, and PANT compounds).

# HPLC analysis of triketide pyrone reactions

MatB reactions were set up using the following conditions: 5 mM CoA or 5 mM D-pantetheine or 5 mM NAC, 100 mM HEPES pH 7.5, 100 mM MgCl<sub>2</sub>, 20 mM ATP (in 100 mM HEPES pH 7.5), 50 mM sodium methylmalonate pH 7.2, 10% glycerol, and 8 nM MatB. Reactions were incubated at 22°C overnight and analyzed by HPLC to ensure formation of mmCoA, mmPANT, or mm-S-NAC. EryMod6TE reactions were set up by adding (2S, 3R)-2-methyl-3-oxopentanoate-S-NAC (diketide-S-NAC) (10 mM, final), S. coelicolor methylmalonyl-CoA epimerase (12 nM, final), and EryMod6TE (4 nM, final) to the MatB reaction, resulting in a 2-fold dilution of the MatB reaction. Reactions were incubated at 22°C for 16 hours. Samples were analyzed by HPLC (Waters) using a

Varian Microsorb-MV 300-5  $C_{18}$  column and a Waters 2998 photodiode array detector. The mobile phases consisted of water containing 0.1% TFA (solvent A) and methanol containing 0.1% TFA (solvent B). A linear gradient (flow rate = 1 ml/min) of 0%–100% B over 15 minutes was used. Reactions were monitored at 254 nm (adenine, CoA), 235 nm (thioester bond, NAC, and PANT compounds), and 290 nm (triketide pyrone).

## Mass spectrometry analysis

CoA-linked extender units were collected from HPLC runs and isolated *in vacuo*. Samples were subjected to low-resolution negative-ESI mass spectrometry performed with a Finnigan LCQ ion trap mass spectrometer with the needle voltage set to 3 kV and the capillary temperature set to 120°C.

## LC/MS analysis

The triketide pyrone and two products from the hm-S-NAC reaction were collected, isolated *in vacuo*, resuspended in 25% acetonitrile and 0.1% formic acid, and subjected to positive-ESI LC/MS (Agilent Technologies 1200 Series HPLC with a Gemini  $C_{18}$  column coupled to an Agilent Technologies 6130 quadrupole mass spectrometer). Mobile phases consisted of water containing 0.1% formic acid (solvent A) and acetonitrile containing 0.1% formic acid (solvent B). A linear gradient (flow rate = 0.7 ml/min) of 5%–95% B over 12 minutes was used.

## **Accession numbers**

Coordinates were deposited with PDB Codes 3NYR and 3NYQ for mCoA/AMP-bound and (2R)-mmCoA/AMP-bound MatB, respectively.

Scheme 2-1 MatB reaction mechanism.

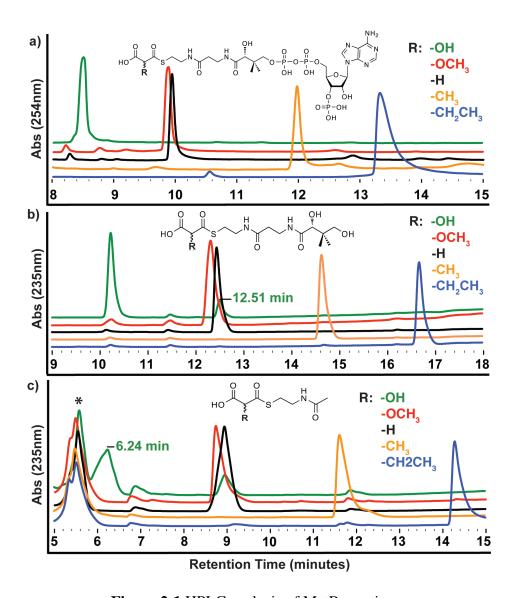


Figure 2-1 HPLC analysis of MatB reactions.

(a–c) MatB synthesized 15 polyketide extender units valuable for *in vitro* PKS reactions. Retention times (min) of extender units formed using CoA, PANT, and NAC as the thiol acceptor are indicated. The observed masses (calculated masses) from the hmCoA, oxCoA, mCoA, mmCoA, and emCoA reactions were 867.8(868.1), 881.9(882.1), 851.7(852.1), 865.9(866.1), and 879.9(880.1), respectively. The two peaks observed for hm-S-NAC at 6.24 and 8.97 min were analyzed by positive-ESI LC/MS and correspond to hm-S-NAC (222.2(222.0)) and m-S-NAC (206.2(206.1)); the presence of m-S-NAC is may be from a malonate impurity in the tartronic acid stock. Colors indicated bound products: green – hydroxymalonyl-; red – methoxymalonyl-; black – malonyl-; orange – methylmalonyl-; blue - ethylmalnoyl-; \*AMP.

**Scheme 2-2** Enzymatic synthesis of triketide pyrone by EryMod6TE.

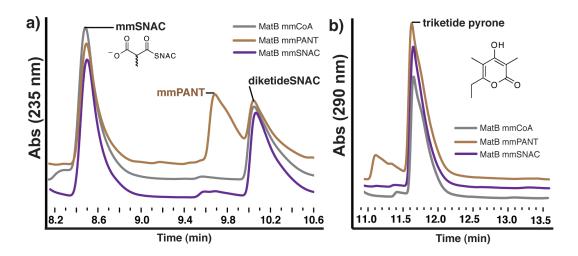


Figure 2-2 HPLC analysis of the formation of triketide pyrones.

a) HPLC analysis of Mod6TE reactions at 235 nm. The formation of mm-S-NAC (from the transthioesterification of diketide-S-NAC) may be used as a diagnostic tool to follow an in vitro PKS reaction. b) HPLC analysis of Mod6TE reactions at 290 nm. Triketide pyrone was synthesized from all three different extender units, (2S)-mmCoA/PANT/NAC. The identity of triketide pyrone was confirmed by positive-ESI LC/MS with an observed mass (calculated mass) of 169.2(169.1).

**Table 2-1** Data collection and refinement of *S. coelicolor* MatB structures.

	T		I		
	MatB	MatB	SeMatB		
	(mCoA)	(2R-mmCoA)	(2R-mmCoA)		
	3NYR	3NYQ	(2K-IIIIICOA)		
Space Group	C222(1)	C222(1)	C222(1)		
Cell dimensions					
a, b, c (Å)	73.2, 86.6,	73.4, 86.8,	72 0 97 4 154 4		
a, b, c (A)	153.1 153.4		73.9, 87.4, 154.4		
$a = \beta = g (^{\circ})$	90	90	90		
Resolution (Å)	50-1.45	50-1.43	50-1.75		
R <sub>merge</sub>	0.051 (0.439)	0.062 (0.292)	0.080 (0.202)		
I/σ(Î)	23.1 (2.6)	15.5 (3.9)	30.8 (7.1)		
Completeness (%)	99.8 (99.4)	98.5 (82.4)	97.8 (94.7)		
Redundancy	7.1 (4.5)	6.8 (3.8)	3.9 (4.0)		
Refinement					
Resolution (Å)	50-1.45	50-1.43	50-1.75		
Unique reflections	621,987	671,166	537,849		
$R_{\text{work}}/R_{\text{free}}$	0.1997/0.2247	0.1951/0.2161	0.2026/0.2353		
Ramachandran	99.34(0.66)	00.24(0.66)	99.55(0.45)		
Allowed(Outlier)(%)	99.34(0.00)	99.34(0.66)			
Number of Atom					
Protein	3378	3378	3320		
Extender unit	28	29	29		
AMP	23	23	N/A		
Water	347	404	338		
B factors (Å <sup>2</sup> )					
Protein	16.72	17.40	19.66		
Extender unit	14.60	14.29	21.06		
AMP	21.94	23.79	N/A		
Water	26.50	27.64	27.75		
Rmsd					
Bond lengths (Å)	0.010	0.013	0.011		
Bond angles (°)	1.80	1.77	1.45		
Highest Res. Bin (Å)	1.49-1.45	1.47-1.43	1.78-1.75		
<u> </u>	•		•		

AMP: Adenosine monophosphate; RMSD: root-mean-square deviations

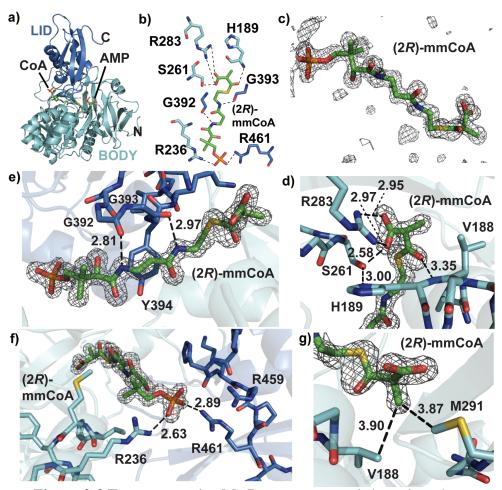


Figure 2-3 Ternary complex MatB structure reveals bound product.

a) Overall structure of the mmCoA/AMP/MatB ternary complex. N-terminal body (light blue), C-terminal lid (dark blue) and (2R)-mmCoA and AMP (green). b) LIGPLOT representation of interactions observed between (2R)-mmCoA and surrounding MatB residues. c) Simulated-annealing omit map of (2R)-mmCoA (contoured at  $\sigma = 3.0$ ). d) H189, S261, and R283, interact with methylmalonyl moiety of (2R)-mmCoA. A hydrogen-bonding network is formed between S261, R283, and the (2R)-mmCoA carboxylate (distances in Å,  $2F_o$ - $F_c$  map contoured at  $\sigma = 1.6$ ). e) Carbonyls from G392 and G393 hydrogen bond with the pantetheine amide groups (distances in Å,  $2F_o$ - $F_c$  map contoured at  $\sigma = 1.6$ ). f) R236 and R461 form salt-bridges with the (2R)-mmCoA  $\beta$ -phosphate. No density was observed for the  $3_o$ ,5 $_o$ -diphosphoadenosine; however, the positively charged pantetheine tunnel entrance likely interacts with both  $\alpha$ - and  $\beta$ -phosphate groups (distances in Å,  $2F_o$ - $F_c$  map contoured at  $\sigma = 1.6$ ). g) V188 and M291 form a hydrophobic pocket near the  $\alpha$ -methyl substituent of (2R)-mmCoA (distances in Å,  $2F_o$ - $F_c$  map contoured at  $\sigma = 1.6$ ).

 Table 2-2 Core motifs conserved in adenylate-forming enzymes.

Core	Function	Adenylate-Forming Enzyme Consensus Sequence	Corresponding MatB Sequence
<b>A1</b>	Structural	L(TS)YxEL	(29)LTYAEL(34)
<b>A2</b>	Structural	LKAGxAYL(VL)P(LI)D	(69)LLAGVAAVPLN(79)
A3	Stabilize pyrophosphate leaving group	LAYxxYTSG(ST)TGxPKG	(137)PALVVYTSGTT- GPPKG(152)
A4	Asp interacts with α-amino groups of peptides	FDxS	(187)HVHG(190)
A5	Stabilize product	NxYGPTE	(282)ERYGMTE(288)
<b>A6</b>	Unknown	GELxIxGxG(VL)ARGYL	(334)GEIQVRGPNLFT- EYL(348)
<b>A7</b>	Stabilize ATP/AMP ribose	Y(RK)TGDL	(364)FRTGDM(369)
A8	Essential for adenylation	GRxDxQVKIRGxRIELGEI E	(382)GRKATDLIKSGGY- KIGAGEIE(402)
A9	Unknown	LPxYM(IV)P	(454)LAPHKRP(460)
A10	Stabilize ATP/AMP ribose	NGK(VL)DR	(474)MGKIMK(479)

AMP: adenosine monophosphate; ATP: adenosine triphosphate

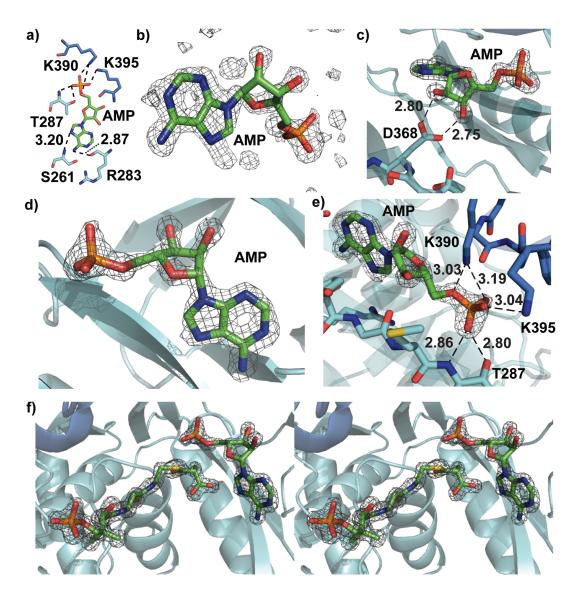


Figure 2-4 Ternary complex MatB structure reveals AMP leaving group bound.

a) LIGPLOT diagram of AMP and MatB interactions (distances in Å). b) Simulated-annealing omit map of AMP (contoured at  $\sigma = 3.0$ ). c) Interaction between conserved D368 and hydroxyl groups of AMP ( $2F_o$ - $F_c$  map contoured at  $\sigma = 1.6$ ). d)  $2F_o$ - $F_c$  electron density map surrounding AMP (contoured at  $\sigma = 1.6$ ). e) Ionic interactions between the AMP  $\alpha$ -phosphate and both K390 and K395. The backbone amide and hydroxyl group of T287 form hydrogen bonds with the AMP  $\alpha$ -phosphate (distances in Å,  $2F_o$ - $F_c$  map contoured at  $\sigma = 1.6$ ). f) Stereodiagram of (2R)-mmCoA and AMP within the active site (distances in Å,  $2F_o$ - $F_c$  map contoured at  $\sigma = 1.6$ ).

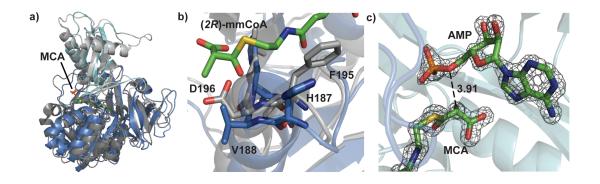


Figure 2-5 Engineering MatB substrate specificity and product stereochemistry.

a) Structural alignment of MatB (blues) and DltA (greys, PDB code: 3E7W). DltA is an A-domain that ligates D-alanine onto the phosphopantheinyl arm of a peptidyl carrier protein. b) D196 (gray) of the FDxS motif in DltA ionically interacts with the of D-alanine amino group. The MatB mutant V188D could catalyze the formation of aminomalonyl-CoA. c) The distance between the AMP leaving group and the  $\alpha$ -carbon is 3.91Å. Re-engineering MatB to form (2S)- $\alpha$ -substituted extender units may not be possible because the  $\alpha$ -substituent of a (2S)- $\alpha$ -substitued-malonyl adenylate intermediate would clash with the AMP phosphoryl moiety. MatB is hypothesized to synthesize (2R)-mmCoA, (2R)-emCoA, (2R)-hmCoA, and (2R)-oxCoA (2F<sub>o</sub>-F<sub>c</sub> map contoured at  $\sigma$  = 1.6).

# Chapter 3 Employing a Polyketide Synthase Module and Thioesterase in the Semipreparative Biocatalysis of Diverse Triketide Pyrones

#### ABSTRACT

To demonstrate the potential of polyketide synthase (PKS) modules as biocatalysts that can synthesize polyketides in vitro, the terminal module and thioesterase (TE) from the erythromycin PKS were employed in the multimilligram, chemoenzymatic syntheses of diverse triketide pyrones. Methylmalonyl-S-NAC extender units were generated by the promiscuous malonyl-CoA ligase Streptomyces coelicolor MatB. Initiating the reaction with  $\beta$ -ketoacyl-N-acetylcysteamine thioester diketides yielded the anticipated triketide pyrones through chain extension and cyclization. TE-mediated hydrolysis of the substrates prevented quantitative yields; however, a reaction with 200 mg of diketide substrate generated 20 mg of pyrone product. Trace quantities of a particular pyrone were observed in all reactions and was hypothesized to result from extender unit decarboxylation and subsequent "stuttering". To harness this polymeraselike activity, reactions were initiated with diverse acyl-N-acetylcysteamine thioester compounds to yield the anticipated triketide pyrones, one harboring a terminal alkyne chemical handle. This biocatalytic system enables more informative analysis of in vitro PKS reactions by HPLC, NMR, and crystallography and sets the stage for the preparative generation of polyketides such as chiral building blocks valuable in the synthesis of natural products and pharmaceuticals.

# Introduction

Trace quantities of triketide products have been generated, often from radiolabeled extender units, in enzymology studies of PKS modules fused to TEs (ModTEs). Larger quantities of triketide products have been generated by feeding diketide substrates to bacteria expressing ModTEs (Regentin, 2004). Here we show that the terminal module and TE of the erythromycin PKS (EryMod6TE) is capable of generating triketides in vitro on a scaled-up level from simple synthetic diketides and enzymatically generated building blocks. The triketide pyrone products could be readily observed by HPLC and characterized by NMR. Encouragingly, EryMod6TE possesses broad substrate specificity, which enables it to be employed as a general biocatalyst. A scaled-up reaction with a diketide revealed that the yields were limited to ~12\% due to the competing side-reaction of TE-catalyzed hydrolysis of the diketide substrate. TEs from other synthases were fused to EryMod6; while they were capable of forming pyrones, they also catalyzed hydrolysis. We discovered that EryMod6TE is capable of "stuttering", a phenomenon in which one module catalyzes the addition of multiple extender units to a growing polyketide. This ability was harnessed biocatalytically to generate the anticipated pyrones from monoketides.

### RESULTS

## Triketide pyrones from diketide substrates

The promiscuous malonyl-CoA ligase *Streptomyces coelicolor* MatB was utilized to generate the extender unit methylmalonyl-S-NAC (**3.1**) (Hughes, 2011). As methylmalonyl groups can be transferred from **3.1** onto PKSs, costly methylmalonyl-CoA is no longer necessary for scaled-up *in vitro* PKS reactions (Scheme 3-1). As the

described biocatalytic schemes are designed to be both rapid and inexpensive, neither overexpressed MatB nor EryMod6TE was purified from the lysate of the expression host.

Each β-ketoacyl-*S*-NAC **3.2–3.7** was added to MatB-generated extender unit **3.1** and lysate from *Escherichia coli* K207-3 cells that overexpressed EryMod6TE (Scheme 3-1) (Murli, 2003; Piasecki, 2011). After incubating the reactions overnight at ambient temperature, triketide pyrones **3.8–3.13** were extracted, purified, and analyzed by reverse phase chromatography, LC/MS, high-resolution mass spectrometry, and <sup>1</sup>H NMR (Table 3-1). The C-2 methyl hydrogens and the C-4 hydrogen produced diagnostic peaks in the <sup>1</sup>HNMR spectra at 1.9 ppm and 5.8 ppm, respectively. Pyrones **3.8–3.13** were further characterized by their reverse phase chromatographic retention times and absorbance spectra ( $\lambda_{max}$  = 288.7 nm for **3.8**, **3.9**, and **3.11–3.13**,  $\lambda_{max}$  = 292.3 nm for **3.10**) (Figure 3-1).

