The Thesis Committee for Andrew David Harper
Certifies that this is the approved version of the following thesis:

In Vitro Polyketide Biocatalysis: Triketide Building-Blocks and Enzymology

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Adrian Keatinge-Clay

David Hoffman

In vitro Polyketide Biocatalysis: Triketide Building-Blocks and Enzymology

by

Andrew David Harper, BS

Thesis

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

Master of Arts

The University of Texas at Austin

May 2013

Acknowledgements

I would like to thank Adrian Keatinge-Clay for his invaluable help, support, and advice. It has been a privilege to work for and learn from him. It has also been an honor to work with each of the other Keatinge-Clay lab members.

I owe Jianting Zheng immense thanks for years of tutelage regarding Actinomycetes, molecular biology, natural products, and anything else I asked about.

I would like to thank Constance Bailey for her collaboration with me on experiments and writing, as well as generously sharing her broader knowledge of chemistry with me. She increased the yields in this project and aided in purification and characterization of triketide lactones discussed here.

I would like to thank Amanda Hughes and Shawn Piasecki. My work would not be possible without theirs, and I always found them transparent and generous collaborators. Amanda Hughes also trained me in chemical syntheses and use of her fluorescent HPLC assay.

Thanks to Darren Gay for help with troubleshooting and designing experiments. His insights and encouragement were invaluable.

I must thank the outstanding students who chose to work with me as part of their undergraduate experience. Angela Edwards aided in initial development of this preparative, cell-free polyketide biocatalytic system. Andrew Tran and Seong-Meen Yoon aided in exploring the capabilities of the system once it was in place. Joshua Detelich donated compounds that he had synthesized.

I also owe many thanks to Chris Fage, Eta Isiorho-Mansoorabadi, Jessica Momb, and Abram Axelrod for their camaraderie and encouragement.

Abstract

In Vitro Polyketide Biocatalysis: Triketide Building-Blocks and Enzymology

Andrew David Harper, MA

The University of Texas at Austin, 2013

Supervisor: Adrian Keatinge-Clay

Polyketide products are useful compounds to research and industry but can be difficult to access due to their richness in stereogenic centers. Type I polyketide synthases offer unique engineering opportunities to access natural stereocontrol and resultant complex compounds. The development of a controlled *in vitro* platform based around type I polyketide synthases is described. It has been used to produce a small library of polyketide fragments on an unprecedented and synthetically-relevant scale and explore polyketide synthase enzymology.

Table of Contents

List of Tables

List of Figures

Introduction

POLYKETIDES AND THEIR SYNTHASES

Polyketides are structurally diverse secondary metabolites whose broad utility has generated intense interest in their production. Type I polyketide synthases (PKSs) have their many catalytically independent enzymes fused into multidomain assembly-line stations, called modules. At each module, a defined set of chemical reactions is performed before passing the product of that module to the next module. Modules and enzymes within a module cooperate in series to synthesize diverse molecules, often rich in stereogenic centers (Weissman 2009). In this way, the arrangement of domains with the gene cluster controls the content and layout of the synthase and thus the structure of the product (Hertweck 2009; Cane 2010). The metabolic cost entailed in encoding and producing these very long lines of enzymes must be justified by evolutionary advantages for the organisms that produce them. In addition to the direct advantage of a compound produced by any specific type I PKS, the modular configuration of these biosynthetic gene clusters confers the advantage of rapid evolution of new compounds. This is because they are amenable to duplication, insertion, deletion, and rearrangement to produce new compounds in a way that non-modular biosynthetic clusters are not (Fischbach et al. 2008).

Polyketide synthases in this project generally assemble their complex scaffolds from simple propionyl groups to begin the assembly line and a methylmalonyl group to extend the product at each module. The domains relevant to this thesis include the ketosynthase (KS), the acyltransferase (AT), the acyl carrier protein (ACP), the ketoreductase (KR), and the thioesterase (TE).

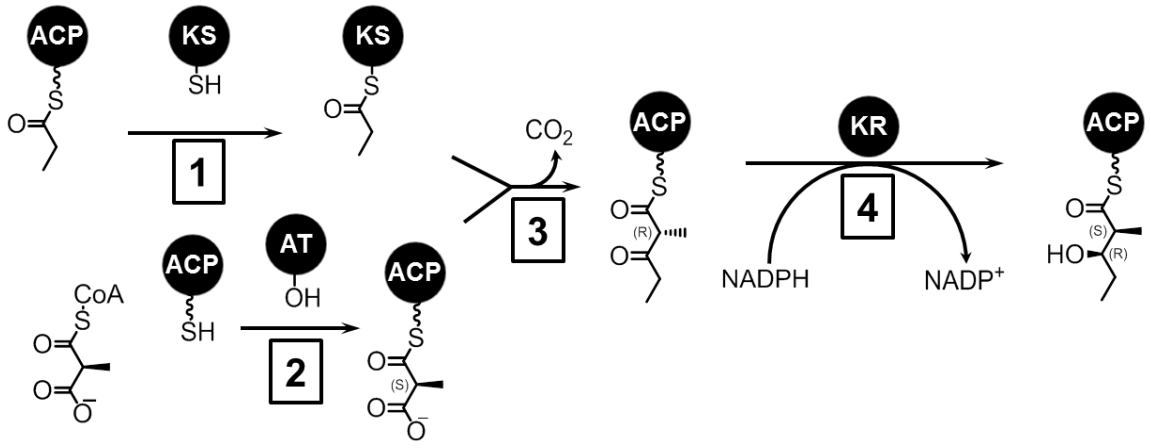

Figure 1. Select steps in polyketide biosynthesis.