HPLC analysis of triketide pyrone production under identical reaction conditions enabled an estimate of the relative titers obtained (assuming similar extinction coefficients for the pyrone products) and provided insight as to the optimal diketide substrate for EryMod6TE-catalyzed pyrone formation. Previous experiments demonstrated that EryKS6 accepts the diketide (2*S*, 3*R*)-3-hydroxy-2-methylpentanoyl-*S*-NAC (Wu, 2000). Our experiments show that each of the β-ketoacyl-*S*-NAC substrates 3.2–3.7 are also accepted, which indicates that the substrate specificity of EryKS6 is very broad. Based on the integrations of product peaks, reactions with 3.2 yielded the least triketide pyrone product, while reactions with 3.4 yielded the most.

To examine the scalability of these *in vitro* reactions, a large-scale reaction was prepared using 200 mg of diketide **3.4**. The reaction was complete after 24 hours, as monitored by reverse phase chromatography. After ethyl acetate extraction and silica gel purification, 20 mg of pyrone **3.11** was obtained, representing an overall yield of 12%.

We hypothesized that the majority of substrate **3.4** was hydrolyzed by EryTE. While we seek to increase the overall yield of this process, the reactants and the catalysts are easily and inexpensively obtained so sub-quantitative yields are tolerable. A <sup>13</sup>C NMR of **3.11** was readily acquired (Appendix 1).

# Triketide pyrones produced by "stuttering"

In each of the reactions with diketide substrates and **3.1**, some quantity of pyrone **3.10** was produced. We hypothesized that this pyrone was derived from the decarboxylation of extender unit **3.1**, priming of EryKS6 by the resulting propionyl-S-NAC, two extension cycles (*i.e.* stuttering), and cyclization. Stuttering, the phenomenon of a module catalyzing more than one condensation, is considered normal in the biosynthesis of the polyketides stigmatellin (Gaitatzis, 2002), borrelidin (Olano, 2004), aureothin (He, 2003), and lankacidin (Mochizuki, 2003) but abnormal in the biosynthesis of such polyketides as epothilone (Hardt, 2001) and erythromycin (Wilkinson, 2000) (a small quantity of octaketide is produced by the erythromycin PKS due to stuttering by the fourth module). The isolated fifth module of the pikromycin PKS has been demonstrated to stutter *in vitro*, yielding a reduced triketide lactone when supplied with methylmalonyl-CoA and NADPH (Beck, 2003).

With the goal of harnessing the stuttering activity of EryMod6TE to biocatalytically generate triketide pyrones from monoketide priming units, acyl-S-NACs **3.14–3.18** were incubated overnight with MatB-generated **3.1** and EryMod6TE lysate (Scheme 3-2). Triketide pyrone products **3.19–3.22** were extracted and analyzed by reversed-phase chromatography, LC/MS, and high-resolution mass spectrometry (an <sup>1</sup>H NMR spectrum was also obtained for **3.20**) (Figure 3-2, Table 3-1, and Appendix 1). An estimation of pyrone production through peak integration indicates that butyryl-S-NAC

(3.16) is converted to pyrone through the stuttering activity of EryMod6TE to the greatest extent. Crotonyl-S-NAC was not detectably converted. These observations suggest that the greater flexibility of the butyryl group enables it to be accepted by EryKS6 more readily than the rigid crotonyl group.

# **TE-catalyzed substrate hydrolysis**

Quantitative yields of triketide pyrones were not obtained from diketide-S-NACs **3.2–3.7** and acyl-S-NACs **3.14–3.18** due to a side-reaction: TE-mediated thioester hydrolysis (Sharma, 2007). To determine how quickly different substrates hydrolyze under the employed reaction conditions, extender unit **3.1** was excluded from four EryMod6TE reactions and hydrolysis was monitored by reverse phase chromatography (Figure 3-3). Hydrolysis half-lives were measured to be  $\sim$ 5.9 h for  $\alpha$ -substituted  $\beta$ -ketoacyl-S-NAC **3.7**,  $\sim$ 4.4 h for acyl-S-NAC **3.15**,  $\sim$ 3.8 h for acyl-S-NAC **3.16**, and  $\sim$ 2.0 h for  $\beta$ -ketoacyl-S-NAC **3.4** (Appendix 1).

We hypothesized that other TEs might not hydrolyze substrates as rapidly as EryTE, permitting a more complete conversion of a diketide substrate to a triketide product. The TE domains from the amphotericin PKS (Caffrey, 2003), the radicicol PKS (Reeves, 2008), and the calcium-dependent antibiotic NRPS (Bentley, 2002) were fused to EryMod6 to create EryMod6AmpTE, EryMod6RadTE, and EryMod6CdaTE (Figure 3-4). Each of the four enzyme complexes was incubated with 3.4 and MatB-generated 3.1. After 24 hours, each reaction had gone to completion, as judged by the disappearance of 3.4. Pyrone formation was analyzed by reverse phase chromatography. Each of the TEs catalyzed pyrone formation; however, none of the new constructs produced more triketide pyrone than native EryMod6TE (Figure 3-5, Appendix 1). Thus EryTE outperformed the other TEs, exhibiting the highest ratio of triketide pyrone cyclized to

diketide substrate hydrolyzed. In order to obtain higher yields of triketide pyrone products, we are currently investigating the use of acyl-CoA ligases that will regenerate the thioester substrates hydrolyzed by TEs.

### **DISCUSSION**

This study sets the stage for utilizing PKS modules biocatalytically to produce preparative quantities of valuable polyketides. While our goal is to produce longer polyketides through in vitro biocatalysis, we currently aim to generate reduced triketide lactone chiral building blocks useful in the syntheses of medicinally- relevant polyketides (Hanessian, 1993). Such molecules would be most amenable to synthetic efforts if a chemical handle were present at the terminus of the polyketide fragment. We demonstrated the feasibility of introducing such functionality through the generation of pyrone **3.19**, which contains a terminal alkyne. The diversity of triketides pyrones formed using EryMod6TE as a biocatalyst reveals the substrate promiscuity of its enzymatic domains and illustrates the remarkable biosynthetic potential of such PKS enzymes in vitro. While other classes of PKSs may be useful in the biocatalytic generation of pyrones (e.g. chalcone synthase [Ferrer, 1999], 6-methylsalicylic acid synthase [Beck 1990]), they lack the versatility of modular PKSs and are incapable of synthesizing reduced polyketides with multiple stereocenters. Further studies of TE-catalyzed hydrolysis and employing acyl-CoA ligases to regenerate hydrolyzed substrate should enable higher yields of polyketide products.

In addition to accelerating the total synthesis of medicinally relevant natural products, the described biocatalytic systems will accelerate PKS enzymology. The chemical structures of polyketides naturally report on the activities of the enzymes that

produce them. Since polyketides can now be obtained through *in vitro* studies in quantities sufficient for characterization by NMR and small-molecule crystallography, much more will be learned about the PKS enzymes that catalyze their formation.

### MATERIALS AND METHODS

# Synthesis of compounds utilized

# *N-acetylcysteamine (NAC)*

Into a flask containing 20 mL deionized water was dissolved acetyl-N-acetylcysteamine (NAC) thioester (1 mL, 9.4 mmol, 1 eq.) and sodium hydroxide (1.1 g, 28.3 mmol, 3 eq.). The reaction was stirred for 3 hours at 23 °C and product was extracted with ethyl acetate (80 mL x 3). The organic layer was dried with MgSO<sub>4</sub>, and the product was isolated *in vacuo*.

# **β-ketoacyl-S-NAC** diketides

The syntheses of diketides **3.2–3.4**, and **3.7** were performed as previously described (Piasecki, 2011). Diketides **3.5** and **3.6** were synthesized using the following protocol: into a round-bottom flask containing anhydrous dichloromethane (50 mL) cooled to 0 °C was added pyridine (3.44 mL, 42 mmol, 2 eq.) and Meldrum's acid (3.0 g, 20.8 mmol, 1 eq.). Hexanoyl chloride (3.49 mL, 25 mmol, 1.2 eq.) or valeryl chloride (2.97 mL, 25 mmol, 1.2 eq.) was added dropwise to solution and stirred overnight at 23 °C. The reaction was washed with 0.1 M HCl (20 mL x 1) and the organic layer was dried, filtered, and concentrated *in vacuo*. The dried product (20.8 mmol, 1 eq.) and NAC (2.2 mL, 20.8 mmol, 1 eq.) was then refluxed in dry toluene (50 mL) for 5 hours. The

final product was extracted with ethyl acetate (150 mL x 3). The organic layer was dried with MgSO<sub>4</sub>, and the product was isolated *in vacuo*.

3-oxoheptanoyl-N-acetylcysteamine thioester (3.5) <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  5.89 (bs, 1H), 3.70 (s, 2H), 3.45 (q, J = 6.3, 2H), 3.10 (t, J = 6.5, 2H), 2.54 (t, J = 7.4, 2H), 1.98 (s, 3H), 1.58 (m, J = 7.6, 2H), 1.32 (m, J = 7.3, 2H), 0.91 (t, J = 7.3, 3H).

3-oxooctanoyl-N-acetylcysteamine thioester (**3.6**) <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 5.89 (bs, 1H), 3.70 (s, 2H), 3.47 (q, J = 6.4, 2H), 3.10 (t, J = 6.6, 2H), 2.53 (t, J = 7.3, 2H), 1.98 (s, 3H), 1.60 (m, J = 7.4, 2H), 1.26-1.35 (m, 4H), 0.89 (t, J = 7.2, 3H).

## 4-pentynoyl-S-NAC

Into a round-bottom flask containing anhydrous dicholoromethane cooled (50 mL) to 0 °C was dissolved 4-pentynoic acid (50 mg, 0.51 mmol, 1.0 eq), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (146.7 mg, 0.77 mmol, 1.5 eq.), 4-dimethylaminopyridine (12.5 mg, 0.10 mmol, 0.2 eq.) and NAC (81.2 μL, 0.77 mmol, 1.5 eq.). The reaction was stirred at 0 °C for 30 minutes, warmed to room temperature, and stirred for an additional 16 hours. The reaction was quenched using deionized water (50 mL), and the resulting mixture was extracted with ethyl acetate (150 mL x 3). The organic layer was dried with MgSO<sub>4</sub>, and the product was isolated *in vacuo*.

4-pentynoyl-*N*-acetylcysteamine thioester (3.14) <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  5.79 (bs, 1H), 3.45 (q, J = 6.5, 2H), 3.07 (t, J = 6.6, 2H), 2.82 (t, J = 7.2, 2H), 2.55 (td, J = 2.6, 7.2, 2H), 2.00 (t, J = 2.7, 1H), 1.97 (s, 3H).

# acyl-S-NAC monoketides

Into a round-bottom flask containing anhydrous dicholoromethane (8.5 mL) was added TEA (70  $\mu$ L, 0.50 mmol, 1.2 eq.), NAC (45  $\mu$ L, 0.42 mmol, 1 eq.), and either 37  $\mu$ L propionyl chloride, 44  $\mu$ L butyryl chloride, 50  $\mu$ L valeryl chloride, or 59  $\mu$ L hexanoyl

chloride (0.42 mmol, 1 eq.). The reaction was stirred on ice for 30 minutes, warmed up to room temperature, and stirred for an additional hour. The reaction was quenched using deionized water (20 mL), and the resulting mixture was extracted with ethyl acetate (30 mL x 3). The organic layer was dried with MgSO<sub>4</sub>, and the product was isolated *in vacuo*.

propionyl-*N*-acetylcysteamine thioester (**3.15**) <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  5.86 (bs, 1H), 3.44 (q, J = 6.3, 2H), 3.03 (t, J = 6.6, 2H), 2.61 (t, J = 7.5, 2H), 1.97 (s, 3H), 1.19 (t, J = 7.5, 3H).

butyryl-*N*-acetylcysteamine thioester (**3.16**) <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  5.80 (bs, 1H), 3.44 (q, J = 6.5, 2H), 3.03 (t, J = 6.2, 2H), 2.56 (t, J = 7.6, 2H), 1.96 (s, 3H), 1.70 (m, J = 7.4, 2H), 0.96 (t, J = 7.4, 3H).

valeryl-*N*-acetylcysteamine thioester (**3.17**) <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  5.82 (bs, 1H), 3.43 (q, J = 6.2, 2H), 3.02 (t, J = 6.4, 2H), 2.58 (t, J = 7.7, 2H), 1.96 (s, 3H), 1.64 (m, J = 7.5, 2H), 1.35 (m, J = 7.5, 2H), 0.91 (t, J = 7.3, 3H).

hexanoyl-*N*-acetylcysteamine thioester (**3.18**) <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  5.91 (bs, 1H), 3.42 (q, J = 6.8, 2H), 3.01 (t, J = 6.3, 2H), 2.56 (t, J = 7.3, 2H), 1.95 (s, 3H), 1.65 (m, J = 6.9, 2H), 1.30 (m, J = 3.8, 4H), 0.88 (t, J = 7.2, 3H).

# Cloning

Restriction enzyme sites within the EryMod6TE plasmid were utilized for the construction of EryMod6, EryMod6AmpTE, EryMod6CdaTE, and EryMod6RadTE (Scheme 3-3, adapted from Menzella, 2005).

# Protein expression and purification

MatB and EryMod6TE expression plasmids were prepared as previously described (Hughes, 2011). The MatB expression plasmid was transformed into *Escherichia coli* BL21(DE3) while EryMod6, EryMod6TE, EryMod6AmpTE,

EryMod6CdaTE, and EryMod6RadTE expression plasmids were transformed into *E. coli* K207-3 cells (Murli, 2003). Starter cultures (50 mL) were grown overnight and used to inoculate Luria broth (2 L, 37 °C) supplemented with 50 mg/L kanamycin. The culture was cooled (15° C) when cells reached OD600 = 0.4, and protein expression was induced with 0.5 mM IPTG. After 16 hours, cells were harvested by centrifugation, resuspended in 50 mL lysis buffer (10% glycerol, 0.5 M NaCl, 100 mM HEPES pH 7.5), and sonicated. Cellular debris was removed via centrifugation (30,000 x g, 45 minutes). Both MatB and the five EryMod6TE lysates were dialyzed against 1 L of lysis buffer (4 °C). After 6 hours, the buffer was exchanged and a second round of dialysis was performed for an additional 16 hours of dialysis. Lysates were flash-frozen and stored at (-80) °C until further use.

# Methylmalonyl-S-NAC MatB reaction

The following reaction conditions were used for the enzymatic synthesis of methylmalonyl-S-NAC extender unit **3.1**: 300 mM HEPES pH 7.5, 120 mM adenosine triphosphate, 240 mM MgCl<sub>2</sub>, 100 mM methylmalonic acid, 110 mM NAC, and 10% v/v MatB (in dialyzed lysate). The reaction was incubated at 23 °C overnight. Synthesis of **3.1** was verified by reverse phase HPLC (Hughes et al., 2011). The reaction was flash-frozen and stored at (-80) °C until further use.

# Enzymatic synthesis of triketide pyrones by EryMod6TE

Reactions were set up using the following conditions (4 mL total volume): 5% v/v DMSO, 300 mM HEPES pH 7.5, 10 mM 2-7, 25 mM MatB-generated **3.1**, and 25% v/v EryMod6TE (in dialyzed lysate). After 6 hours at 23 °C, reactions were spiked with additional 10 mM **3.2–3.7** and incubated at 23 °C overnight. Triketide pyrone products were extracted using ethyl acetate (8 mL x 3) and isolated in vacuo. The pyrone product

was purified via flash chromatography using silica gel (with a 0.5 cm top layer of copper sulfate-impregnated silica gel to remove unreacted NAC) and 50% ethyl acetate in hexanes. Fractions were analyzed by TLC, the triketide pyrone products made visible through iodine staining. Product-containing fractions were combined and evaporated. Samples were resuspended in 50 μL pure ethanol and further purified by reverse phase HPLC (Waters) using a Phenomenex Bondclone 10-C<sub>18</sub> column and a Waters 2998 photodiode array detector set at 290 nm. The mobile phases consisted of water containing 0.1% TFA (solvent A) and methanol containing 0.1% TFA (solvent B), and a linear gradient (flow rate = 2 mL/min) of 20-80% B over 10 minutes was used. Triketide pyrone peaks were collected, isolated in vacuo, and submitted for <sup>1</sup>H NMR characterization (400 MHz, CDCl<sub>3</sub>, pulse width = 90°, number of scans = 64). The molecular weights of compounds 3.8–3.13 were verified using LC/MS and high-resolution mass spectrometry (HR-MS).

HR-MS Data

**3.8**: ESI; [M + H]+ calc. m/z = 141.05462, obs. m/z = 141.05435

**3.9**: ESI; [M + H]+ calc. m/z = 155.07027, obs. m/z = 155.07007

**3.10**: ESI; [M + H]+ calc. m/z = 169.08592, obs. m/z = 169.08583

**3.11**: ESI; [M + H]+ calc. m/z = 169.08592, obs. m/z = 169.08599

**3.12**: ESI; [M + H]+ calc. m/z = 183.10157, obs. m/z = 183.10136

**3.13**: ESI; [M + H]+ calc. m/z = 197.11722, obs. m/z = 197.11698

## Scaled-up synthesis of triketide pyrone 3.11

A scaled-up synthesis of pyrone **3.11** was set up using the following conditions (50 mL total volume): 5% v/v DMSO, 200 mM HEPES pH 7.5, 21 mM **3.4** (200 mg, 1.0 mmol), 25 mM MatB-generated **3.1** (25 mL, 1.25 mmol), and 25% v/v EryMod6TE (in

12.5 mL dialyzed lysate). After 16 hours, 100 µL of the reaction was quenched with 100 μL pure ethanol, and then vortexed, centrifuged (10,000 x g, 1 min), and analyzed by HPLC as outlined in the previous section. A significant quantity of diketide 3.4 was still unreacted, so the reaction was spiked with an additional 2 mL of dialyzed EryMod6TE lysate and an additional 0.2 mmol of MatB-generated 3.1. The reaction was incubated for an additional 16 hours and analyzed in the same fashion as described above. Only trace amounts of compound 3.4 remained. Triketide pyrone 3.11 was extracted with ethyl acetate (150 mL x 3), dried with MgSO<sub>4</sub>, and isolated in vacuo. The pyrone was further purified by silica gel flash chromatography (with a 2 cm top layer of copper sulfateimpregnated silica gel) using 50% ethyl acetate in hexanes. Fractions were analyzed by TLC (iodine stain). Pyrone-containing fractions were combined and evaporated. The residue was dried on high-vacuum for 3 hours and the quantity of pyrone product produced was determined to be 20 mg, which represents a yield of 11.9%. Pyrone 3.11 was characterized both by <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>, pulse width = 30°, number of scans = 8) as well as <sup>13</sup>C NMR (400 MHz, CDCl<sub>3</sub>, pulse width = 30°, number of scans = 64).

# Enzymatic synthesis of triketide pyrones via "stuttering"

Reactions to produce triketide pyrones through the stuttering of EryMod6TE were set up using the following conditions (250 µL total volume): 5% v/v DMSO, 300 mM HEPES 7.5, 10 mM **3.14–3.18**, 40 mM MatB-generated **3.1**, and 25% v/v EryMod6TE (in dialyzed lysate). After 6 hours at 23 °C, reactions were spiked with additional 10 mM **3.14–3.18** and incubated at 23 °C for an additional 18 hours. Triketide pyrones were extracted with ethyl acetate (750 µL x 3) and isolated *in vacuo*. Products were resuspended in ethanol and analyzed via reversed-phase HPLC (Waters) using a Varian

Microsorb-MV 100-5  $C_{18}$  column and a Waters 2998 photodiode array detector set at 290 nm. The mobile phases consisted of water containing 0.1% TFA (solvent A) and methanol containing 0.1% TFA (solvent B). A linear gradient (flow rate = 1 mL/min) of 20-80% B over 10 minutes was employed. The molecular weights of **3.19–3.22** were characterized by LC/MS and high-resolution mass spectrometry. The <sup>1</sup>H NMR of pyrone **3.20** was determined (400 MHz, CDCl<sub>3</sub>, pulse width = 90°, number of scans = 64).

**HR-MS** Data

**3.19:** ESI; [M + H]+ calc. m/z = 193.08592, obs. m/z = 193.08561

**3.20**: ESI;  $[M + H]_{+}$  calc. m/z = 183.10154, obs. m/z = 183.10157

**3.21**: ESI; [M + H]+ calc. m/z = 197.11722, obs. m/z = 197.11716

**3.22**: ESI; [M + H]+ calc. m/z = 211.13287, obs. m/z = 211.13261

# **TE-catalyzed hydrolysis reactions**

The following reaction conditions were used to monitor hydrolysis by EryMod6TE (250 μL total volume): 5% v/v DMSO, 300 mM HEPES pH 7.5, 10 mM 3.4, 3.7, 3.15, or 3.16, and 25% v/v EryMod6TE (in dialyzed lysate). At time points 0, 2, 4, 6, and 8 hours, 25 μL of the reaction was quenched with 25 μL of pure ethanol, vortexed, and centrifuged (10,000 rcf, 5 min). Time point samples (10 μL) were analyzed by reversed-phase HPLC (Waters) using a Varian Microsorb-MV 100-5 C<sub>18</sub> column and a Waters 2998 photodiode array detector set at 235 nm. The mobile phases consisted of water containing 0.1% TFA (solvent A) and methanol containing 0.1% TFA (solvent B), and a linear gradient (flow rate = 1 mL/min) of 20-80% B over 10 minutes was used. Peaks were integrated using Waters software. All peak areas for time trials were normalized to peak area at time = 0 hr. Each reaction was set up in triplicate, and the entire experiment was duplicated.

## **Activity assay of the five EryMod6TE variants**

Reactions to assay the five EryMod6TE variants for pyrone formation were set up with the following conditions (500 μL total volume): 300 mM HEPES pH 7.5, 10 mM of diketide **3.4**, 20 mM MatB-generated **3.1**, and 25% v/v EryMod6AmpTE (in dialyzed lysate), or EryMod6RadTE (in dialyzed lysate), EryMod6CdaTE (in dialyzed lysate), EryMod6NoTE (in dialyzed lysate), or EryMod6TE (in dialyzed lysate). Reactions were incubated at 23 °C overnight. After 24 hours, 100 μL from each reaction was quenched with 100 μL pure ethanol, vortexed, and centrifuged (10,000 x g, 1 minute). 20 μL samples were analyzed by reversed-phase HPLC (Waters) using a Varian Microsorb-MV 100-5 C<sub>18</sub> column, monitoring at both 235 nm (to detect diketide **3.4**) and 290 nm (to detect triketide pyrone **3.11**) with a Waters 2998 photodiode array detector. The mobile phases consisted of water containing 0.1% TFA (solvent A) and methanol containing 0.1% TFA (solvent B). A linear gradient (flow rate = 1 mL/min) of 20-80% B over 10 minutes was used. Triketide pyrone peak areas were integrated using Waters software.

## LC/MS analysis

Pyrones 3.8–3.13 and 3.19–3.22 were resuspended in 25% acetonitrile and 0.1% formic acid and subjected to positive-ESI LC/MS (Agilent Technologies 1200 Series HPLC with a Gemini  $C_{18}$  column coupled to an Agilent Technologies 6130 quadrupole mass spectrometer). Mobile phases consisted of water containing 0.1% formic acid (solvent A) and acetonitrile containing 0.1% formic acid (solvent B). A linear gradient (flow rate = 0.7 mL/min) of 5-95% B over 12 minutes was employed.

**Scheme 3-1** Semipreparative synthesis of triketide pyrones from diketides and methylmalonyl-*S*-NAC.

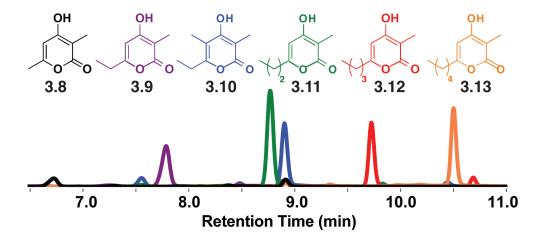


Figure 3-1 Enzymatic synthesis of triketide pyrones from diketides.

Reversed-phase chromatography traces of triketide pyrones synthesized by EryMod6TE from  $\beta$ -ketoacyl-S-NACs.

**Table 3-1** <sup>1</sup>H NMR and LC/MS characterization of triketide pyrones.

Cmpd.	$R_1$	$\mathbf{R}_2$	C-2 methyl	Mass: (exp.) obs.
3.8	2.21 (3H, s)	5.82 (1H, s)	1.96 (3H, s)	(141.0546) 141.0544
3.9	2.50 (2H, q, J = 7.6) 1.22 (3H, t, J = 7.6)	5.81 (1H, s)	1.96 (3H, s)	(155.0703) 155.0701
3.10	2.55 (2H, q, J = 7.2) 1.21 (3H, t, J = 7.6)	1.98 (3H, s)	1.97 (3H, s)	(169.0859) 169.0858
3.11	2.43 (2H, t, J = 7.5) 1.68 (2H, m, J = 7.5) 0.95 (3H, t, J = 7.4)	5.83 (1H, s)	1.96 (3H, s)	(169.0859) 169.0860
3.12	2.45 (2H, q, J = 7.2) 1.63 (2H, m, J = 7.2) 1.36 (2H, m, J = 7.2) 0.92 (3H, t, J= 7.4)	5.84 (1H, s)	1.96 (3H, s)	(183.1016) 183.1014
3.13	2.44 (2H, t, J = 7.6) 1.64 (2H, m, J = 7.7) 1.32 (4H, m, J = 7.2) 0.90 (3H, t, J= 7.0)	5.83 (1H, s)	1.96 (3H, s)	(197.1172) 197.1170
3.20	2.49 (2H, t, J = 7.7) 1.68 (2H, m, J = 7.6) 0.95 (3H, t, J = 7.4)	1.98 (3H, s)	1.96 (3H, s)	(183.1015) 183.1016

**Scheme 3-2** Semipreparative synthesis of triketide pyrones from monoketides and methylmalonyl-*S*-NAC *via* stuttering.

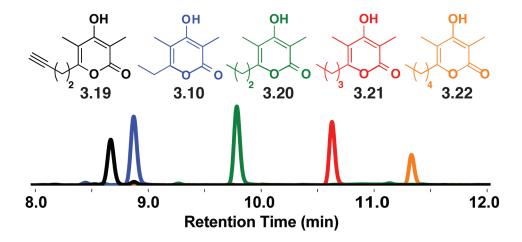


Figure 3-2 Enzymatic synthesis of triketide pyrones from monoketides.

Reversed-phase chromatography traces of triketide pyrones synthesized by EryMod6TE from acyl-S-NACs *via* stuttering.

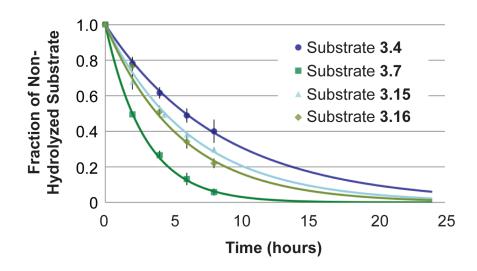


Figure 3-3 TE-mediated hydrolysis of priming units 3.4, 3.7, 3.15 and 3.16.

Substrates were incubated with EryMod6EryTE, and TE-mediated hydrolysis was monitored by HPLC (peak integrations are normalized to 1 at time = 0 hours).

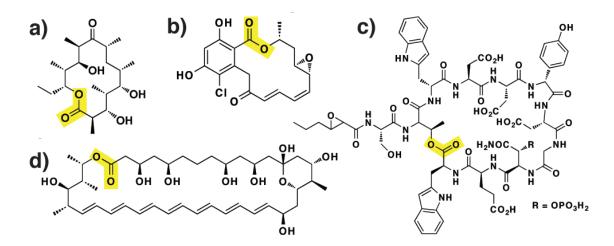
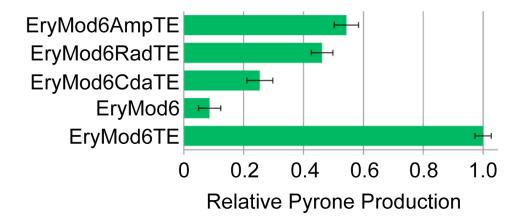


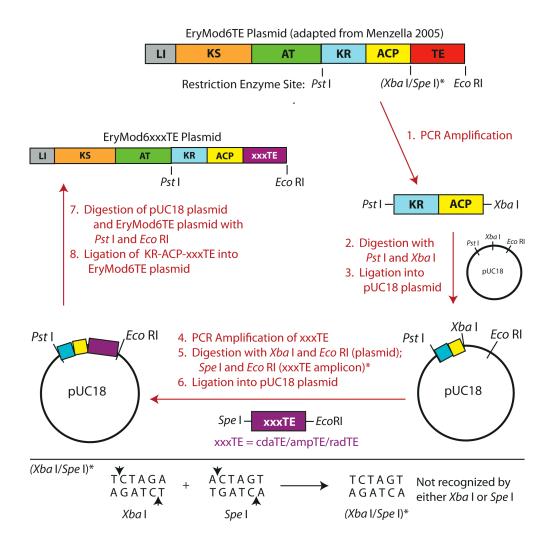
Figure 3-4 Polyketides and TE cyclization.

Molecules cyclized by TEs of a) the erythromycin PKS, b) the radicicol PKS, c) the calcium-dependent antibiotic NRPS, and d) the amphotericin PKS. The esters formed through cyclization are highlighted in yellow.



**Figure 3-5** Domain swapping of EryTE leads to the formation of triketide pyrones.

Enzymatic synthesis of pyrone **3.11** by four EryMod6TEs. Quantities were normalized relative to the pyrone produced by EryMod6TE.



**Scheme 3-3** Engineering EryMod6TE plasmid constructs for domain swapping experiments.

# Chapter 4 Investigating the Reactivities of a Polyketide Synthase Module through Fluorescent Click Chemistry

#### **ABSTRACT**

A method using fluorescent click chemistry has been developed to rapidly, sensitively, and inexpensively quantify products generated by polyketide synthase modules. The utility of this method was demonstrated through experiments designed to investigate the reactivity of a module to various substrates and reaction conditions. In a system where the terminal module and thioesterase of the erythromycin synthase generates reduced diketides, ethanethiol was found to be as effective as N-acetylcysteamine in delivering both priming units and extender units. Insights into the competing reactivities within the module-thioesterase were revealed through triketide lactone products generated from a  $\beta$ -ketoacyl ethanethioester substrate. The addition of phosphate increased enzymatic turnover, while the addition of glycerol resulted in thioesterase-mediated glycerolysis of priming units.

#### Introduction

There are many challenges associated with investigating the operation of PKS modules, let alone employing them for *in vitro* synthesize desired compounds: substrates such as NADPH and methylmalonyl-CoA are costly, modules are often difficult to express and purify, and the products generated are typically nonchromophoric. Traditionally, nanogram to microgram quantities of product are generated from radiolabeled precursors and visualized by autoradiography - thin layer chromatography (radio-TLC) (Menzella, 2005). While GC/MS has been employed, derivitization is sometimes necessary and signal-to-noise can be low (Valenzano, 2009). Recently, advances in biocatalytic methods have aided in scaling up module-catalyzed reactions – a

glucose-fueled NADPH-regeneration system can provide a constant supply of reduced nicotinamide coenzyme, and extender units like methylmalonyl-*S-N*-acetylcysteamine (mm-*S*-NAC) can be enzymatically generated by MatB for the price of ATP (Wong, 1985; Hughes, 2011). However, the major bottleneck to investigating the reactivity of modules remains analyzing *in vitro* reactions.

Previously, we used fluorine NMR to follow substrates containing a terminal trifluoromethyl group through reactions catalyzed by overexpressed PKS enzymes in the commonly-employed biocatalytic setting of dialyzed cell lysate (Piasecki, 2012). This method requires both sophisticated instrumentation and knowledge of the chemical shifts of generated products. We sought a more general bioorthogonal approach that would allow visualization of polyketide products by HPLC. Since fluorescent alkyne probes and click chemistry have been used to locate azide-containing glycolipids in the background of a zebrafish (Laughlin, 2008), we reasoned that polyketide products containing a unique chemical functionality could be selectively labeled by a complementary fluorophore in the background of bacterial cell lysate.

Here we showcase a quantitative detection method that reports the fate of alkynyl priming units in module-catalyzed reactions within cell lysate (Figure 4-1). Copper(I)-catalyzed azide alkyne cycloaddition (CuAAC) with sulforhodamine B azide enables normally nonchromophoric polyketide products to be sensitively visualized and identified using fluorescence spectroscopy and mass spectrometry (Rostovtsev, 2002; Prasad, 2012). Employing this method, the incorporation of ethanethiol-linked priming and extender units into reduced diketide compounds by the terminal module–TE of the erythromycin PKS (EryMod6TE) was investigated. Ethanethiol handles were found to be as effective as *N*-acetylcysteamine (NAC) handles. Competing reactivities of EryMod6TE were observed through the conversion of a β-ketoacyl ethanethioester

priming unit to triketide lactones. We also noted an increase in enzymatic turnover in a single-module catalyzed reaction with increased phosphate concentrations and thioesterase-mediated glycerolysis of priming units in reaction buffers containing glycerol.

#### RESULTS

### Bioorthogonal detection of polyketide products in cell lysate

All experiments reported here investigated the reactivity of EryMod6TE within *E. coli* cell lysate. Performing reactions in cell lysate saves the time and expense associated with protein purification. To obtain EryMod6TE-containing lysate, overexpressing cells are pelleted, resuspended, and sonicated before cell debris is removed by centrifugation. Subsequent dialysis places EryMod6TE in an appropriate reaction buffer and removes metabolites. EryMod6TE is incubated with synthetic priming and extender units before initiating CuAAC with sulforhodamine B azide. Resulting adducts are analyzed by HPLC, and LC/MS is performed as necessary to assign masses to peaks. In the experiments reported here, a fluorimeter was employed for optimal sensitivity, although UV/vis detection also provides excellent signal.

#### An ethanethiol-linked extender unit

To decrease costs associated with both investigating the reactivities of modules and biocatalytically producing polyketides, the use of a simplified methylmalonyl-CoA analog as an *in vitro* extender unit was investigated (Figure 4-2). Previous *in vitro* reactions have employed enzymatically- or synthetically-generated **4.1** (mm-S-NAC) (Hughes, 2011; Pohl, 1998); however, the one-step, multigram synthesis of methylmalonyl ethanethioester (mm-S-Et) **4.2** is much more economical (Figure 4-2a).

Thus, EryMod6TE was incubated with priming unit **4.3**, extender unit **4.1** or **4.2**, and the NADPH regeneration system to produce reduced diketide **4.5** (Figure 4-2b; Appendix 2). After 16 hrs, reactions were subjected to CuAAC with sulforhodamine B azide and analyzed by HPLC (Figures 4-2c-d; Appendix 2).

The analysis revealed approximately equal quantities of **4.5\*** (asterisks denote sulforhodamine adducts) were generated in reactions employing **4.1** or **4.2**; ~35% of the priming unit **4.3** was accepted, extended, reduced, and hydrolyzed by EryMod6TE to afford reduced diketide **4.5.** Analysis also showed that ~50% of priming unit **4.3** was hydrolyzed to carboxylic acid **4.4** (Figures 4-2c-d). TE-catalyzed hydrolysis of priming units has been previously implicated in the low yields (~10%) of EryMod6TE-catalyzed reactions (Hughes, 2012; Harper, 2012). Remaining **4.3** from the overnight incubations may not have reacted due to the gradual loss of EryMod6TE activity - precipitate was noticeable in reactions after 6 hrs.

To demonstrate the use of extender unit **4.2** in the preparative, biocatalytic syntheses of polyketides, 17 mg of **4.3** and 13 mg of **4.2** were incubated with EryMod6TE lysate and the NADPH regeneration system. After 16 hrs, the reaction was acidified, extracted with ethyl acetate, and chromatographed to afford 3 mg of reduced diketide **4.5** (22% yield; Appendix 2). The difference between analytical and preparative reaction yields is likely due to losses during purification.

### **Ethanethiol-linked priming units**

Since the NAC handle of extender unit **4.1** could be replaced by ethanethiol, we investigated replacing the NAC handle of priming unit **4.3** with ethanethiol. Thus, EryMod6TE was incubated with the NADPH regeneration system, extender unit **4.2**, and either NAC-linked **4.3** or ethanethiol-linked **4.9** (Figures 4-3a-b). After 16 hrs reactions

were subjected to CuAAC with sulforhodamine B azide and analyzed by HPLC (Figure 4-3c; Appendix 2). A comparison of the amount of **4.5**\* formed from priming units **4.3** and **4.9** revealed that EryMod6TE did not significantly discriminate between NAC or ethanethiol priming unit handles. Building on these results, the promiscuity of EryMod6TE towards NAC- and ethanethiol-linked priming units of varying chain length was assayed. The analysis showed less than 5% difference in product formation between reduced diketide products **4.13**\* and **4.14**\* (Figure 4-3d). Thus EryMod6TE does not significantly discriminate between priming units based on chain length or handle.

## Reactivity of EryMod6TE towards a β-ketoacyl ethanethioester

Having analyzed reactions employing priming units **4.3** and **4.6-4.10** in the module-catalyzed generation of reduced diketides, we were interested in probing the reactivity EryMod6TE with a more advanced priming unit, such as β-ketoacyl ethanethioester **4.15**. The chemical synthesis of **4.15** was performed by *C*-acylation of 5-hexynoic acid with **4.2** through a reaction developed by Masamune and coworkers (Brooks, 1979). Briefly, **4.2** was reacted with magnesium ethoxide to form its magnesium salt, 5-hexynoic acid was reacted with carbonyldiimidizole to form its imidazolide, and these intermediates were reacted with one another. EryMod6TE was incubated for 16 hrs with priming unit **4.15**, extender unit **4.2**, and the NADPH regeneration system. The reaction was subjected to CuAAC with sulforhodamine B azide and analyzed by HPLC (Figures 4-4a–c; Appendix 2).

Interestingly, **4.15** was reduced by EryKR6 yielding at least two reduced stereoisomers (Figures 4-4a-c). EryKS6 accepted at least one stereoisomer of **4.17** since extension followed by cyclization yielded ketolactone **4.19** while extension followed by

another round of EryKR6-mediated reduction prior to cyclization yielded hydroxylactone **4.20**.

### Phosphate increases substrate turnover

Increased turnover numbers for triketide lactone formation by the bimodular DEBS1+TE (the first polypeptide of the erythromycin PKS fused to EryTE) were observed with increases in phosphate concentration (Pieper, 1996). To determine whether phosphate has a similar effect on single-module constructs, EryMod6TE was incubated with 0–100 mM phosphate, priming unit **4.3**, extender unit **4.2**, and the NADPH regeneration system (Figure 4-5a). After 2 and 16 hrs, reactions were subjected to CuAAC with sulforhodamine B azide and analyzed by HPLC (Figures 4-5b–c).