The steps of polyketide biosynthesis most relevant to this project are (figure 1, Adapted from Cane 2010):

1. the KS is acylated by an upstream ACP
2. the AT attaches a methylmalonyl group to the ACP as a thioester
3. the KS catalyzes a decarboxylative condensation to extend the product, leaving it attached to the ACP as a β-keto-thioester
4. the KR may at this point reduce the β-keto group, simultaneously setting stereochemical configuration of α- and β-carbons. The product shown here is B2-type (Keatinge-Clay 2007).
5. (not shown) this process may be repeated by downstream modules until the resultant scaffold reaches a terminal TE which releases is from the ACP.

There are two different ways in which type I PKS gene clusters are valuable to the goal of broadening access to untapped biochemical abilities. Firstly, they code for the ability to manufacture a huge number of pre-evolved, bioactive small molecules. Polyketides and their derivatives have been repeatedly reappropriated from toxins into powerful and successful agents in pharmacy, agriculture, and biotechnology based on

their antibacterial, antifungal, anticancer, insecticide, and other bioactive properties. Secondly, type I PKS gene clusters arrange all the reactions necessary to polyketides in a customizable configuration. This could provide access to customized polyketide products as well as natural ones. It has long been a goal of investigators to use polyketide synthases to access new polyketides, designed to order, by engineering their synthases (Weissman and Leadlay 2005).

CHALLENGES IN ACCESSING POLYKETIDE PRODUCTS

Access to a library of thousands of pre-evolved, presumably bioactive compounds is a worthwhile goal. The most direct access route would be to simply purify these polyketides from organisms that are already producing them. Progress along this route has been frustrated by the cryptic origins of many polyketides, the often limited feasibility of culturing the organism once it is discovered, and the notoriously low titers of the desired compounds produced in culture.

The most well studied polyketide producing system is certainly the erythromycin synthase (Khosla et al. 2007). However, even within this model system, fundamental concepts such as interactions within the synthase (Kapur et al. 2010; Kapur et al. 2012) or rational routes to *in vivo* overproduction (Lum et al. 2004; Peano et al. 2012) are only beginning to be understood. Given this *status quo* in the model system of type I PKSs, simultaneously overcoming all the same hurdles in thousands of polyketide producing organisms is not currently a practical approach. Current heterologous PKS production technologies in *E. coli* also suffer from similarly impractical yields and so offer no immediate help (Zhang et al. 2010).

Chemical synthesis offers an organism-independent route to utilizing polyketide products. But while these complex total syntheses offer enriching challenges to synthetic

chemists, their large scale syntheses often prove challenging and can preclude the development of even the most promising compounds (Wender et al. 2011).

THE IMPORTANCE OF CHIRAL BUILDING BLOCKS

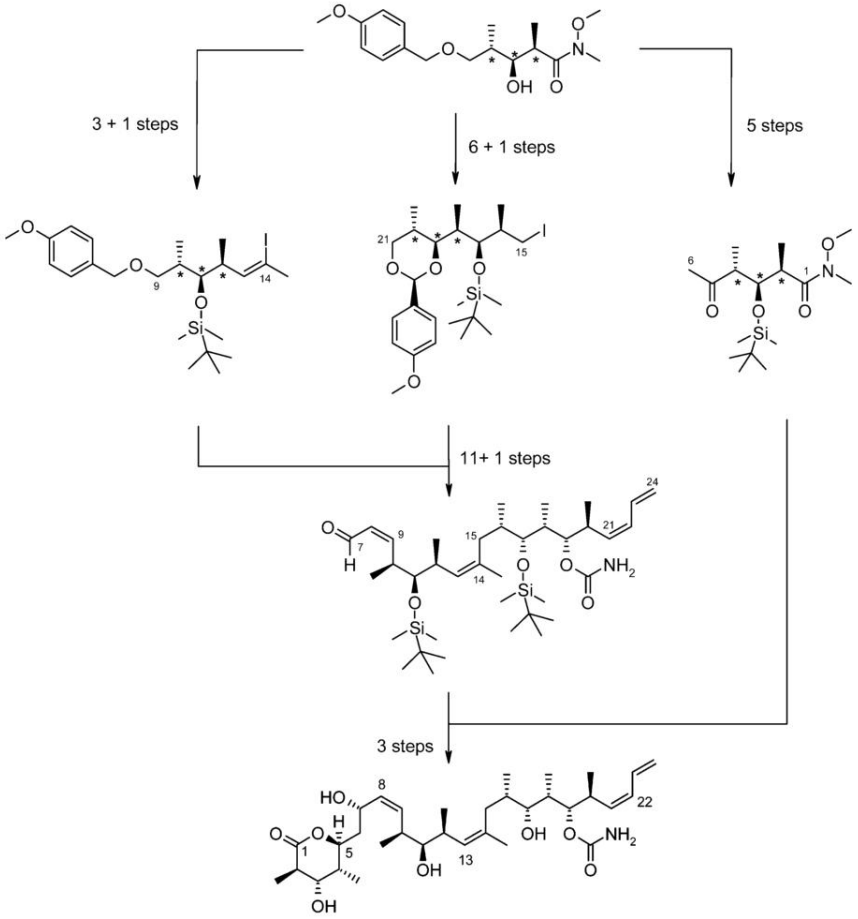

Figure 2. The use of chiral intermediates in a large-scale discodermolide total synthesis.