After 2 hours, 20% of **4.3** was converted to **4.5** in the absence of phosphate, while 31% had been converted in the presence of 100 mM phosphate (Figure 4-5b). After 2 hrs, 43% of **4.3** remained unreacted in the absence of phosphate, whereas only 20% remained unreacted in the presence of 100 mM phosphate. After 16 hrs, no **4.3** remained and no precipitate was observed in reactions containing at least 50 mM phosphate (Figure 4-5c).

### **TE-catalyzed glycerolysis**

Previously, glycerol has been used to prevent precipitation of EryMod6TE in dialyzed lysate (Hughes, 2012; Harper, 2012); however, HPLC analysis of EryMod6TE-catalyzed reactions in the presence of glycerol yielded a large, unanticipated peak (4.21\*, Figure 4-6b). LC/MS and <sup>1</sup>H NMR analysis confirmed that this peak resulted from priming unit glycerolysis (Figure 4-6a; Appendix 2). A panel of single and double knockout mutants of EryMod6TE was assayed for the ability to generate 4.21 (ΔKS, ΔAT, ΔTE, ΔKS+AT, ΔKS+TE, and ΔAT+TE; catalytic nucleophiles replaced by alanines). Through these experiments, TE activity was linked to the generation of 4.21

(Figure 4-6c; Appendix 2). TE-catalyzed glycerolysis was confirmed when **4.21\*** was observed from the incubation of purified EryTE with **4.3** in a glycerol-containing buffer (Appendix 2).

#### **DISCUSSION**

A method for monitoring *in vitro* polyketide synthesis has been developed with several advantages over traditional techniques such as radio-TLC and GC/MS. Polyketide products, which are typically nonchromophoric, are made brightly fluorescent in a simple, rapid, inexpensive, and bioorthogonal manner upon CuAAC with sulforhodamine B azide. The terminal alkyne moiety is easily incorporated into synthetic priming units, and based on the studies reported here, is well-tolerated by PKS enzymes.

We employed this method to compare the reaction profiles of EryMod6TE either incubated with ethanethiol-linked priming and extender units or with the equivalent NAC-linked substrates. Since no significant difference exists between these handles in their ability to deliver substrates *in vitro*, those investigating the reactivities of modules or using them biocatalytically may benefit from employing the ethanethiol handle.

Several insights into the reactivity of EryMod6TE were made from its incubation with the  $\beta$ -ketoacyl ethanethioester priming unit **4.15.** Even though this substrate possesses an ethanethiol handle, it was reduced to both *syn-* and *anti-* products by EryKR6 (this KR also exhibits little stereocontrol towards  $\alpha$ -methyl,  $\beta$ -ketopentanoyl-S-NAC) (Siskos, 2005). However, at least one of the reduced  $\alpha$ -methyl,  $\beta$ -hydroxyacyl stereoisomers is extended, as evidenced by the significant production of ketolactone **4.19** and hydroxylactone **4.20**. The relative quantities of **4.19** and **4.20** may indicate that EryKR6 kinetically outcompeted EryTE for the ACP-bound intermediate, as observed in

preparative reactions with substrates not containing a terminal alkyne (Harper, 2012). Based on previous studies in which  $\beta$ -ketoacyl-S-NAC substrates were converted into pyrones by EryMod6TE (Hughes, 2012), some pyrone formation was anticipated from **4.15**; however, pyrone was not generated at a detectable level. This suggests that the  $\alpha$ -methyl,  $\beta$ -ketoacyl group is infrequently extended by EryKS6.

Reaction conditions were investigated using the described method. As reported for DEBS1+TE, phosphate had a beneficial effect on the activity of EryMod6TE. In the presence of 50 mM phosphate, the priming unit substrate was completely turned over by EryMod6TE, apparently owing to enhanced enzyme stability. Reaction conditions that include glycerol should be used with caution since glycerol was found to attack acyl-TE intermediates, effecting the glycerolysis of the priming unit substrate. To our knowledge, this is the first time TE-mediated glycerolysis activity has been reported.

TE-catalyzed hydrolysis was responsible for the largest loss of the priming unit substrate (>50%). Before the PKS module-catalyzed synthesis of new compounds becomes routine, reaction yields must be increased. Ideally, priming units would not be substrates for TE hydrolysis. In recent biocatalytic studies employing modular constructs from the pikromycin PKS, greater than 60% yields were obtained in experiments employing thiophenol-linked priming units (Hansen, 2013). We are now investigating a panel of alkyne-containing priming units that display decreased TE-mediated hydrolysis while remaining substrates for KSs.

The described method will help to answer many fundamental questions of singleand bimodular systems *in vitro*. Through its bioorthogonal nature, it may also be useful in the quantitation of polyketide intermediates and products *in vivo*. The precursor-directed biosynthesis of 15-propargyl erythromycin A was accomplished using a priming unit containing a terminal alkyne (Harvey, 2012), and directed evolution by Schulz and coworkers led to the incorporation of a terminal alkyne-containing extender unit by EryAT6 (Sundermann, 2012). Applying the described method to such systems would be very informative, especially for engineered versions of the erythromycin synthase that do not produce products expected of them. In general, we anticipate the described method to greatly accelerate the investigation and optimization of PKS-catalyzed biosynthesis of complex polyketide products.

#### MATERIALS AND METHODS

### Synthesis of compounds utilized

## Methylmalonyl-N-acetylcysteamine thioester

*Mono-tert-butyl methylmalonic acid*. To methyl meldrum's acid (8.5 g, 138 mmol) was added *tert*-butanol (40 mL) and pyridine (2 mL). The mixture was heated at 80 °C for 24 hrs. The reaction was then cooled to 23 °C and concentrated *in vacuo* to furnish mono-*tert*-butyl methylmalonic acid.

Tert-butyl-methylmalonyl-N-acetylcysteamine. To a stirred solution of 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (EDCI; 3.9 g, 21 mmol, 1.1 eq) in dichloromethane (DCM, 60 mL) at 23 °C was added triethylamine (TEA; 2.9 mL, 21 mmol, 1.1 eq). To the yellow solution was added the mono-tert-butyl methylmalonic acid (3.0 g, 19 mmol, 1.0 eq) followed by N-acetylcysteamine (NAC; 2.0 mL, 21 mmol, 1.1 eq). The resulting solution was stirred for 48 hrs, then diluted with ethyl acetate (50 mL) and washed with 0.1 M HCl (4 x 10mL). The organic layer was dried over sodium sulfate and concentrated *in vacuo*.

Methylmalonyl-S-N-acetylcysteamine. To a stirred solution of tert-butyl-methylmalonyl-S-NAC (500 mg, 1.8 mmol, 1.0 eq) in DCM (9 mL) at 0 °C under argon

was added 4.0 M HCl in dioxane (4.5 mL, 18 mmol, 10.0 eq). After 18 hrs, the reaction was concentrated *in vacuo* and purified by dry flash column chromatography (silica, eluting with 10:1 DCM:methanol) to afford compound 1. <sup>1</sup>H NMR characterization was consistent with a previously reported synthesis (Pohl 1998).

methylmalonyl-*N*-acetylcysteamine thioester (mm-*S*-NAC, **4.1**) <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  3.71 (q, 1H, J = 7.2), 3.48 (q, 2H, J = 6.0), 3.10 (dtd, 2H, J = 47.6, 14.2, 6.4), 1.98 (s, 3H), 1.45 (d, 3H, J = 7.2)

### Methylmalonyl-ethanethiol thioester

To a round-bottom flask charged with methylmalonic acid (10 g, 5 mmol, 1.0 eq), was added 4-dimethylaminopyridine (DMAP; 1 g, 8.5 mmol, 0.1 eq) followed by toluene (250 mL). Upon stirring, phosphoryl chloride (POCl<sub>3</sub>; 8.8 mL, 93 mmol, 1.1 eq) was added followed by the dropwise addition of ethanethiol (6.9 mL, 93 mmol, 1.1 eq). The reaction was heated to 60 °C under an argon atmosphere. After 18 hrs, the reaction was cooled to 23 °C and concentrated *in vacuo*. The residue was suspended in DCM, and purified via dry flash column chromatography (silica, eluting with an ethyl acetate:hexanes gradient from 1:9 to 1:3) to furnish compound 2.

Methylmalonyl ethanethiol thioester (4.2) <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  3.67 (q, 1H, J = 7.1), 2.94 (q, 2H, J = 7.4), 1.47 (d, 3H, J = 7.2), 1.28 (t, 3H, J = 7.4)

#### NAC- and ethanethiol-linked monoketides

To a round-bottom flask charged with 75 mL of anhydrous DCM cooled to 0 °C under an atmosphere of argon was dissolved EDCI (1.054 g, 5.5 mmol, 1.1 eq), TEA (773  $\mu$ L, 5.5 mmol, 1.1 eq), 4-pentynoic acid (6 and 8, 490  $\mu$ L, 5.0 mmol, 1.0 eq) or 5-hexynoic acid (3 and 9, 562  $\mu$ L, 5.0 mmol, 1.0 eq) or 6-heptynoic acid (7 and 10, 633  $\mu$ L, 5.0 mmol, 1.0 eq), and catalytic DMAP (307 mg, 2.5 mmol, 0.5 eq). The reaction was

stirred on ice for 10 min before the addition of NAC (3, 6–7, 504 μL, 4.75 mmol, 0.95 eq) or ethanethiol (8–10, 342.3 μL, 4.75 mmol, 0.95 eq) and was allowed to warm to room temperature. After 16 hrs, the reaction was concentrated *in vacuo* and the remaining oil was resuspended in 50 mL of water. The product was extracted with ethyl acetate (1 x 150 mL) and the organic layer was washed with 0.1 M HCl (2 x 50 mL) and saturated sodium bicarbonate (1 x 50 mL). The organic layer was dried over anhydrous sodium sulfate, and product was concentrated *in vacuo*. Products were purified by dry flash column chromatography (silica, eluting with 10:90 ethyl acetate:hexanes) to afford title compounds 3, 6–10. Products were characterized by ¹H NMR.

5-hexynoyl-S-N-acetylcysteamine (4.3) <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  5.80 (bs, 1H), 3.44 (q, 2H, J = 6.0), 3.04 (t, 2H, J = 6.6), 2.74 (t, 2H, J = 7.6), 2.26 (dt, 2H, J = 6.9, 2.6), 1.99 (t, 1H, J = 2.6), 1.97 (s, 3H), 1.89 (m, 2H, J = 7.7)

4-pentynolyl-S-N-acetylcysteamine (4.6) <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  5.78 (bs, 1H), 3.44 (q, 2H, J = 6.0), 3.06 (t, 2H, J = 6.6), 2.81 (t, 2H, J = 7.3), 2.54 (dt, 2H, J = 7.2 and 2.6), 1.99 (t, 1H, J = 2.6), 1.96 (s, 3H)

6-heptynoly-S-N-acetylcysteamine (4.7) <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  5.80 (bs, 1H), 3.43 (q, 2H, J = 6.1), 3.03 (t, 2H, J = 6.6), 2.61 (t, 2H, J = 7.3), 2.21 (dt, 2H, J = 7.1, 2.7), 1.97 (s, 3H), 1.96 (t, 1H, J = 2.7), 1.79 (m, 2H, J = 7.6), 1.60-1.53 (m, 2H, J = 7.1)

4-pentynoyl-S-N-ethanethiol thioester (4.8) <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  2.90 (q, 2H, J = 7.4), 2.77 (t, 2H, J = 7.2), 2.53 (dt, 2H, J = 6.8 and 2.6), 1.98 (t, 1H, J = 2.6), 1.25 (t, 3H, J = 7.4)

5-hexynoyl-S-ethanethiol (4.9) <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  2.88 (q, 2H, J = 7.4), 2.68 (t, 2H, J = 7.2), 2.25 (dt, 2H, J = 6.9 and 2.8), 1.98 (t, 1H, J = 2.7), 1.88 (m, 2H, J = 7.8), 1.25 (t, 3H, J = 7.4)

<u>6-heptynoyl-S-ethanethiol</u> (**4.10**) <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  2.87 (q, 2H, J = 7.4), 2.57 (t, 2H, J = 7.3), 2.21 (dt, 2H, J = 7.0, 2.6), 1.95 (t, 1H, J = 2.6), 1.79 (m, 2H, J = 7.8), 1.60-1.53 (m, 2H, J = 7.5), 1.25 (t, 3H, J = 7.4)

### $\beta$ -ketoacyl-S-ethanethiol diketide

To a round-bottom flask charged with a stir bar and **4.2** (50 μL, 0.37 mmol, 2.0 eq) under an atmosphere of argon was added anhydrous tetrahydrofuran (THF; 1 mL) and magnesium ethoxide (21.4 mg, 0.19 mmol, 1.0 eq). To a separate round-bottom flask charged with a stir bar and 5-hexynoic acid (63 μL, 0.56 mmol, 3.0 eq) under an atmosphere of argon was added anhydrous THF (3 mL) and carbonyldiimidazole (CDI; 90.8 mg, 0.56 mmol, 3.0 eq). Reactions were stirred at room temperature. After 1 hr, the reaction generating the magnesium salt of methylmalonyl ethanethioester was concentrated *in vacuo*. To the residue, the solution of the acyl imidazole was added, and the reaction was stirred under argon for an additional 16 hrs. The reaction was concentrated *in vacuo* and resuspended in 0.1 M HCl (10 mL) and ethyl acetate (50 mL). The organic layer was washed with 0.1 M HCl (1 x 30 mL) and then with saturated sodium chloride (1 x 10 mL). The organic layer was dried over anhydrous sodium sulfate, and product was isolated *in vacuo*. The product was purified by dry flash column chromatography (silica, eluting with 5:95 ethyl acetate:hexanes) to afford compound 15.

(2RS)-methyl-3-oxo-hept-6-ynoyl-S-ethanethiol (4.15) <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  3.74 (q, 2H, J = 7.1), 2.92 (q, 2H, J = 7.4), 2.79 (m, 2H, J = 27.6), 2.22 (m, 2H), 1.94 (t, 1H, J = 2.6), 1.38 (d, 3H, J = 7.1), 1.26 (t, 3H, J = 7.4)

### Sulforhodamine B azide

Synthesis of 2-azidoethanol followed previously described method (Lu, 2005). A 25 mL Schlenk flask was charged with 2-azidoethanol (17 mg, 0.19 mmol, 1.1 eq), TEA

(2.6 μL, 0.19 mmol, 1.1 eq) and DCM (5 mL). The resulting solution was cooled to 0 °C and allowed to stir for 10 min. A sulforhodamine B acid chloride solution in 5 mL DCM (100 mg, 0.17 mmol, 1.0 eq) was subsequently added dropwise via syringe. The resulting solution was stirred at 0 °C for 30 min and then allowed to stir at room temperature. Upon reacting for 12 hrs, the product was isolated *in vacuo* to afford a metallic green residue. The product was first to elute upon purification by dry flash column chromatography (silica, eluting with 90:10 chloroform:methanol).

### **Engineering EryMod6TE mutant plasmids**

Formation of mutant plasmids were engineering using an efficient one-step, site-directed mutagenesis protocol (Liu, 2008) of EryMod6TE plasmid with the following 6 primers (bold depicts overlapping region and alanine residue is underlined): (1) ΔKS C→A forward 5'-ACCTGTGCCGTCGGAGAAGGTCGGACCACCGGAATGTGA-GGCGAACACCG-3', (2) ΔKS C→A reverse 5'-GGCTTAGAAGGTCCGGCGGTAA-CCGGGACACGGCAGCCTCTTCCAGCC-3', (3) ΔAT S→A forward 5'-CGGCAA- TAGCCAGTACGAGTCCCGCTTTAACGCCGCCGGCAGCAC-3' (4) ΔAT S→A reverse 5'-CGTGTGGGGTGAGCCCGTCAGCCGTTATCGGTCAT-GCTCAGGGC-3', (5) ΔTE S→A forward 5'-CAGGCCACGCCGCCGCCGCCACT-CATGGCCTATGCACTCGCGAC-3', and (6) ΔTE S→A reverse 5'- CGTGGGAACC-ACTGTTTGGAAAGCACCATCGTCCGGTGCGGCGCCGCC3'.

### $\Delta KS$ , $\Delta AT$ , $\Delta TE$ Plasmids

PCR reaction conditions to engineer ΔKS, ΔAT and ΔTE plasmids was performed using the following conditions (50 mL total volume reactions): 1X Phusion GC buffer, 5 ng EryMod6TE plasmid, 1 mM each corresponding forward and reverse primers, 200 mM dNTPs, 3% v/v DMSO and 3 U Phusion DNA polymerase.

#### ∆KS+AT Plasmid

PCR reaction conditions to engineer  $\Delta KS+AT$  plasmid was performed using the following conditions (50 mL total volume reactions): 1X Phusion GC buffer, 5 ng  $\Delta AT$  plasmid, 1 mM each  $\Delta KS$  forward and reverse primers, 200 mM dNTPs, 3% v/v DMSO and 3 U Phusion DNA polymerase.

#### ∆KS+TE Plasmid

PCR reaction conditions to engineer ΔKS+TE plasmid was performed using the following conditions (50 mL total volume reactions): 1X Phusion GC buffer, 5 ng ΔTE plasmid, 1uM each ΔKS forward and reverse primers, 200 mM dNTPs, 3% v/v DMSOand 3 U Phusion DNA polymerase.

#### ∆AT+TE Plasmid

PCR reaction conditions to engineer  $\Delta$ AT+TE plasmid was performed using the following conditions (50 mL total volume reactions): 1X Phusion GC buffer, 5 ng  $\Delta$ TE plasmid, 1 mM each  $\Delta$ AT forward and reverse primers, 200 mM dNTPs, 3% v/v DMSO and 3 U Phusion DNA polymerase.

#### Protein expression, dialysis, and purification

The formation of EryMod6TE plasmid was described previously, and site-directed mutagenesis of EryMod6TE plasmid was performed to engineer mutant plasmids (Appendix 2) (Liu, 2008). EryMod6TE expression plasmids were transformed into  $E.\ coli$  K207-3 cells (Murli, 2003), while the glucose dehydrogenase (GDH) and the EryTE expression plasmids were each transformed into  $E.\ coli$  BL21(DE3) cells (Piasecki, 2012). Starter cultures (60 mL) were grown overnight and used to inoculate Luria broth supplemented with 50 mg/L kanamycin (6 L, 37 °C). Cultures were cooled to 15 °C after cells reached OD<sub>600</sub> = 0.4, and protein expression was induced (0.5 mM IPTG). After 16

hrs, cells were harvested by centrifugation (4,500 x g, 10 min) and resuspended in 40 mL lysis buffer (0.5 M NaCl, 30 mM HEPES pH 7.5). Cells were lysed via sonication, and cellular debris was removed (centrifugation at 30,000 x g, 60 min). EryMod6TE and GDH lysates were dialyzed against 1 L lysis buffer at 4 °C. After 4 hrs, the buffer was exchanged with 1 L fresh lysis buffer, and allowed to dialyze for an additional 16 hrs. Lysates were centrifuged (4,000 x g, 10 min), flash-frozen, and stored at -80 °C until further use. EryTE was further purified for TE-mediated hydrolysis experiments by passing the cell lysate over a nickel-NTA column equilibrated with lysis buffer. The column was washed with lysis buffer containing 15 mM imidazole (25 mL) and protein was eluted with lysis buffer containing 150 mM imidazole (5 mL).

### **EryMod6TE reaction conditions**

The following stock solutions were prepared: 1.5 M Tris-HCl pH 7.5, 5 M NaCl, 3 M glucose, 100 mM NADP<sup>+</sup>, 0.5 M **4.1** and **4.2** in DMSO, and 0.1 M **4.3**, **4.6-4.10** in DMSO. Tris-HCl pH 7.5 (150 mM), NaCl (100 mM), glucose (500 mM), NADP<sup>+</sup> (10 mM), extender unit **4.1** or **4.2** (5 mM), priming unit **4.3** or **4.6-4.10** (2.5 mM), pure DMSO (5% v/v), water, dialyzed GDH lysate (10% v/v), and dialyzed EryMod6TE lysate (25% v/v) were combined (200 mL reaction volume) in a microcentrifuge tube. After addition of each dialyzed enzyme lysate, reactions were gently mixed by pipetting and then centrifuged (3000 x g, 5 sec). Reactions were incubated at 23 °C for 16 hrs before being subjected to CuAAC with sulforhodamine B azide. Negative control reactions contained all substrates except for EryMod6TE dialyzed lysate. Reactions were performed in duplicate.

### Preparative in vitro synthesis of reduced diketide 4.5

Tris-HCl pH 7.5 (150 mM), NaCl (100 mM), glucose (500 mM), NADP<sup>+</sup> (10 mM), extender unit **4.2** (5 mM), priming unit **4.3** (2.5 mM), pure DMSO (total 5% v/v), water, dialyzed GDH lysate (10% v/v), and dialyzed EryMod6TE lysate (25% v/v) were combined in a conical tube in a total volume of 25 mL. After the addition of each dialyzed lysate, reactions were gently mixed by pipetting and then centrifuged (3000 x g, 5 sec). After 16 hrs, the reaction was washed using diethyl ether (1 x 50 mL). The aqueous layer was then acidified using 12 N HCl (2 mL) and products were extracted with ethyl acetate (3 x 150 mL). The organic layer was dried over sodium sulfate and concentrated *in vacuo*. Diketide **4.5** was purified by dry flash column chromatography (silica, eluting with 10:89:1 ethyl acetate:hexanes:acetic acid).

### **Phosphate-containing reaction conditions**

Tris-HCl pH 7.5 (150 mM), NaCl (100 mM), extender unit **4.2** (5 mM), priming unit **4.3** (2.5 mM), pure DMSO (total 5% v/v), Na<sub>2</sub>HPO<sub>4</sub> pH 7 (0, 10, 50 or 100 mM), water, dialyzed GDH lysate (10% v/v) and dialyzed EryMod6TE lysate (25% v/v) were combined (200 mL total volume) in a microcentrifuge tube. After addition of each dialyzed lysate, reactions were gently mixed by pipetting and then centrifuged (3000 x g, 5 sec). Reactions were incubated at 23 °C for 16 hrs before being subjected to CuAAC with sulforhodamine B azide. Negative control reactions contained all substrates except for EryMod6TE dialyzed lysate. Reactions were performed in duplicate.

### **Glycerol-containing reaction condition**

Reactions were prepared as described in "Conditions for EryMod6TE reactions" except for the use of 10% v/v glycerol and EryMod6TE knockouts (25% v/v; ΔKS, ΔAT, ΔTE, ΔKS+AT, ΔKS+TE, and ΔAT+TE). In the experiment with purified EryTE, Tris-

HCl pH 7.5 (150 mM), NaCl (100 mM), priming unit **4.3** (2.5 mM), glycerol (10% v/v), water, and nickel-purified EryTE (25% v/v) were added (200 mL total volumne) to a microcentrifuge tube. Reactions were incubated at 23 °C for 16 hrs before being subjected to CuAAC with sulforhodamine B azide.

#### **CuAAC** reaction conditions

The following stock solutions were prepared: 10 mM sulforhodamine B azide in DMSO, 1 M sodium ascorbate, and 0.5 M copper(II) sulfate. DMSO (50% v/v), sulforhodamine B azide (1 mM), EryMod6TE reaction (0.75 mM), and sodium ascorbate (40 mM) were combined (48 mL total volume) in a microcentrifuge tube. Reactions were vortexed and centrifuged (3000 x g, 5 sec). Copper(II) sulfate (20 mM) was then added and immediately vortexed for 3 seconds to initiate the click reaction. Samples were incubated at 23°C for 1 hr and stored at -80 °C until further use.

### **HPLC** and fluorescence detection

Reactions subjected to CuAAC were centrifuged to remove debris (20,000 x g, 2 min). Samples (20 mL) were analyzed using a Waters Symmetry® C18 3.5  $\mu$ m 4.6 x 75 mm column on a Beckman Coulter System Gold® 126 Solvent Module equipped with a Jasco FP-2020 Plus Intelligent Fluorescence Detector ( $\lambda_{ex}$  = 565 nm,  $\lambda_{em}$  = 586 nm). The mobile phases consisted of degassed, deionized water + 0.1% TFA (solvent A) and degassed acetonitrile + 0.1% TFA (solvent B). A linear gradient (flow rate = 1 mL/min) of 75% to 55% B over 40 min followed by 10 min at 100% B was used to analyze reactions. Data was analyzed using 32 Karat Software.

#### LC/MS analysis

Whole reactions or compounds collected from HPLC runs were subjected to negative-ESI LC/MS (Agilent Technologies 1200 Series HPLC with a Gemini C18

column coupled to an Agilent Technologies 6130 quadrupole mass spectrometer). Mobile phases consisted of water containing 0.1% formic acid (solvent A) and acetonitrile containing 0.1% formic acid (solvent B). A linear gradient (flow rate = 0.7 mL/min) of 5%–95% B over 12 min was used.

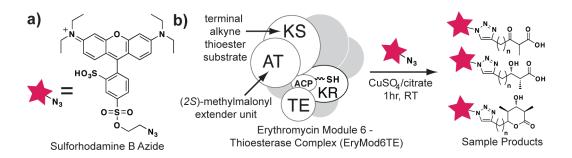
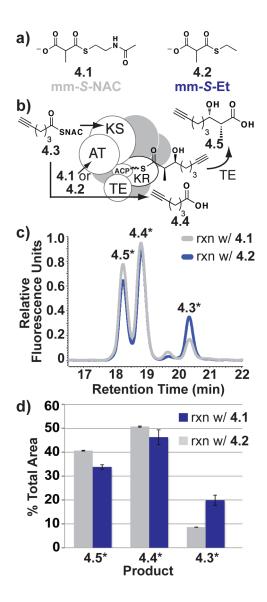


Figure 4-1 Lighting up polyketide products through fluorescent click chemistry.

a) The sulforhodamine B azide fluorophore. b) Terminal alkyne priming units and methylmalonyl extender units are incubated with PKS constructs. Polyketide products are analyzed by copper(I)-catalyzed azide alkyne cycloaddition (CuAAC, click chemistry) with the fluorophore and adducts were monitored by reverse-phase HPLC. (KS, ketosynthase; AT, acyltransferase; KR, ketoreductase; ACP, acyl carrier protein; TE, thioesterase).



**Figure 4-2** A comparison of *in vitro* extender units.

a) Methylmalonyl-S-N-acetylcysteamine (mm-S-NAC, **4.1**) and methylmalonyl-S-ethanethiol (mm-S-Et, **4.2**). b) Reduced diketide **4.5** is produced by EryMod6TE from priming unit **4.3**, extender unit **4.1** or **4.2**, and the NADPH regeneration system. EryTE-mediated hydrolysis of **4.3** yields **4.4**. c) Fluorimeter-coupled HPLC analysis of reactions ( $\lambda_{ex} = 565$  nm,  $\lambda_{em} = 586$  nm). Adduct masses were determined by LC/MS. d) EryMod6TE uses extender units **4.1** and **4.2** interchangeably, as determined through a comparison of the peak areas of **4.3\*–4.5\*** (asterisks indicate sulforhodamine adducts).

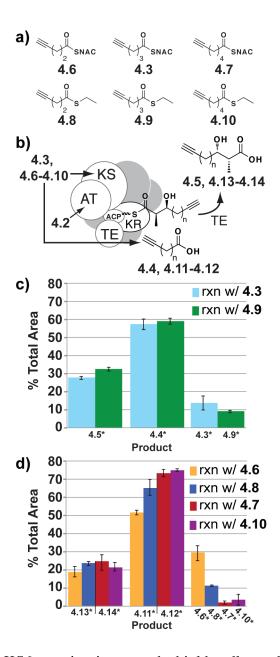
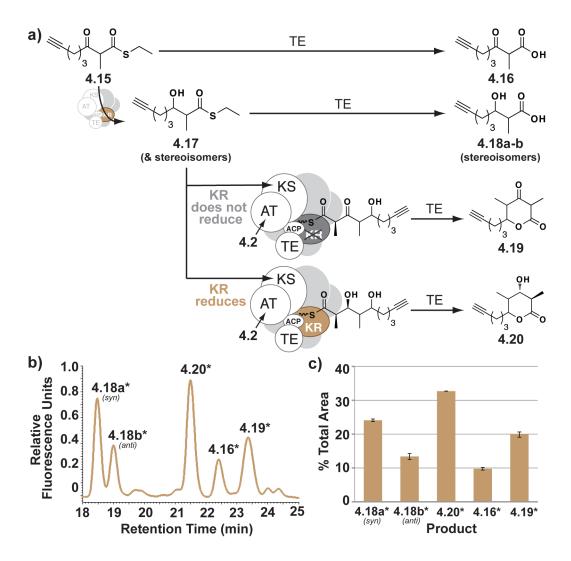


Figure 4-3 EryKS6 promiscuity towards thiol handles and chain lengths.

a) Terminal alkyne thioester priming unit substrates. b) Reduced diketides **4.5**, **4.13**, and **4.14** are generated by EryMod6TE from priming units shown in the previous panel, extender unit **4.2**, and the NADPH regeneration system. EryTE-catalyzed hydrolysis of the priming units yields carboxylic acids **4.4**, **4.11**, and **4.12**. c) Equivalent product profiles are observed for NAC-linked priming unit **4.3** and ethanethiol-linked priming unit **4.9**. d) Equivalent product profiles are observed for substrates of varying chain length, regardless of thiol handle.



**Figure 4-4** EryMod6TE reactivity towards an advanced priming unit.

a)  $\beta$ -Ketoacyl ethanethioester **4.15** is reduced by EryKR6. At least one of the resulting stereoisomers (**4.17**) is extended by EryKS6. EryTE-mediated cyclization yields ketolactone **4.19**, while EryKR6-mediated reduction followed by EryTE-mediated cyclization yields hydroxylactone **4.20**. EryTE also hydrolyzes diketides **4.15** and **4.17** to yield **4.16** and **4.18a-b**, respectively. b) Reverse-phase HPLC analysis of reactions ( $\lambda_{ex} = 565 \text{ nm}$ ,  $\lambda_{em} = 586 \text{ nm}$ ). Diastereomer(s) **4.18a** are *syn* and diastereomer(s) **4.18b** are *anti*, as the retention time of **4.18a\*** matches that of **4.5\***. c) A bar graph compares the peak areas from the previous panel.

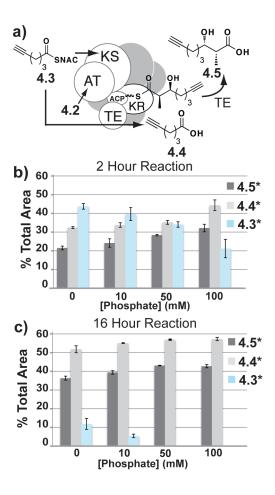


Figure 4-5 Phosphate enhances turnover.

a) EryMod6TE operating as a reduced diketide synthase. b) A comparison of peak areas for 2 hr reactions shows increased substrate turnover with phosphate concentration. c) A comparison of peak areas for 16 hr reactions shows complete substrate turnover when the phosphate concentration is at least 50 mM.

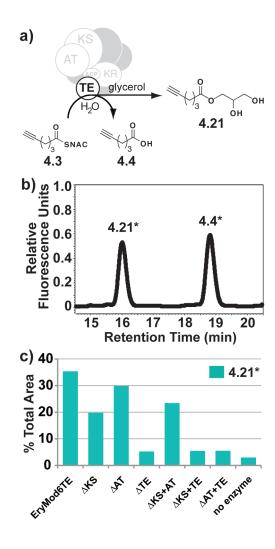


Figure 4-6 EryTE-mediated glycerolysis of priming units.

a) TE-mediated hydrolysis has been implicated in low biocatalytic product yields. Here EryTE is also shown to catalyze glycerolysis in reaction buffers containing glycerol. b) An HPLC trace ( $\lambda_{ex} = 565$  nm,  $\lambda_{em} = 586$  nm) reveals the generation of both the hydrolysis product **4.4** and the glycerolysis product **4.21** from priming unit **4.3**. c) EryMod6TE mutants containing an active EryTE produce significant quantities of glycerolysis product **4.21**.

## **Chapter Five: Conclusions**

Here, we demonstrated the S. coelicolor MatB-catalyzed synthesis of polyketide extender units malonyl-CoA, methylmalonyl-CoA, ethylmalonyl-CoA, methoxymalonyl-CoA, and hydroxymalonyl-CoA, marking the first enzymatic syntheses of methoxymalonyl-CoA and hydroxymalonyl-CoA. Equivalent PANT- and NAC-linked extender units were also generated, demonstrating that MatB was not only promiscuous toward malonyl derivatives, but also thiol acceptors. Crystal structures of two MatB ternary product complexes were obtained with malonyl-CoA and AMP and (2R)methylmalonyl-CoA and AMP bound (1.45 Å and 1.43 Å resolution, respectively). These are the first structures of a diacid-activating adenylate-forming enzyme. The strong electron density surrounding the protein, thioester product, and AMP provides some of the most complete snapshots of product formation in the enzyme superfamily. The orientation of the (2R)-methylmalonyl-CoA  $\alpha$ -methyl group reveals the stereoselectivity of MatB as well as the pocket that enables MatB to be promiscuous toward the αsubstituents of the diacid. MatB generated CoA-, PANT- and NAC-linked methylmalonyl extender units were employed in the synthesis of a triketide pyrone, exemplifying the use of MatB for in vitro PKS reactions.

We showed EryMod6TE was capable of generating triketides *in vitro* on a scaled-up level from simple synthetic diketides and enzymatically generated building blocks. The triketide pyrone products were readily observed by HPLC and characterized by NMR. Encouragingly, EryMod6TE possesses broad substrate specificity, which enabled it to be employed as a general biocatalyst. A scaled-up reaction with a diketide priming unit revealed that the yields were limited to ~12% due to the competing side-reaction of

TE-mediated hydrolysis of the diketide substrate. TEs from other synthases were fused to EryMod6; while they were capable of forming pyrones, they also catalyzed hydrolysis. We discovered that EryMod6TE was capable of "stuttering", a phenomenon in which one module catalyzes the addition of multiple extender units to a growing polyketide. This ability was harnessed biocatalytically to generate the anticipated pyrones from monoketide substrates.

We also showcased a quantitative detection method that reports the fate of alkynyl priming units in module-catalyzed reactions within cell lysate. CuAAC with sulforhodamine B azide enabled normally nonchromophoric polyketide products to be sensitively visualized and identified using fluorescence spectroscopy and mass spectrometry. Employing this method, the incorporation of ethanethiol-linked priming and extender units into reduced diketide compounds by EryMod6TE was investigated. Ethanethiol handles were found to be as effective as NAC- handles. Competing reactivities of EryMod6TE were observed through the conversion of a β-ketoacyl ethanethioester priming unit to triketide lactones. We also noted an increase in enzymatic turnover in a single-module catalyzed reaction with increased phosphate concentrations and TE-mediated glycerolysis of priming units in reaction buffers containing glycerol.

Taken together, the works described in this dissertation have helped the scientific community advance towards procuring a PKS-generated library of stereochemically diverse molecules *in vitro*. We increased economic feasibility of preparative-scale synthesis of polyketide compounds by employing MatB-generated extender units. We demonstrated the utility of EryMod6TE as a general biocatalyst of polyketide compounds by successfully characterizing 10 triketide pyrone compounds, attesting to the scalability of *in vitro* PKS reactions. Lastly, we developed a new, bioorthogonal, rapid, inexpensive and sensitive method to detect polyketide products upon CuAAC with sulforhodamine B

azide. This method was used to optimize reaction conditions and test the reactivites of domains within EryMod6TE. The work presented here has helped set the stage for the preparative, *in vitro* PKS synthesis of a library of stereoisomeric polyketides which could serve as chiral synthons to generate large quantities of stereochemically complex natural products and natural product analogs, with the ultimate goal of providing novel therapeutics to the infirm.

# **Appendix 1. Supporting Information for Chapter 3**

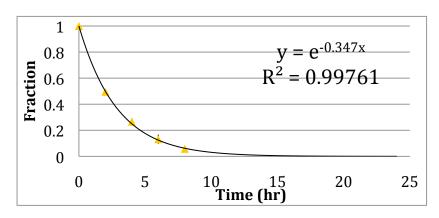
### HPLC DATA FOR TE-MEDIATED HYDROLYSIS OF POLYKETIDE SUBSTRATES

TE-mediated hydrolysis of 3.4

	HPLC Peak Area at 235 nm		
Time (hr)	Trial 1	Trial 2	Trial 3
0	25032051	26011949	24493101
2	12485244	12913755	12019588
4	6711618	7548863	5860547
6	2661936	4447764	2882410
8	1164408	2098361	1156503

Normalized Peak Area			
Trial 1	Trial 2	Trial 3	
norm.	norm.	norm.	
1	1	1	
0.50	0.50	0.49	
0.27	0.29	0.24	
0.11	0.17	0.12	
0.05	0.08	0.05	

Time (hr)	Average	St. Dev.
0	1	0
2	0.50	0.00
4	0.27	0.03
6	0.13	0.03
8	0.06	0.02



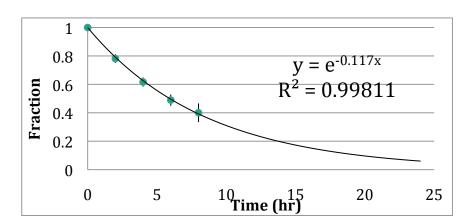
Peak areas for EryMod6EryTE hydrolysis of **3.4** were determined using Waters Breeze<sup>TM</sup> 2 System Software from reverse-phase chromatograms monitored at 235 nm. For each time trial, peak areas were normalized to 1 at time = 0 hour. Averages were plotted in Microsoft Excel and fit to an exponential curve.