The large-scale synthesis of discodermolide (Figure 2, excerpted from Mickel et al. 2004) demonstrates the remarkable utility of chiral intermediates in the synthesis of complex polyketide targets. However, the chiral pool of such intermediates is usually limited to sugars, amino acids, and Roche esters (Hilterhaus and Liese 2007) which are not always suitable. Polyketide fragments or molecules resembling them would be more

relevant. In addition to usefulness as chiral precursors in the synthesis of natural polyketides, such polyketide fragments would be rich in stereogenic sp^3 carbons and therefore desirable for building new molecules with these qualities, which are more likely to be specific protein binders (Clemons et al. 2010).

Simultaneously inspired by the challenges in synthesizing complex polyketide-related molecules, and the stereocontrol and accessibility of polyketide synthases as catalysts, we set about to assemble an *in vitro* polyketide biosynthetic platform. We did this simultaneously produce chiral intermediates for more complex syntheses and to better understand basic questions about these remarkable catalysts.

THE CURRENT STATE OF *IN VITRO* POLYKETIDE SYNTHESIS

Structural exploration has increased insight into PKS function at the domain level (Khosla et al. 2007) but many questions about mechanisms and interactions within the context of a modular or multimodular system remain unanswered. Dissecting the immense complexity of these so-called "megasynthases" is certainly no small task, but *in vitro* experiments have also faced frustrating limitations leading many studies to explore questions amenable to *in vivo* strategies.

One of these limitations is the complex biological cofactors used by these synthases. Their simple carbon extender units are delivered *in vivo* as CoA thioesters, which are difficult to access synthetically and costly to purify. This is reflected in their commercial price. Methylmalonyl-CoA, for example, is available from Sigma Aldrich for approximately $75 per milligram. A ^{14}C-labelled version is also available, but its cost is not publicly available. Likewise the reducing power of polyketide synthases is supplied by the complex cofactor NADPH, which costs several dollars per milligram and is generally used stoichiometrically.

This limited accessibility of substrates leads investigators to make *in vitro* PKS experiments extremely small scale and necessarily reliant on detection methods sensitive to micrograms to nanograms of product (Kapur et al. 2010; Kapur et al. 2012). These high-sensitivity methods are generally low-throughput and do not provide rigorous characterization of products.

The largest exception in the field is the common use of N-acetyl-cysteamine (NAC) thioesters as a truncated mimics of upstream acyl-ACPs in acylating the active sites of KS domains (Kinoshita et al. 2003; Lee et al. 2011; Zhang et al. 2010). However, it is not common practice to use a NAC thioester to acylate the active site of an AT domain as a truncated CoA mimic. While one AT has been shown to accept extender units linked to NAC instead of CoA (i.e. methylmalonyl-*S*-NAC instead of methylmalonyl-*S*-CoA) (Pohl et al. 1998), these NAC-based extender units are not commercially available, and the practice has not yet been commonly adopted.

PREVIOUS *IN VITRO* PKS ACCOMPLISHMENTS OF THE KEATINGE-CLAY GROUP

During an attempt to use the *Streptomyces coelicolor* malonyl-CoA ligase (MatB) to generate methylmalonyl-CoA *in vitro*, Hughes and Keatinge-Clay (2011) discovered that this promiscuous enzyme could generate a not only methylmalonyl-CoA but also methylmalonyl-pantetheine and methylmalonyl-*S*-NAC. Moreover, this NAC-based extender unit was used by the terminal module and thioesterase of the erythromycin synthase (EryMod6TE) to extend a diketide-*S*-NAC to produce an achiral triketide pyrone. It was also demonstrated that EryMod6TE can produce a variety of achiral triketide pyrones by both extending a variety of non-natural β-keto-diketides (Figure 3) and perform two extensions on monoketide NAC thioesters using NAC-based extender units (Hughes et al. 2012).