TE-mediated hydrolysis of 3.7

	HPLC Peak Area at 235 nm		
Time (hr)	Trial 1	Trial 2	Trial 3
0	8870433	7759841	13221525
2	6849638	5871402	10787740
4	5283320	4659478	8638482
6	4258585	3526882	7045881
8	3011532	3037601	6216343

Normalized Peak Area			
pp1 norm.	pp2 norm.	pp3 norm.	
1	1	1	
0.77	0.76	0.82	
0.60	0.60	0.65	
0.48	0.45	0.53	
0.34	0.39	0.47	

Time (hr)	Average	St. Dev.
0	1	0
2	0.78	0.03
4	0.62	0.03
6	0.49	0.04
8	0.40	0.07



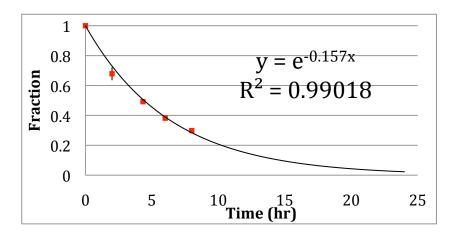
Peak areas for EryMod6EryTE hydrolysis of **3.7** were determined using Waters Breeze<sup>TM</sup> 2 System Software from reverse-phase chromatograms monitored at 235 nm. For each time trial, peak areas were normalized to 1 at time = 0 hour. Averages were plotted in Microsoft Excel and fit to an exponential curve.

TE-mediated hydrolysis of 3.15

	HPLC Peak Area at 235 nm			
Time (hr)	Trial 1 (p1) Trial 2 (p2) Trial 3 (j			
0	9018668	9011877	9812933	
2	6445475	6221437	6189349	
4.33	4471662	4387909	4841047	
6	3548422	3398701	3674209	
8	2815218	2660332	2822005	

Normalized Peak Area			
p1 norm.	p2 norm.	p3 norm.	
1	1	1	
0.71	0.69	0.63	
0.50	0.49	0.49	
0.39	0.38	0.37	
0.31	0.30	0.29	

Time	Average	St. Dev.
(hr)		
0	1	0
2	0.68	0.04
4.33	0.49	0.00
6	0.38	0.01
8	0.30	0.01



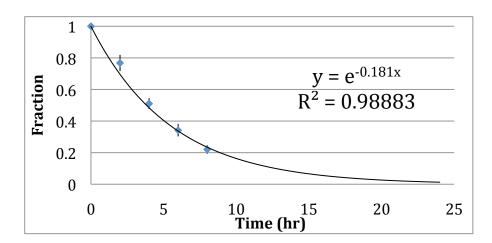
Peak areas for EryMod6EryTE hydrolysis of **4.15** were determined using Waters Breeze-<sup>TM</sup> 2 System Software from reverse-phase chromatograms monitored at 235 nm. For each time trial, peak areas were normalized to 1 at time = 0 hour. Averages were plotted in Microsoft Excel and fit to an exponential curve.

TE-mediated hydrolysis of 3.16

	HPLC Peak Area at 235 nm			
Time (hr)	Trial 1 (b1)	Trial 2 (b2)	Trial 3 (b3)	
0	12390224	7221369	8228404	
2	10218321	5236959	6198739	
4	6806394	3512810	4075931	
6	4690121	2174455	2846014	
8	3065041	1391035	1816802	

Normalized Peak Area			
b1 norm. b2 norm. b3 norm			
1	1	1	
0.82	0.73	0.75	
0.55	0.49	0.50	
0.38	0.30	0.35	
0.25	0.19	0.22	

Time (hr)	Average	St. Dev.
0	1	0
2	0.77	0.05
4	0.51	0.03
6	0.34	0.04
8	0.22	0.03



Peak areas for EryMod6EryTE hydrolysis of 4.16 were determined using Waters Breeze-TM 2 System Software from reverse-phase chromatograms monitored at 235 nm. For each time trial, peak areas were normalized to 1 at time = 0 hour. Averages were plotted in Microsoft Excel and fit to an exponential curve.

#### HPLC PEAK AREAS FOR ACTIVITY ASSAY OF ERYMOD6TE VARIANTS

#### **Experiment 1**

Raw HPLC Peak Areas of 3.11 for EryMod6TE variants

Trial	EryMod6TE	EryMod6- CdaTE	EryMod6- RadTE	EryMod6- AmpTE
1a	5261282	1314142	2428190	3077927
2a	5528033	1392116	2490892	2839655
3a	5500844	1431234	2607389	2932649

AVERAGE

5430053

Normalized HPLC Peak Areas to EryMod6TE Average Peak Area

Trial	EryMod6TE	EryMod6- CdaTE	EryMod6- RadTE	EryMod6- AmpTE
1a	0.969	0.242	0.447	0.567
2a	1.018	0.256	0.459	0.523
3a	1.013	0.264	0.480	0.540
<b>AVERAGE</b>	1.000	0.254	0.462	0.543

**Experiment 2** 

Raw and Normalized HPLC Peak Areas of 3.11 for EryMod6

	Raw HPLC Peak Area		ea Normalized HPLC Peak Are	
Trial	EryMod6TE	EryMod6	EryMod6TE	EryMod6
1b	4468906	378312	1.033	0.087
2b	4244724	382120	0.982	0.088
3b	4259734	356552	0.985	0.082
DACE	1321155		1 000	0.096

**AVERAGE** 

4324455

1.000 0.086

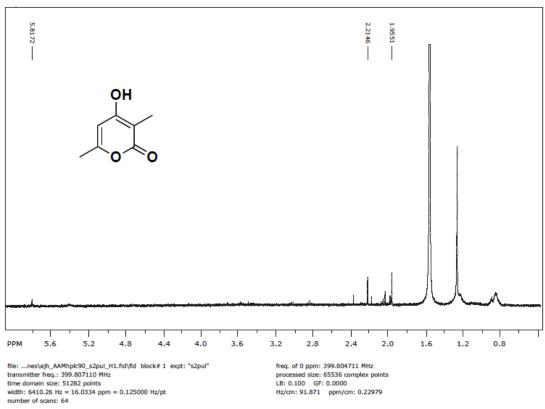
#### Combined Normalized Data

Trial	EryMod6TE	EryMod6- CdaTE	EryMod6- RadTE	EryMod6- AmpTE
1a	0.969	0.242	0.447	0.567
2a	1.018	0.256	0.459	0.523
3a	1.013	0.264	0.480	0.540
<b>AVERAGE</b>	1.000	0.254	0.462	0.543
Standard Dev.	0.027	0.043	0.036	0.041

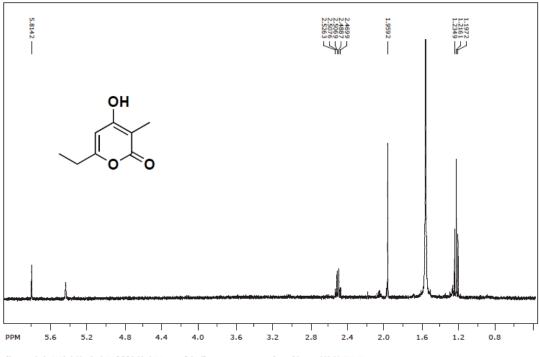
Trial	EryMod6TE	EryMod6
1b	1.033	0.087
2b	0.982	0.088
3b	0.985	0.082
AVERAGE	1.000	0.086
Standard Dev.	0.029	0.037

#### <sup>1</sup>H NMR CHARACTERIZATION OF TRIKETIDE PYRONE COMPOUNDS

### Compound 3.8

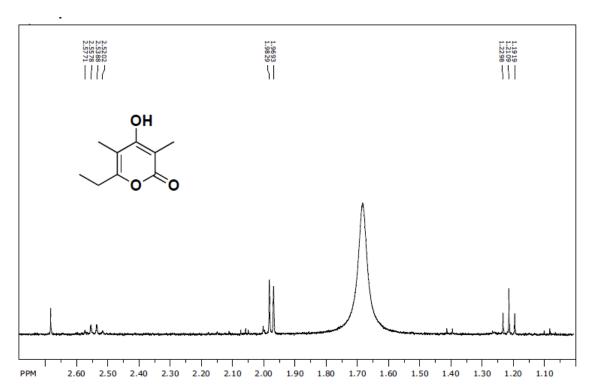


freq. of 0 ppm: 399.804711 MHz processed size: 65536 complex points LB: 0.100 GF: 0.0000 Hz/cm: 91.871 ppm/cm: 0.22979



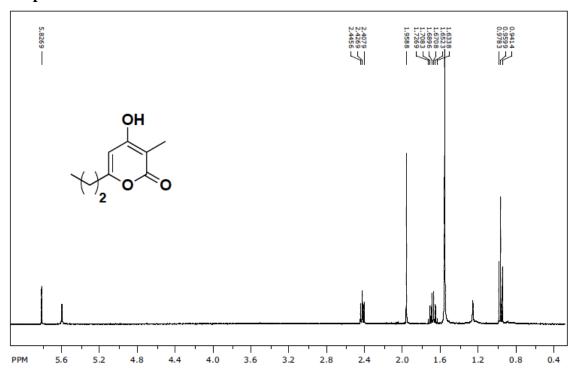
file: ...nes\ajh\_PAMhpic90\_s2pul\_H1.fid\fid block# 1 expt: "s2pul" transmitter freq.: 399.807110 MHz time domain size: 51282 points width: 641.026 Hz = 16.0334 ppm = 0.125000 Hz/pt number of scans: 64

freq. of 0 ppm: 399.804711 MHz processed size: 65536 complex points LB: 0.100 GF: 0.0000 Hz/cm: 92.089 ppm/cm: 0.23033



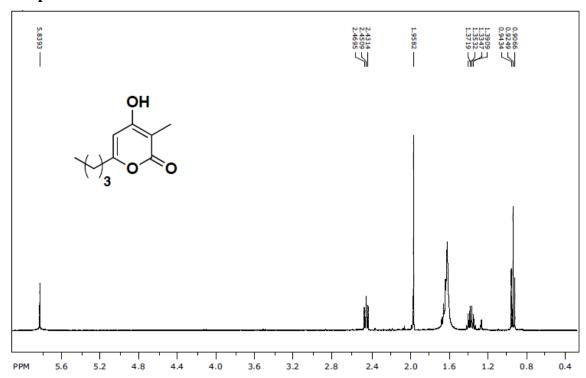
 $\label{eq:file:....afh_PPMhplc90new_s2pul_H1.fid\fid block# 1 expt: "s2pul" transmitter freq.: 399.807110 MHz time domain size: 51282 points width: 6410.26 Hz = 16.0334 ppm = 0.125000 Hz/pt number of scans: 64$ 

freq. of 0 ppm: 399.804710 MHz processed size: 65536 complex points LB: 0.100 GF: 0.0000 Hz/cm: 28.587 ppm/cm: 0.07150



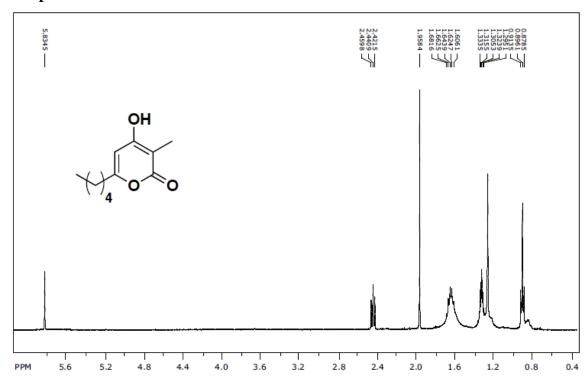
file: ...nes\ajh\_BAMhpkS0\_s2pul\_H1.fid\\\\fid\block\#1 expt: "s2pul" transmitter freq.: 399.807110 MHz time domain size: 51282 points width: 6410.26 Hz = 16.0334 ppm = 0.125000 Hz/pt number of scans: 64

freq. of 0 ppm: 399.804711 MHz processed size: 65536 complex points LB: 0.100 GF: 0.0000 Hz/cm: 94.053 ppm/cm: 0.23525



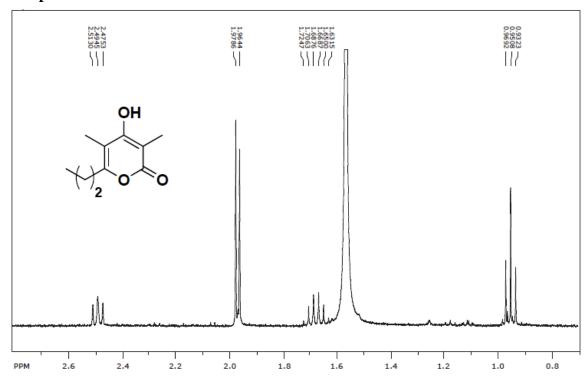
 $file: ....nes\ajh\_VAMhplc90\_s2pul\_H1.fid\fid\ block\#\ 1 \ expt: "s2pul" transmitter freq.: 399.807110 MHz time domain size: 51282 points width: 6410.26 Hz = 16.0334 ppm = 0.125000 Hz/pt number of scans: 64$ 

freq. of 0 ppm: 399.804710 MHz processed size: 65536 complex points LB: 0.100 GF: 0.0000 Hz/cm: 93.835 ppm/cm: 0.23470



file: ...nes\ajh\_HAMhplc90\_s2pul\_H1.fid\fid block# 1 expt: "s2pul" transmitter freq.: 399.807110 MHz time domain size: 51282 points width: 6410.26 Hz = 16.0334 ppm = 0.125000 Hz/pt number of scans: 64

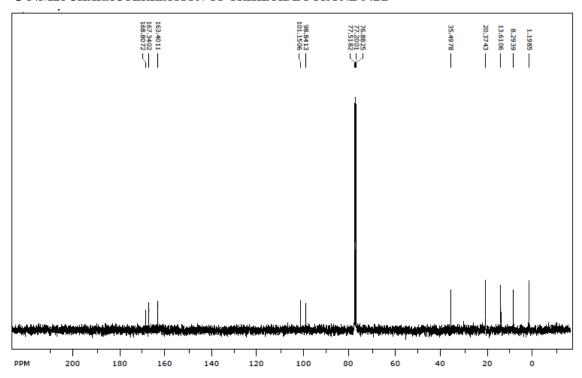
freq. of 0 ppm: 399.804711 MHz processed size: 65536 complex points LB: 0.100 GF: 0.0000 Hz/cm: 93.181 ppm/cm: 0.23306



file: ...es\ajh\_BMMhpic902\_s2pul\_H1.fid\fid block# 1 expt: "s2pul" transmitter freq.: 399.807110 MHz time domain size: 51282 points width: 6410.26 Hz = 16.0334 ppm = 0.125000 Hz/pt number of scans: 64

freq. of 0 ppm: 399.804711 MHz processed size: 65536 complex points LB: 0.100 GF: 0.0000 Hz/cm: 33.824 ppm/cm: 0.08460

### $^{13}\mbox{C}$ NMR characterization of triketide pyrone 3.11



file: ...top\ajh\_20mgBAMC\_s2pul\_C13.fid\fid block# 1 expt: "s2pul" transmitter freq.: 100.541530 MHz time domain size: 63750 points width: 2450-980 Hz = 243.7779 ppm = 0.384468 Hz/pt number of scans: 64

freq. of 0 ppm: 100.530954 MHz processed size: 65536 complex points LB: 2.000 GF: 0.0000 Hz/cm: 980.392 ppm/cm: 9.75112

# **Appendix 2. Supporting Information for Chapter 4**

### HPLC PEAK AREAS FOR CUAAC REACTIONS

Compound **4.1** reaction analysis

Trial 1

	RET Time	AREA	% Total AREA
4.5*	18.727	15787002	40.5
4.4*	19.294	19825587	50.9
4.3*	20.825	3343245	8.6

Trial 2

	RET Time	AREA	% Total AREA
4.5*	18.679	16962289	40.8
4.4*	19.249	21039755	50.6
4.3*	20.800	3606768	8.7

Compound 1			
AVERAGE STDEV			
4.5*	40.6	0.2	
4.4*	50.7	0.2	
4.3*	8.6	0.1	

#### Compound 4.2 reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.5*	18.447	11912678	33.1
4.4*	19.030	17413347	48.5
4.3*	20.539	6613850	18.4

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.5*	18.699	15345004	34.5
4.4*	19.270	19658811	44.1
4.3*	20.787	9527100	21.4

Compound <b>4.2</b>				
AVERAGE STDEV				
4.5*	33.8	0.9		
4.4*	46.3	3.0		
4.3*	19.9	2.1		

### Compound 4.1 CONTROL Reaction Analysis

### Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.4*	19.444	8176075	20.6
4.3*	20.910	31563023	79.4

### Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.4*	19.490	6260615	17.5
4.3*	21.007	29561878	82.5

Compound 4.1 CONTROL			
AVERAGE STDEV			
4.4*	19.0	2.2	
4.3*	81.0	2.2	

# Compound 4.2 CONTROL reaction analysis

### Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.4*	19.444	5960798	15.8
4.3*	20.977	31720433	84.2

### Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.4*	19.485	5514672	15.0
4.3*	21.087	31295599	85.0

Compound <b>4.2</b> CONTROL			
	AVERAGE STDEV		
4.4*	15.4	0.6	
4.3*	84.6	0.6	

# Compound **4.6** reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.13*	17.809	9013751	20.7
4.11*	18.680	22850133	52.6
4.6*	20.194	11593317	26.7

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.13*	17.592	8347381	16.6
4.11*	18.484	25467750	50.8
4.6*	19.930	16329898	32.6

	Compound <b>4.6</b>			
AVERAGE STDEV				
4.13*	18.7	2.9		
4.11*	51.7	1.3		
4.6*	29.6	4.2		

### Compound 4.8 reaction analysis

Trial 1

	Ret. Time	AREA	% Total
	(min)		AREA
4.13*	17.867	7061148	20.6
4.11*	18.714	23420844	68.3
4.8*	35.014	3821288	11.1

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.13*	17.842	10990448	26.6
4.11*	18.752	25630547	62.1
4.8*	35.048	4675599	11.3

Compound <b>4.8</b>				
AVERAGE STDEV				
4.13*	23.6	4.3		
4.11*	65.2	4.4		
4.8*	11.2	0.1		

### Compound 4.6 CONTROL reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.11*	18.737	172204	0.5
4.6*	20.024	37911066	99.5

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.11*	18.465	2424183	6.1
4.6*	19.745	37329938	93.9

Compound <b>4.6</b> CONTROL			
	AVERAGE STDEV		
4.11*	3.3	4.0	
<b>4.6*</b> 96.7		4.0	

# Compound 4.8 CONTROL reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.11*	18.755	212419	0.6
4.8*	33.681	36043103	99.4

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.11*	18.827	2710007	6.1
4.8*	33.460	41899558	93.9

Compound 4.8 CONTROL			
	AVERAGE STDEV		
4.11*	3.3	3.9	
4.8*	96.7	3.9	

# Compound 4.3 reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
	(111111)		AKEA
4.5*	18.620	15828462	28.3
4.4*	19.092	33896564	60.6
4.3*	20.719	6183995	11.1

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.5*	18.464	11380788	27.1
4.4*	19.105	23650841	56.3
4.3*	20.509	6981340	16.6

	Compound 4.3			
	STDEV			
4.5*	27.7	0.9		
4.4*	58.5	3.1		
4.3*	13.8	3.9		

### Compound 4.9 reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.5*	18.692	14020189	31.5
4.4*	19.374	26711265	60.0
4.9*	36.719	3814999	8.6

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.5*	18.697	14966330	32.9
4.4*	19.342	26192166	57.6
4.9*	36.746	4290108	9.4

Compound <b>4.9</b>				
	AVERAGE	STDEV		
4.5*	32.2	1.0		
4.4*	58.8	1.6		
4.9*	9.0	0.6		

### Compound 4.3 CONTROL reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.4*	19.359	6049863	15.9
4.3*	20.760	31884840	84.1

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.4*	19.330	6555358	16.5
4.3*	20.849	33124632	83.5

Compound 4.3 CONTROL		
AVERAGE STDEV		
4.4*	16.2	0.4
4.3*	83.8	0.4

### Compound 4.9 CONTROL reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.4*	19.300	594660	4.0
4.9*	36.803	14175016	96.0

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.4*	19.314	500041	2.8
4.9*	36.754	17355117	97.2

Compound <b>4.9</b> CONTROL		
AVERAGE STDEV		
4.4*	3.4	0.9
4.9*	96.6	0.9

#### Compound **4.7** reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.14*	20.08	11581283	25.46
4.12*	20.68	32761857	72.03
4.7*	22.69	1138304	2.50

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.14*	20.16	10643821	23.85
4.12*	20.71	33422377	74.89
4.7*	22.70	564438	1.26

Compound 4.7			
	AVERAGE STDEV		
4.14*	24.7	1.1	
4.12*	73.5	2.0	
4.7*	1.9	0.9	

### Compound 4.10 reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.14*	20.16	10535136	22.95
4.12*	20.66	34727449	75.66
4.10*	39.71	637867	1.39

Trial 2

	Ret. Time (min)	AREA	% Total Area
4.14*	20.09	8001594	19.65
4.12*	20.62	30425013	74.71
4.10*	38.95	2299424	5.65

Compound <b>4.10</b>			
AVERAGE STDEV			
4.14*	21.3	2.3	
4.12*	75.2	0.7	
4.10*	3.5	3.0	

# Compound 4.7 CONTROL reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.12*	20.814	23100604	41.1
4.7*	22.619	33168434	58.9

Trial 2

	Ret. Time (min)	AREA	%total AREA
4.12*	20.655	25015046	43.3
4.7*	22.619	32755287	56.7

Compound 4.7 CONTROL			
	AVERAGE STDEV		
4.12*	42.2	1.6	
4.7*	57.8	1.6	

# Compound 4.10 CONTROL reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.12*	20.822	9162547	51.7
4.10*	39.851	8562374	48.3

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.12*	20.795	12754036	62.3
4.10*	39.769	7724401	37.7

Compound <b>4.10</b> CONTROL			
	AVERAGE STDEV		
4.12*	57.0	7.5	
4.10*	43.0	7.5	

### Compound 4.3 – 0 mM phosphate – 2 hrs - reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.5*	18.665	14806840	21.3
4.4*	19.232	22840606	32.8
4.3*	20.822	31899164	45.9

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.5*	18.867	13023723	22.8
4.4*	19.407	19204279	33.5
4.3*	20.987	25014238	43.7

0 mM Phosphate- 2 hrs			
AVERAGE STDEV			
4.5*	22.0	1.0	
4.4*	33.2	0.5	
4.3*	44.8	1.5	

Compound **4.3** – 10 mM phosphate – 2 hrs - reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.5*	18.580	15802446	23.1
4.4*	19.244	23049643	33.7
4.3*	20.774	29464589	43.1

Trial 2

	Ret. Time	AREA	% Total
	(min)	AREA	AREA
4.5*	18.890	13553917	26.4
4.4*	19.424	18227682	35.5
4.3*	21.020	19587930	38.1

10 mM Phosphate- 2 hrs			
AVERAGE STDEV			
4.5*	24.8	2.3	
4.4*	34.6	1.2	
4.3*	40.6	3.5	

Compound **4.3** – 50 mM phosphate – 2 hrs - reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.5*	18.464	16814392	29.4
4.4*	19.019	21047074	36.8
4.3*	20.539	19366060	33.8

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.5*	18.777	15811837	28.8
4.4*	19.325	19405246	35.3
4.3*	20.905	19771894	36.0

50 mM Phosphate- 2 hrs			
AVERAGE STDEV			
4.5*	29.1	0.4	
4.4*	36.0	1.1	
4.3*	34.9	1.5	

Compound 4.3 – 100 mM phosphate – 2 hrs - reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.5*	18.492	20585641	31.5
4.4*	19.125	28318670	43.3
4.3*	20.640	16505850	25.2

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.5*	18.862	19556396	34.4
4.4*	19.404	26989181	47.5
4.3*	21.020	10325964	18.2

100 mM Phosphate - 2 hrs			
AVERAGE STDEV			
4.5*	32.9	2.1	
4.4*	45.4	2.9	
4.3*	21.7	5.0	

# Compound **4.3** – 0 mM phosphate – 16 hrs - reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.5*	18.669	16536904	37.1
4.4*	19.250	23683668	53.1
4.3*	20.779	4360206	9.8

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.5*	18.740	15832878	35.6
4.4*	19.330	22474980	50.5
4.3*	20.864	6184094	13.9

0 mM Phosphate - 16 hrs			
AVERAGE STDEV			
4.5*	36.3	1.1	
4.4*	51.8	1.8	
4.3*	11.8	2.9	

### Compound **4.3** – 10 mM phosphate – 16 hrs - reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.5*	18.712	19279548	40.0
4.4*	19.317	26531272	55.1
4.3*	20.825	2328798	4.8

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.5*	18.382	19807554	38.8
4.4*	19.004	28157616	55.1
4.3*	20.495	3145642	6.2

10 mM Phosphate - 16 hrs			
AVERAGE STDEV			
4.5*	39.4	0.9	
4.4*	55.1	0.0	
4.3*	5.5	0.9	

# Compound **4.3** – 50 mM phosphate – 16 hrs - reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.5*	18.617	24264845	43,2
4.4*	19.374	31903020	56.8

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.5*	18.642	20931892	43.0
4.4*	19.279	27803314	57.0

50 mM Phosphate – 16 hrs		
AVERAGE STDEV		
4.5*	43.1	0.2
4.4*	56.9	0.2

### Compound **4.3** – 100 mM phosphate – 16 hrs - reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.5*	18.745	22205349	42.1
4.4*	19.300	30532660	57.9

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.5*	18.657	19918332	43.3
4.4*	19.212	26066248	56.7

100 mM Phosphate – 16 hrs			
AVERAGE STDEV			
4.5*	42.7	0.9	
4.4*	57.3	0.9	

Compound **4.3** – 0 mM phosphate – 16 hrs CONTROL reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.4*	19.317	6028704	14.9
4.3*	19.945	34549471	85.1

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.4*	19.295	6749666	16.2
4.3*	20.769	34841101	83.8

0 mM Phosphate CONTROL			
	AVERAGE STDEV		
4.4*	15.5	1.0	
4.3*	84.5	1.0	

# Compound **4.3** – 10 mM phosphate – 16 hrs CONTROL reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.4*	19.164	6257779	15.1
4.3*	20.692	35150517	84.9

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.4*	19.279	6576249	15.4
4.3*	20.592	36009401	84.6

10 mM Phosphate CONTROL			
	AVERAGE STDEV		
4.4*	15.3	0.2	
4.3*	84.7	0.2	

#### Compound **4.3** – 50 mM phosphate – 16 hrs CONTROL reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.4*	19.275	8543020	19.9
4.3*	20.612	34470898	80.1

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.4*	19.310	7864313	18.3
4.3*	20.894	35083973	81.7

50 mM Phosphate CONTROL			
	AVERAGE STDEV		
4.4*	19.1	1.1	
4.3*	80.9	1.1	

# Compound **4.3** – 100 mM phosphate – 16 hrs CONTROL reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.4*	19.294	7648977	18.5
4.3*	20.710	33648585	81.5

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.4*	19.274	9948370	23.0
4.3*	20.677	33388648	77.0

100 mM Phosphate CONTROL		
AVERAGE STDEV		
4.4*	20.7	3.1
4.3*	79.3	3.1

# Compound **4.15** reaction analysis

Trial 1

	Ret. Time (min)	AREA	% Total AREA
4.18a*	18.590	13305217	23.8
4.18b*	19.122	7131187	12.8
4.20*	21.575	18279563	32.7
4.16*	22.539	5606334	10.0
4.19*	23.498	11526588	20.6

Trial 2

	Ret. Time (min)	AREA	% Total AREA
4.18a*	18.845	12387653	24.4
4.18b*	19.372	7118427	14.0
4.20*	21.860	16615907	32.7
4.16*	22.774	4820335	9.5
4.19*	23.727	9887625	19.5

Compound <b>4.15</b>			
	AVERAGE	STDEV	
4.18a*	24.1	0.4	
4.18b*	13.4	0.9	
4.20*	32.7	0.0	
4.16*	9.8	0.4	
4.19*	20.0	0.8	

# Compound **4.3** EryMod6TE mutant reaction analysis

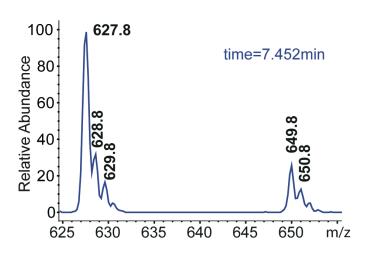
AREA – Trial 1

	EryMod6TE	ΔKS	ΔΑΤ	ΔΤΕ
4.21*	2614798	1554339	2198173	372883
4.4*	826717	1509532	1235584	859962
4.3*	3951371	4797879	3931098	5868967
Total AREA	7392886	7861750	7364855	7101812

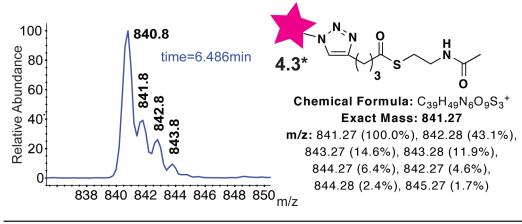
	ΔKS+AT	ΔKS+TE	ΔAT+TE	no EryMod6TE
4.21*	1708062	479870	409617	166666
4.4*	1138868	1127082	993593	0
4.3*	4447477	7212606	6012292	6333238
Total AREA	7294407	8819558	7415502	6499904

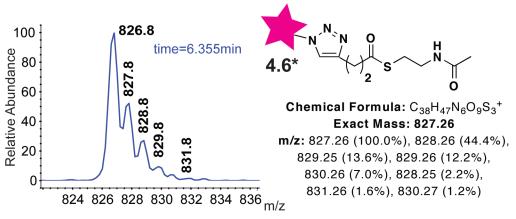
	% Total Area		
	4.21*	4.4*	4.3*
EryMod6TE	35.4	11.2	53.4
ΔKS	19.8	19.2	61.0
ΔΑΤ	29.8	16.8	53.4
ΔΤΕ	5.2	12.1	82.6
ΔKS+AT	23.4	15.6	61.0
ΔKS+TE	5.4	12.8	81.8
ΔAT+TE	5.5	13.4	81.1
no EryMod6TE	3.0	0.0	97.4

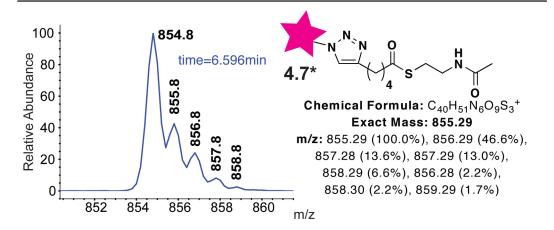
#### LC/MS characterization of sulforhodamine B adducts

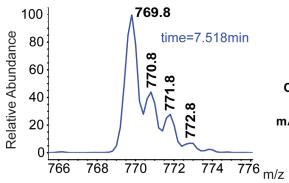


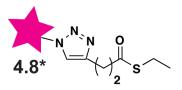
Chemical Formula:  $C_{29}H_{34}N_5O_7S_2^+$ Exact Mass: 628.19 m/z: 628.19 (100.0%), 629.19 (35.1%), 630.19 (11.6%), 630.20 (5.0%), 631.19 (3.1%), 631.20 (1.0%)





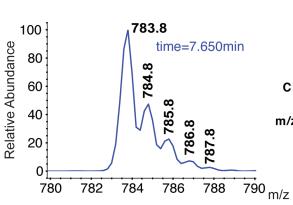


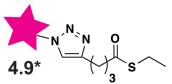




Chemical Formula:  $C_{36}H_{44}N_5O_8S_3^+$ Exact Mass: 770.23

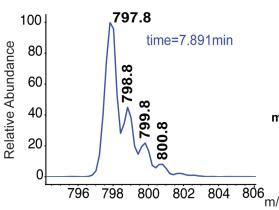
**m/z**: 770.24 (100.0%), 771.24 (39.7%), 772.23 (13.6%), 772.24 (11.0%), 773.23 (5.8%), 771.23 (4.2%), 774.24 (1.4%), 773.24 (1.2%)

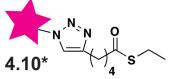




Chemical Formula: C<sub>37</sub>H<sub>46</sub>N<sub>5</sub>O<sub>8</sub>S<sub>3</sub><sup>+</sup> Exact Mass: 784.25

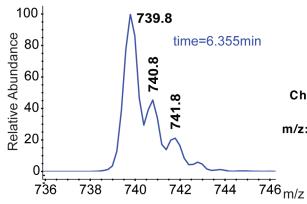
**m/z:** 784.25 (100.0%), 785.25 (44.3%), 786.25 (15.3%), 786.26 (9.8%), 787.25 (6.0%), 787.26 (1.9%), 788.25 (1.6%)

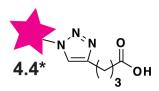




Chemical Formula:  $C_{38}H_{48}N_5O_8S_3^+$ Exact Mass: 798.27

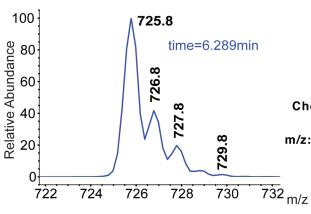
**m/z**: 798.27 (100.0%), 799.27 (44.4%), 800.26 (13.6%), 800.27 (11.8%), 801.27 (6.8%), 799.26 (1.8%), 802.27 (1.5%), 801.28 (1.2%)

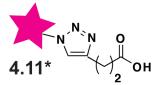




Chemical Formula: C<sub>35</sub>H<sub>42</sub>N<sub>5</sub>O<sub>9</sub>S<sub>2</sub><sup>+</sup> Exact Mass: 740.24

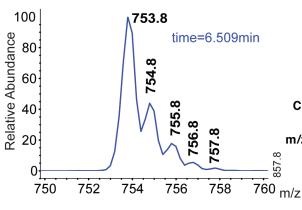
**m/z:** 740.24 (100.0%), 741.25 (38.7%), 742.24 (9.8%), 742.25 (9.8%), 743.24 (3.8%), 741.24 (3.4%), 743.25 (1.8%)

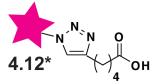




Chemical Formula:  $C_{34}H_{40}N_5O_9S_2^+$ Exact Mass: 726.23

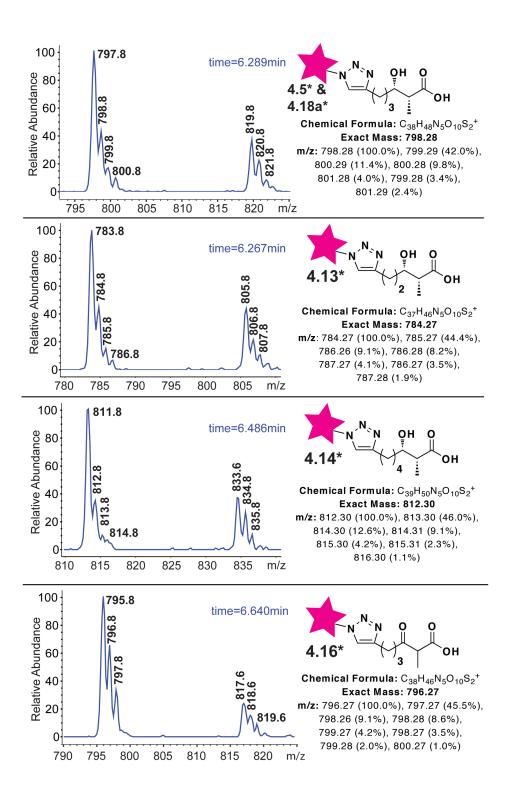
**m/z:** 726.23 (100.0%), 727.23 (39.2%), 728.23 (9.8%), 728.22 (9.1%), 729.23 (4.4%), 727.22 (1.8%)

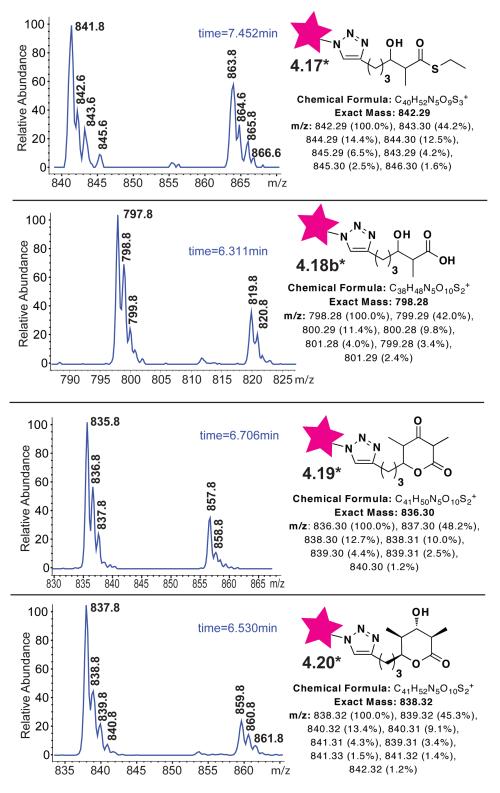




Chemical Formula:  $C_{36}H_{44}N_5O_9S_2^+$ Exact Mass: 754.26

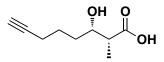
**m/z:** 754.26 (100.0%), 755.26 (43.2%), 756.26 (10.6%), 756.25 (9.1%), 757.26 (3.9%), 757.27 (1.7%)

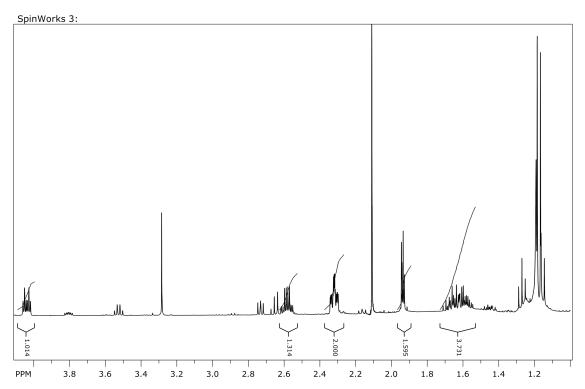




## $^{1}\mathrm{H}\ \mathrm{NMR}\ \mathrm{CHARACTERIZATION}$ of Terminal alkyne-containing compounds

## **Compound 4.5**

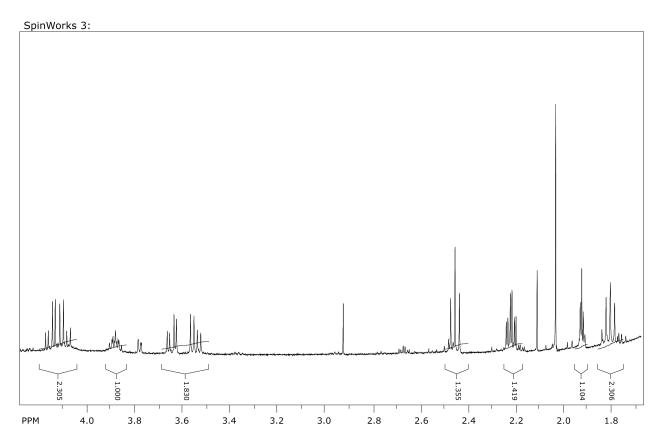




file: ...\ajh0419\_64fracSt\_s2pul\_H1.fid\fid block# 1 expt: "s2pul" transmitter freq.: 399.807110 MHz time domain size: 51282 points width: 6410.26 Hz = 16.0334 ppm = 0.125000 Hz/pt number of scans: 64

freq. of 0 ppm: 399.804738 MHz processed size: 65536 complex points LB: 0.100 GF: 0.0000 Hz/cm: 49.904 ppm/cm: 0.12482

## Compound 4.21



time domain size: 51282 points width: 6410.26 Hz = 16.0334 ppm = 0.125000 Hz/pt number of scans: 8

freq. of 0 ppm: 399.804737 MHz processed size: 65536 complex points LB: 0.100 GF: 0.0000 Hz/cm: 41.854 ppm/cm: 0.10469

## **Bibliography**

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 58, 1948–1954.