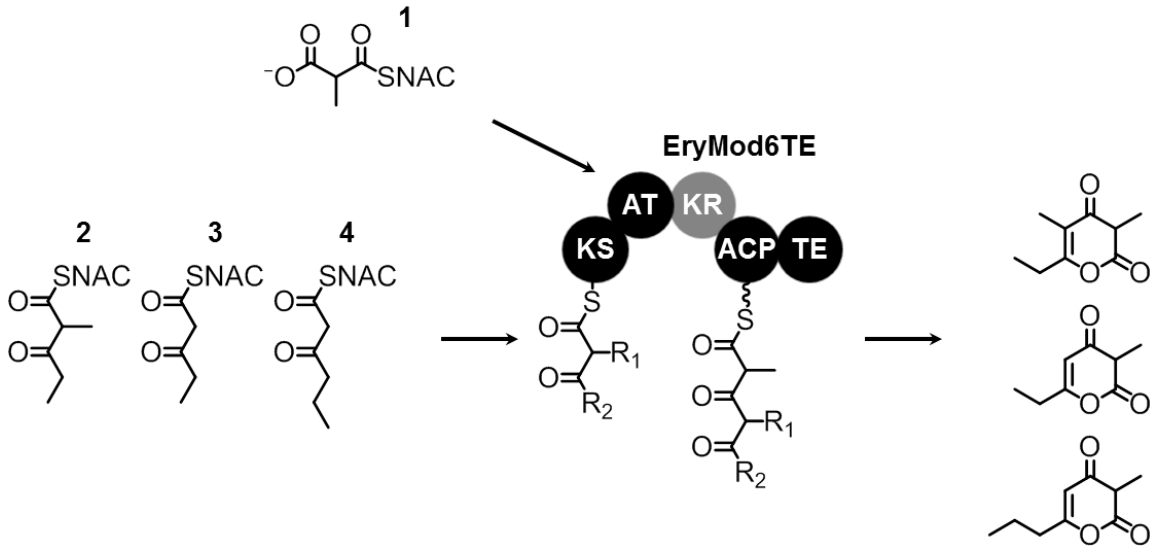

Figure 3. Select pyrones generated by Hughes et al. 2012.

The Keatinge-Clay group was simultaneously investigating KR domains, already shown to be active outside the context of their modules (Siskos et al. 2005). Piasecki et al. (2011) found that a variety of isolated KRs were highly active and exert stereocontrol on diketide-*S*-NAC substrate analogues while reducing them from β-ketoacyl-*S*-NACs to chiral β-hydroxyacyl-*S*-NACs. Additionally, it was shown that NADPH can be supplied to these KRs by regenerating it thousands of times from $NADP^+$ using glucose and *Bacillus subtilis* glucose dehydrogenase (GDH) (Piasecki et al. 2011). This approach drove ketoreduction towards completion, resulting in no measurable residual substrate. Not only could the reactions be initiated with the more economical $NADP^+$, the ratio for cofactor concentration was no longer stoichiometric, making KR biocatalysis far more practical. The concentration of $NADP^+$ was simply maintained at the minimal concentration sufficient for completion of the reactions. In scaling up these reactions to higher concentrations, Piasecki et al. (2011) also demonstrated the utility of dialyzed cell

lysate in place of purified enzyme. In these reactions, dialyzed lysate is equally effective but less time- and resource-intensive to produce than purified enzyme.

Results and Discussion

MODIFICATION OF MATB AND KR REACTIONS

In order to create synthetically-meaningful quantities of polyketide fragments *in vitro*, MatB reactions were concentrated 20-fold, from 5mM to 100mM, to make the most use of the enzyme. The buffering power was also increased and initial pH of the reactions raised to compensate for the release of pyrophosphoric acid during the reaction. Reactions were carried out on the gram scale (5.00 mmol, 1.17 g theoretical yield). Accessibility of MatB products was also greatly increased by locating an economical source of ATP (Meiya Pharmaceuticals $0.20/g) which is generated by an enzymatic method, rather than purifying from an organism.

The KR reactions were likewise concentrated to 100mM and buffer concentrations raised to compensate for the release of gluconic acid. A variety of β-keto-diketide-*S*-NACs known to be extended by EryMod6TE (Hughes at al. 2012) were selected to be reduced to β-hydroxy-diketide-*S*-NAC by a suitable ketoreductase (Piasecki et al. 2011). In this way, chiral diketide-*S*-NACs **5-7** (Figure 4) were produced in separate reactions up to the gram scale.

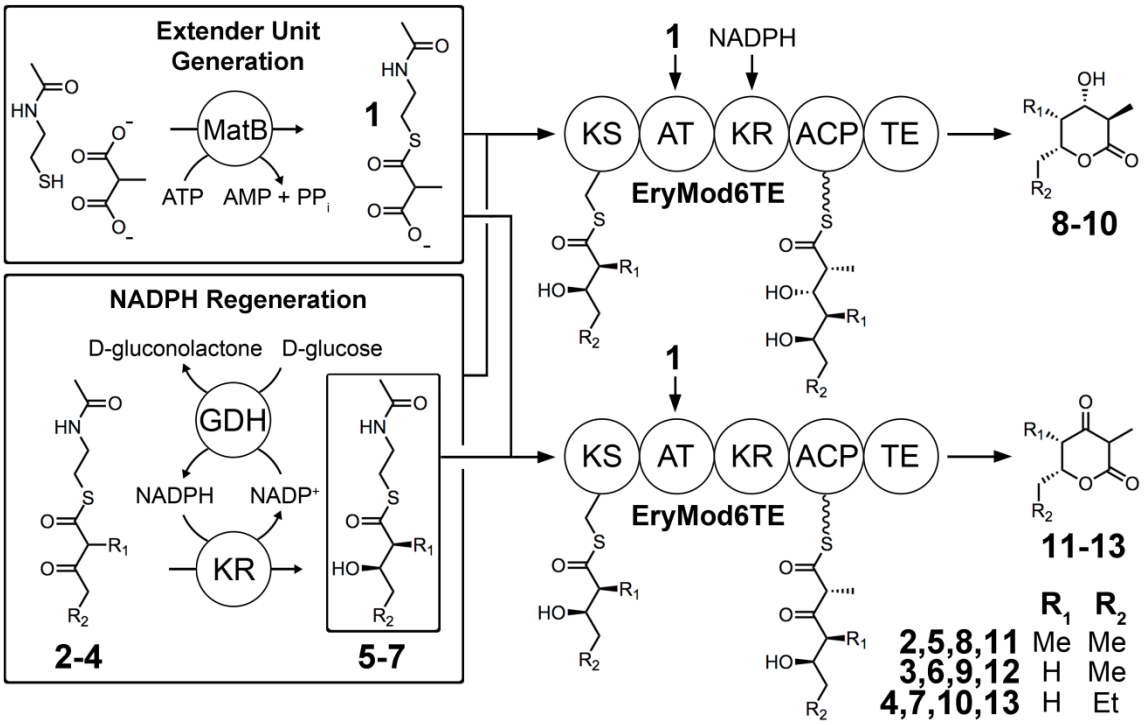

Figure 4. Scheme for *in vitro* polyketide biosynthesis.

GENERATION OF TRIKETIDE LACTONES BY ERYMOD6TE

EryMod6TE had been shown to produce a chiral triketide lactone containing four contiguous stereocenters (**8**) *in vitro*, although only in trace quantities (Castonguay et al. 2007; Chen et al. 2007; Oliynyk et al. 1996; Wiesmann et al. 1995). However, unprecedented amounts of triketide lactones **8** and **11** were generated when dialyzed lysate containing EryMod6TE was supplemented with diketide-*S*-NAC **5** and methylmalonyl-*S*-NAC (**1**) from KR and MatB reactions, respectively (Table 1). Additionally, EryMod6TE promiscuity allowed precursor-directed control of substituents at the R_1 and R_2 positions allowing the generation new lactones **9** and **10**, as well as their ketolactones **12** and **13**. Continued action of the NADPH-regeneration system from the KR reactions allowed the KR of EryMod6TE (EryKR6) to reduce the β-ketoacyl-ACP intermediates to produce reduced triketide lactones **8-10**. If reduced diketides **5-7** were

separated from the NADPH-regeneration system by ethyl acetate (EtOAc) extraction before being added to EryMod6TE lysate, ketolactones **11**-**13** were formed (Table 1).

Diketide precursor	Scale [mmol]	NADPH-regeneration system	Product	Product mass [mg]	Isolated yield [%]
2	0.5	+	**8** **11**	6.3 11	7 13
3	0.5	+	**9**	7.3	9
3	0.5	-	**12**	5.6	7
4	0.5	+	**10**	8.4	10
4	0.5	-	**13**	3.4	4
4	7	+	**10** **13**	77.0 14.1	7 1

Table 1. Biocatalytic syntheses of triketide lactones

It is significant that the ketolactone **11** was the major product of the reaction generating the reduced lactone **8**. Both compounds result from extension of diketide **5** and so share a common α,γ-dimethyl-β-ketoacyl-ACP intermediate which was a more often acted upon by the TE than by the KR. Intermediates resulting from the extension of **6** and **7** were γ-unsubstituted and more often acted upon by the KR than by the TE as ketolactones **12** and **13** were only generated in trace quantities in the presence of NADPH. This previously undocumented substrate preference of EryKR6 in the model erythromycin PKS demonstrates the ability of this cell-free system to measure relative rates of competing reactions within a module.

The multimilligram amounts triketides produced allowed for new levels of characterization. Each compound was characterized by ^{13}C and ^{1}H NMR and MS. A small-molecule crystal structure of **11** was solved by Constance Bailey using triketide lactone produced *in vitro* (Figure 5).