Beck, J., Ripka, S., Siegner, A., Schiltz, E., and Schweizer, E. (1990) The multifunctional 6-methylsalicylic acid synthase gene of *Penicillium patulum*. *Eur. J. Biochem*. 192, 487-498.

Beck, B. J., Aldrich, C. C., Fecik, R. A., Reynolds, K. A., and Sherman, D. H. (2003) Substrate recognition and channeling of monomodules from the pikromycin polyketide synthase. *J. Am. Chem. Soc.* 125, 12551-12557.

Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C. H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M. A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B. G., Parkhill, J., and Hopwood, D. A. (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417, 141-147.

Brooks, D. W., Lu, L. D.-L., and Masamune, S. (1979) *C*-Acylation under virtually neutral conditions. *Angew. Chem. Int. Ed. Engl.* 18, 72-74.

Caffrey, P., Lynch, S., Flood, E., Finnan, S., and Oliynyk, M. (2003) Amphotericin biosynthesis in *Streptomyces nodosus*: deductions from analysis of polyketide synthase and late genes. *Chem. Biol.* 10, 93-94.

Carroll, B. J., Moss, S. J., Bai, L. Q., Kato, Y., Toelzer, S., Yu, T. W., and Floss, H. G. (2002) Identification of a set of genes involved in the formation of the substrate for the incorporation of the unusual "glycolate" chain extension unit in ansamitocin biosynthesis. *J. Am. Chem. Soc.* 124, 4176–4177.

CCP4 (Collaborative Computational Project, Number 4). (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 50, 760–763.

Chan, Y. A., and Thomas, M. G. (2010) Recognition of (2S)-aminomalonyl-acyl carrier protein (ACP) and (2R)-hydroxymalonyl-ACP by acyltransferases in zwittermicin A biosynthesis. *Biochemistry* 49, 3667–3677.

Chan, Y. A., Podevels, A. M., Kevany, B. M., and Thomas, M. G. (2009) Biosynthesis of polyketide synthase extender units. *Nat. Prod. Rep.* 26, 90–114.

Conti, E., Stachelhaus, T., Marahiel, M. A., and Brick, P. (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J. 16*, 4174–4183.

Dayem, L. C., Carney, J. R., Santi, D. V., Pfeifer, B. A., Khosla, C., and Kealey, J. T. (2002) Metabolic engineering of a methylmalonyl-CoA mutase-epimerase pathway for complex polyketide biosynthesis in *Escherichia coli*. *Biochemistry* 41, 5193–5201.

DeLano, W. L. (2002) The PyMOL molecular graphics system user's manual (San Carlos, CA: DeLano Scientific).

Emsley, P., and Cowtan, K. (2004) COOT: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132.

Ferrer, J. L., Jex, J. M., Bowman, M. E., Dixon, R. A., and Noel, J. P. (1999) Structure of chalcone synthase and the molecular basis of plan polyketide biosynthesis. *Nat. Struct. Biol.* 6, 775-784.

Fortman, J. L., Sherman, D. H. (2005) Utilizing the power of microbial genetics to bridge the gap between the promise and the application of marine natural products. *ChemBioChem* 6, 960-978.

Jogl, G. and Tong, L. (2004) Crystal structure of yeast acetyl-coenzyme A synthetase in complex with AMP. *Biochemistry* 43, 1425–1431.

Gaitatzis, N., Silakowski, B., Kunze, B., Nordsiek, G., Blocker, H., Hofle, G., and Muller, R. (2002) The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase. *J. Biol. Chem.* 277, 13082-13090.

Gao, X., Wang, P., and Tang, Y. (2010) Engineered polyketide biosynthesis and biocatalysis in *Escherichia coli*. *Appl. Microbiol*. *Biotechnol*. 88, 1233-1242.

Gokhale, R. S., Tsuji, S. Y., Cane, D. E., and Khosla, C. (1999) Dissecting and exploiting intermodular communication in polyketide synthases. *Science* 284, 482–485.

Gulick, A. M., Starai, V. J., Horswill, A. R., Homick, K. M., and Escalante-Semerena, J. C. (2003) The 1.75Å crystal structure of acetyl-CoA synthetase bound to adenosine-5-propylphosphate and coenzyme A. *Biochemistry* 42, 2866–2873.

Gulick, A. M., Lu, X., and Dunaway-Mariano, D. (2004) Crystal structure of 4-chlorobenzoate:CoA ligase/synthetase in the unliganded and aryl substrate-bound states. *Biochemistry* 43, 8670–8679.

Hanessian, S. (1993) Reflections on the total synthesis of nature products: art, craft, logic, and the chiron approach. *Pure Appl. Chem.* 65, 1189-1204.

Hansen, D. A., Rath, C. M., Eisman, E., Narayan, A., Kittendorf, J. D., Mortison, J. D., Yoon, Y. J., and Sherman, D. H. (2013) Biocatalytic synthesis of pikromycin, methymycin, neomethymycin, novamethymycin, and ketomethymycin. *J. Am. Chem. Soc.* in Press.

Hardt, I. H., Steinmetz, H., Gerth, K., Sasse, F., Reichenbach, H., and Hofle, H. (2001) New natural epothilones from *Sorangium cellulosum*, strains so ce90/B2 and so ce90/D13: isolation, structure elucidation, and SAR studies. *J. Nat. Prod.* 64, 847-856.

Harper, A. D., Bailey, C. B., Edwards, A. D., Detelich, J. F., and Keatinge-Clay, A. T. (2012) Preparative, *in vitro* biocatalysis of triketide lactone chiral building blocks. *ChemBioChem 13*, 2200-2203.

Harvey, C. J. B., Puglisi, J. D., Pande, V. S., Cane, D. E., and Khosla, C. (2012) Precursor directed biosynthesis of an orthogonally functional erythromycin analogue: selectivity in the ribosome macrolide binding pocket. *J. Am. Chem. Soc.* 134, 12259-12265.

He, J., and Hertweck, C. (2003) Iteration as programmed event during polyketide assembly; molecular analysis of the aureothin biosynthesis gene cluster. *Chem. Biol.* 10, 1225-1232.

Herzberg, O., and Moult, J. (1991) Analysis of the steric strain in the polypeptide backbone of protein molecules. *Proteins* 11, 223–229.

Hisanaga, Y., Ago, H., Nakagawa, N., Hamada, K., Ida, K., Yamamoto, M., Hori, T., Arii, Y., Sugahara, M., Kuramitsu, S., Yokoyama, S., and Miyano, M. (2004) Structural basis of the substrate-specific two-step catalysis of long chain fatty acyl-CoA synthetase dimer. *J. Biol. Chem.* 279, 31717–31726.

Hughes, A. J. and Keatinge-Clay, A. T. (2011) Enzymatic extender unit generation for *in vitro* polyketide synthase reactions: structural and functional showcasing of *Streptomyces coelicolor* MatB. *Chem. Biol.* 18, 165-176.

Hughes, A. J., Detelich, J. F., and Keatinge-Clay, A. T. (2012) Employing a polyketide synthase module and thioesterase in the semipreparative biocatalysis of diverse triketide pyrones. *MedChemComm* 3, 956-959.

Kao, C. M., McPherson, M., McDaniel, R. N., Fu, H., Cane, D. E., and Khosla, C. (1998) Alcohol stereochemistry in polyketide backbones is controlled by the ketoreductase domains of modular polyketide synthases. *J. Am. Chem. Soc.* 120, 2478-2479.

Keatinge-Clay, A. T. (2007) A tylosin ketoreductase reveals how chirality is determined in polyketides. *Chem. Biol.* 14, 898–908.

Keatinge-Clay, A. T. (2012) The structures of type I polyketide synthases. *Nat. Prod. Rep.* 29, 1050-1073.

Khosla, C., Tang, Y., Chen, A. Y., Schnarr, N. A., and Cane, D. E. (2007) Structure and mechanism of the 6-deoxyerythronolide B synthase. *Annu. Rev. Biochem.* 76, 195–221.

Kochan, G., Pilka, E. S., von Delft, F., Oppermann, U., and Yue, W. W. (2009) Structural snapshots for the conformation-dependent catalysis by human medium-chain acyl-coenzyme A synthetase ACSM2A. *J. Mol. Biol.* 388, 997–1008.

Laughlin, S. T., Baskin, J. M., Amacher, S. L., and Bertozzi, C. R. (2008) *In vivo* imaging of membrane-associated glycans in developing zebrafish. *Science* 320, 664-667.

Lee, T. V., Johnson, L. J., Johnson, R. D., Koulman, A., Lane, G. A., Lott, J. S., and Arcus, V. L. (2010) Structure of a eukaryotic nonribosomal peptide synthetase

adenylation domain that activates a large hydroxamate amino acid in siderophore biosynthesis. *J. Biol. Chem.* 285, 2415–2427.

Liu, H. and Naismith, J. (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotech*. 8, 91-100.

Lu, X. and Bittman, R. (2005) Synthesis of a photoactivatable (2*S*,3*R*)-sphingosylphosphorylcholine analogue. *J. Org. Chem.* 70, 4746-4750.

Marahiel, M. A. (2009) Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis. *J. Pept. Sci.* 15, 799–807.

Marahiel, M. A., Stachelhaus, T., and Mootz, H. D. (1997) Modular peptide synthesis involved in nonribosomal peptide synthesis. *Chem. Rev.* 97, 2651–2674.

Menzella, H. G., Reid, R., Carney, J. R., Chandran, S. S., Reisinger, S. J., Patel, K. G., Hopwood, D. A., and Santi, D. V. (2005) Combinatorial polyketide biosynthesis by *de novo* design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* 23, 1171–1176.

Mochizuki, S., Hiratsu, K., Suwa, M., Ishii, T., Sugino, F., Yamada, K., and Kinashi, H. (2003) The large linear plasmid pSLA2-L of *Streptomyces rochei* has an unusually condensed gene organization for secondary metabolism. *Mol. Microbiol.* 48, 1501-1510.

Mootz, H. D., Schwarzer, D., and Marahiel, M. A. (2002) Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *ChemBioChem 3*, 490–504.

Murli, S., Kennedy, J., Dayem, L. C., Carney, J. R., and Kealey, J. T. (2003) Metabolic engineering of *Escherichia coli* for improved 6-deoxyerythronolide B production. *J. Ind. Microbiol. Biotechnol.* 30, 500–509.

Nakatsu, T., Ichiyama, S., Hiratake, J., Saldanha, A., Kobashi, N., Sakata, K., and Kato, H. (2006) Structural basis for the spectral difference in luciferase bioluminescence. *Nature* 440, 372–376.

Olano, C., Moss, S.J., Braña, A. F., Sheridan, R. M., Math, V., Weston, A. J., Méndez, C., Leadlay, P. F., Wilkinson, B., and Salas, J. A. (2004) Biosynthesis of the angiogenesis inhibitor borrelidin by *Streptomyces parvulus* Tü4055: insights into nitrile formation. *Mol. Microbiol.* 52, 1745-1756.

Oliynyk, M., Brown, M. J. B., Cortès, J. S., Staunton, J., and Leadlay, P. F. (1996) A hybrid modular polyketide synthase obtained by domain swapping. *Chem. Biol. 3*, 833-839.

Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*. 276, 307–326.

Pfeifer, B. A., and Khosla, C. (2001) Biosynthesis of polyketides in heterologous hosts. *Microbiol. Mol. Biol. Rev.* 65, 106–118.

Piasecki, S. K., Taylor, C. A., Detelich, J. F., Liu, J., Zheng, J., Komsoukaniants, A., Siegel, D. R., Keatinge-Clay, A. T. (2011) Employing modular polyketide synthase ketoreductases as biocatalysts in the preparative chemoenzymatic syntheses of diketide chiral building blocks. *Chem. Biol.* 18, 1331-1340.

Piasecki, S. K. and Keatinge-Clay, A. T. (2012) Monitoring biocatalytic transformations mediated by polyketide synthase enzymes in cell lysate via fluorine NMR. *SynLett* 23, 1840-1842.

Pieper, R., Ebert-Khosla, S., Cane, D. E., and Khosla, C. (1996) Erythromycin biosynthesis: kinetic studies on a fully active modular polyketide synthase using natural and unnatural substrates. *Biochemistry* 35, 2054-2060.

Pohl, N. L., Gokhale, R. S., Cane, D. E., and Khosla, C. (1998) Synthesis and incorporation of an *N*-acetylcysteamine analogue of methylmalonyl-CoA by a modular polyketide synthase. *J. Am. Chem. Soc.* 120, 11206–11207.

Pohl, N. L., Hans, M., Lee, H. Y., Kim, Y. S., Cane, D. E., and Khosla, C. (2001) Remarkably broad substrate tolerance of malonyl-CoA synthetase, an enzyme capable of intracellular synthesis of polyketide precursors. *J. Am. Chem. Soc.* 123, 5822–5823.

Prasad, G., Borketey, L. S., Lin, T.-Y., and Schnarr, N. A. (2012) A mechanism-based fluorescence transfer assay for examining ketosynthase selectivity. *Org. Biomol. Chem.* 10, 6717-6723.

Reeves, C. D., Hu, Z., Reid, R., and Kealey, J. T. (2008) Genes for the biosynthesis of the fungal polyketides hypothemycin from *Hypomyces subiculosus* and radicicol from *Pochonia chlamydosporia*. *Appl. Environ*. *Microbiol*. 74, 5121-5129.

Regentin, R., Kennedy, J., Wu, N., Carney, J. R., Licari, P., Galazzo, J., and Desai, R. (2004) Precursor-directed biosynthesis of novel triketide lactones. *Biotechnol. Prog.* 20, 122-127.

Reger, A. S., Carney, J. M., and Gulick, A. M. (2007) Biochemical and crystallographic analysis of substrate binding and conformational changes in acetyl- CoA synthetase. *Biochemistry* 46, 6536–6546.

Reger, A. S., Wu, R., Dunaway-Mariano, D., and Gulick, A. M. (2008) Structural characterization of a 140 degrees domain movement in the two-step reaction catalyzed by 4-chlorobenzoate:CoA ligase. *Biochemistry* 47, 8016–8025.

Rostovtsev, V. V., Green, L. G., Fokin, V. V., and Sharpless, K. B. (2002) A stepwise Huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem. Int. Ed.* 41, 2596-2599.

Shah, M. B., Ingram-Smith, C., Cooper, L. L., Qu, J., Meng, Y., Smith, K. S., and Gulick, A. M. (2009) The 2.1Å crystal structure of an acyl-CoA synthetase from *Methanosarcina acetivorans* reveals an alternate acyl-binding pocket for small branched acyl substrates. *Proteins* 77, 685–698.

Sharma, K. K., and Boddy, C. N. (2007) The thioesterase domain from the pimaricin and erythromycin biosynthetic pathways can catalyze hydrolysis of simple thioester substrates. *Bioorg. Med. Chem. Lett.* 17, 3034-3037.

Siskos, A. P., Baerga-Ortiz, A., Bali, S., Stein, V., Mamdani, H., Spiteller, D., Popovic, B., Spencer, J. B., Staunton, J., Weissman, K. J., and Leadlay, P. F. (2005) Molecular basis of Celmer's rules: stereochemistry of catalysis by isolated ketoreductase domains from modular polyketide synthases. *Chem. Biol.* 12, 1145-1153.

Smith, S. and Tsai, S. C. (2007) The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat. Prod. Rep.* 24, 1041-1072.

Stachelhaus, T., Mootz, H. D., and Marahiel, M. A. (1999) The specificity conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* 6, 493–505.

Staunton, J., and Weissman K. J. (2001a) Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep. 18*, 380-416.

Staunton, J., and Wilkinson, B. (2001b) Combinatorial biosynthesis of polyketides and nonribosomal peptides. *Curr. Opin. Chem. Biol.* 5, 159–164.

Strieker, M., Tanovic, A., and Marahiel, M. A. (2010) Nonribosomal peptide synthetases: structures and dynamics. *Curr. Opin. Struct. Biol.* 20, 234–240.

Sundermann, U., Bravo-Rodriguez, K., Klopries, S., Kushnir, S., Gomez, H., Sanchez-Garcia, E., and Schulz, F. (2012) Enzyme-directed mutasynthesis: a combined

experimental and theoretical approach to substrate recognition of a polyketide synthase. *ACS Chem. Biol.* 8, 443-450.

Tang, Y., Kim, C. Y., Mathews, I. I., Cane, D. E., and Khosla, C. (2006) The 2.7Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. USA 103*, 11124–11129.

Valenzano, C. R., Lawson, R. J., Chen, A. Y., Khosla, C., and Cane, D. E. (2009) The biochemical basis for stereochemical control in polyketide biosynthesis. *J. Am. Chem. Soc.* 131, 18501-18511.

Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* 8, 127–134.

Walsh, C. T. (2004) Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science* 303, 1805–1810.

Wang, M., Opare, P., and Boddy, C. N. (2009) Polyketide synthase thioesterases catalyze rapid hydrolysis of peptidyl thioesters. *Bioorg. Med. Chem. Lett.* 19, 1413–1415.

Wilkinson, B., Foster, G. G., Rudd, B. A. M., Taylor, N. L., Blackaby, A. P., Sidebottom, P. J., Cooper, D. J., Dawson, M. J., Buss, A. D., Gaisser, S., Bohm, I. U., Rowe, C. J., Cortes, J. S., Leadlay, P. F., and Staunton, J. (2000) Novel octaketide macrolides related to 6-deoxyerythronolide B provide evidence for iterative operation of the erythromycin polyketide synthase. *Chem. Biol.* 7, 111-117.

Wong, C., Drueckhammer, D. G., and Sweers, H. M. (1985) Enzymatic vs. fermentative synthesis: thermostable glucose dehydrogenase catalyzed regeneration of NAD(P)H for use in enzymatic synthesis. *J. Am. Chem. Soc.* 107, 4028-4031.

Wu, N., Kudo, F., Cane, D. E., and Khosla, C. (2000) The FK520 gene cluster of *Streptomyces hygroscopicus* var. ascomyceticus (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. *J. Am. Chem. Soc.* 122, 4847-4852.

Wu, R., Cao, J., Lu, X., Reger, A. S., Gulick, A. M., and Dunaway-Mariano, D. (2008) Mechanism of 4-chlorobenzoate:coenzyme A ligase catalysis. *Biochemistry* 47, 8026–8039.

Yonus, H., Neumann, P., Zimmermann, S., May, J. J., Marahiel, M. A., and Stubbs, M. T. (2008) Crystal structure of DltA. *J. Biol. Chem.* 283, 32484–32491.