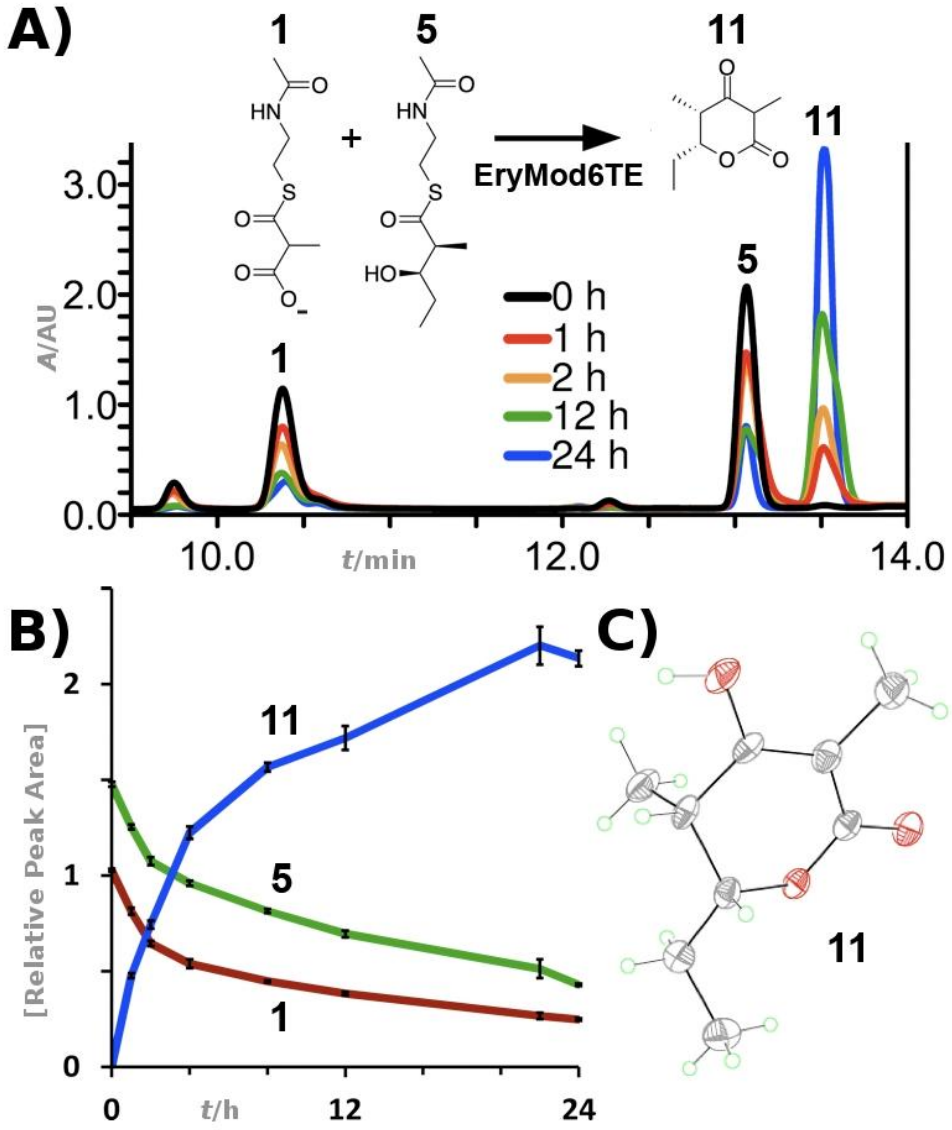

Figure 5. Analytical advantages of *in vitro* PKS biocatalysis on an unprecedented scale.

A) Stacked plot of an HPLC time course (λ=242 nm) of an EryMod6TE reaction measuring both reactants (**1** and **5**) and main product (**11**). Some timepoints have been omitted for readability. B) A plot of relative peak area (arbitrary units) versus time, following the same reaction. C) A crystal structure of **11** in its enol tautomer (crystal structure details available in Harper et al. 2012).

Monitoring progress of ketolactone-forming reactions with HPLC

Although the β-ketoesters of ketolactones **11**-**13** are weak chromophores (λ_{max}=248 nm), they were made in concentrations sufficient to track their formation by HPLC and photodiode array detector. Because β-hydroxyacyl-*S*-NACs and methylmalonyl-*S*-NAC are also UV active (λ_{max}=233 nm and λ_{max}=235 nm, respectively), all reactants and the desired product could be tracked through the course of reactions generating **11** (Figure 6), **12** (Figure 7), and **13** (Figure 8). Peak areas of all these species were charted through the course of 24 h reactions to better display the relative rate of their formation (Figure 9)

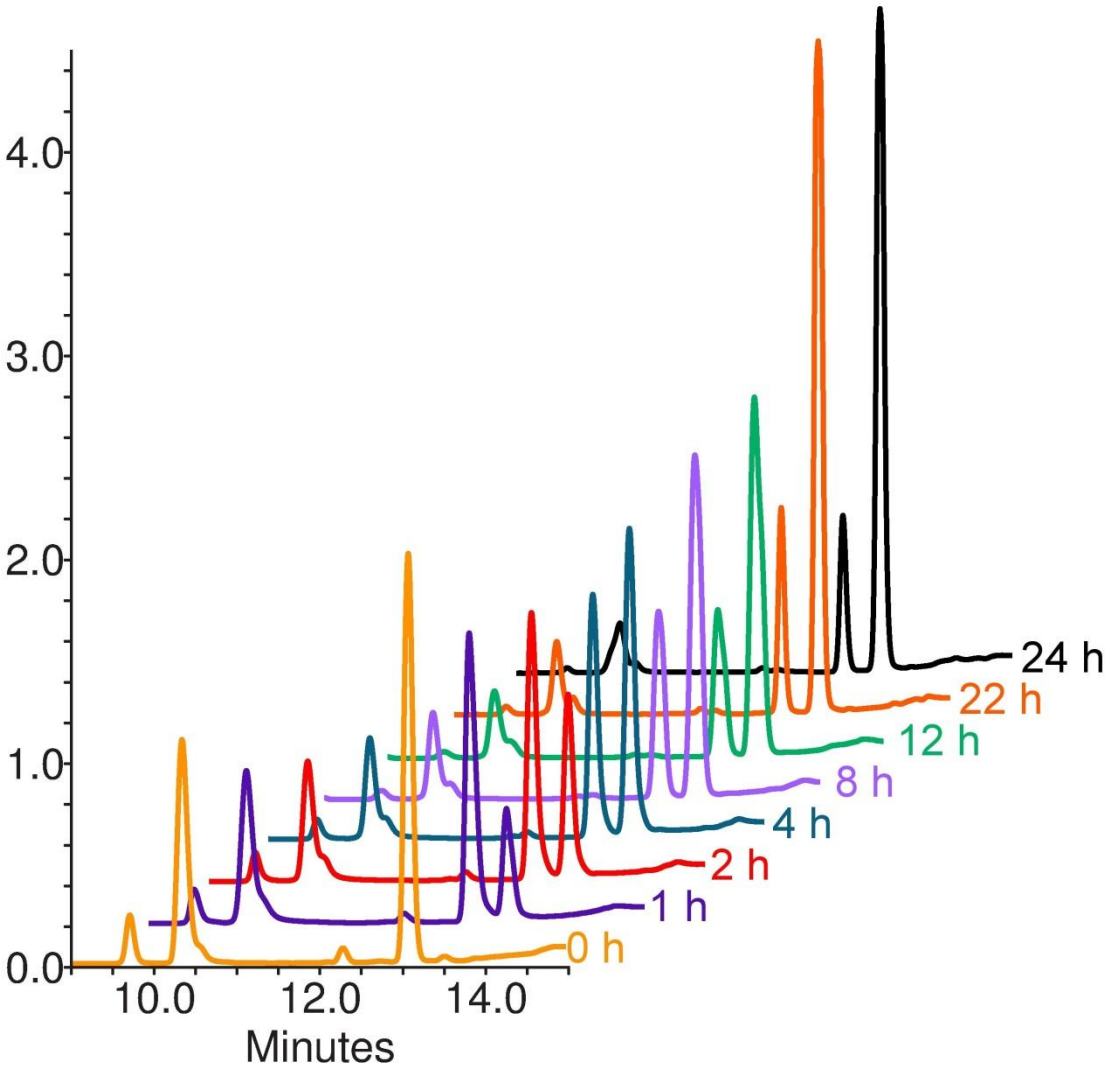

Figure 6. Stacked plot of an HPLC time course (λ=242 nm) of an EryMod6TE reaction generating **11** over 24 h.

A complete version of the abbreviated time course appearing in figure 5A. The decrease of methylmalonyl-*S*-NAC **1** and (2*S*,3*R*)-3-hydroxy-2-methylpentanoyl-*S*-NAC **5** as well as formation of **11** can be followed over the course of the reaction.

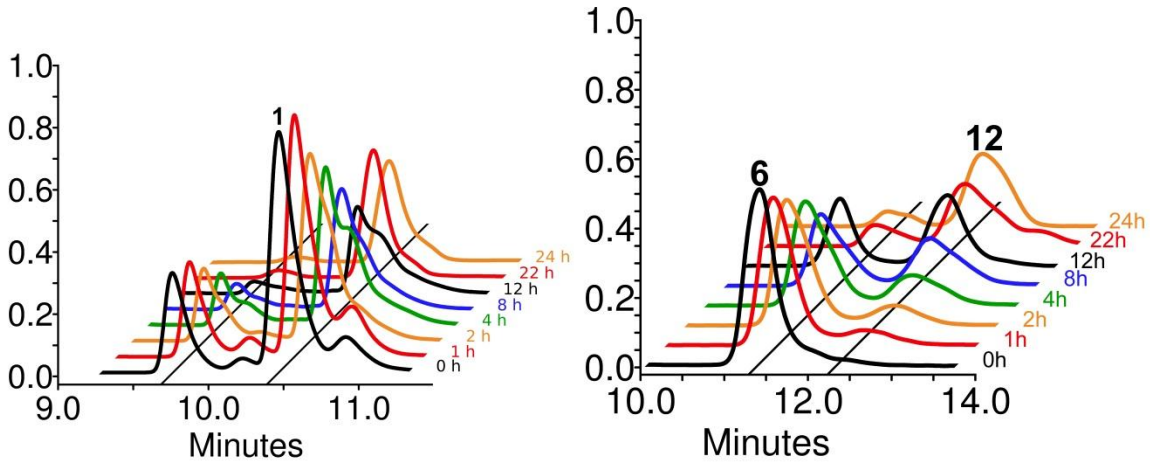

Figure 7. Stacked plot of an HPLC time course (λ=242 nm) of an EryMod6TE reaction generating **12** over 24 h.

The decrease of methylmalonyl-*S*-NAC **1** and (3*R*)-3-hydroxy-pentanoyl-*S*-NAC **6** as well as formation of **12** can be followed over the course of the reaction.

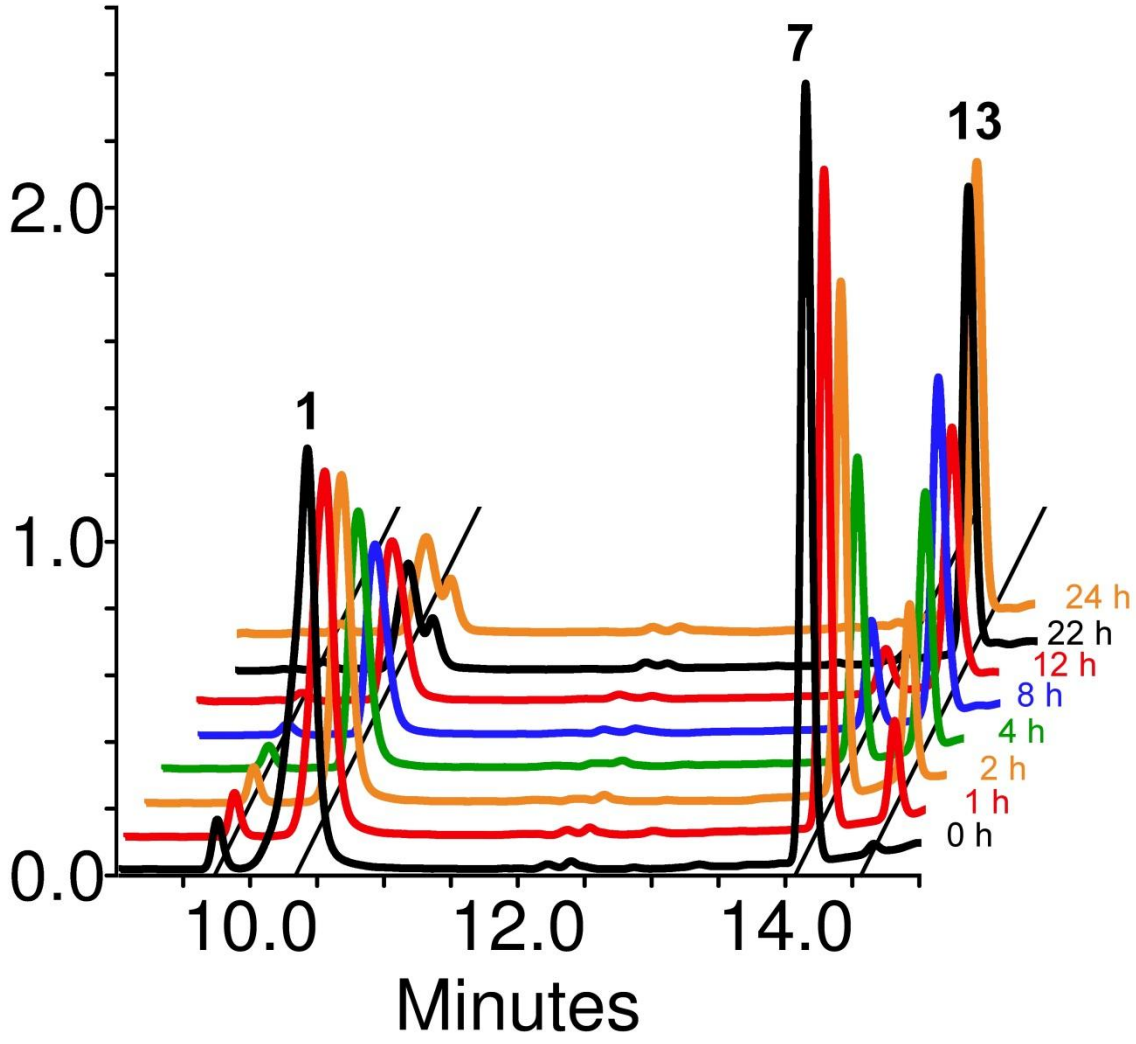

Figure 8. Stacked plot of an HPLC time course (λ=242 nm) of an EryMod6TE reaction generating **13** over 24 h.

The decrease of methylmalonyl-*S*-NAC **1** and (3*R*)-3-hydroxy-hexanoyl-*S*-NAC **7** and formation of **13** can be followed over the course of the reaction.

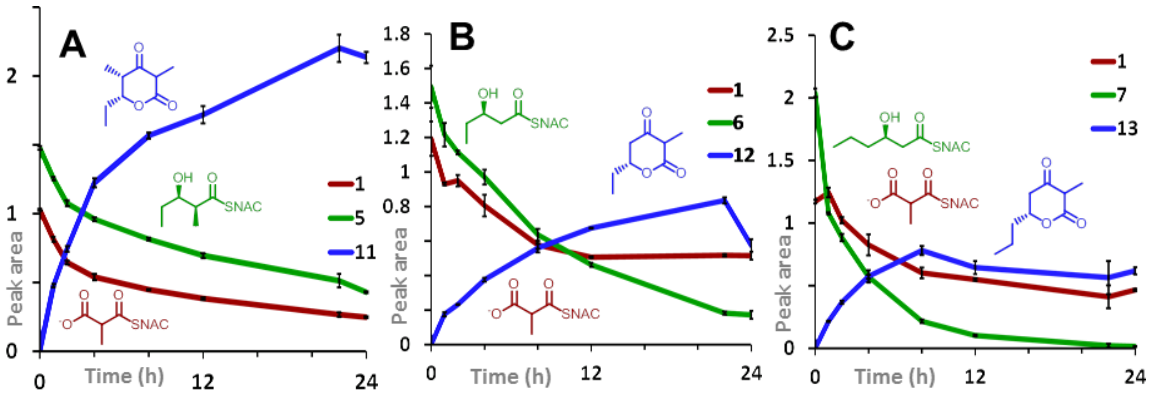

Figure 9. Triplicate time courses of all three ketolactone-producing reactions.

Plots of HPLC peak area (arbitrary units) over reaction time (hours). In each plot, the extender unit methylmalonyl-*S*-NAC 1 is in red, the reduced diketide *S*-NAC is in green, and the triketide ketolactone is in blue. Reactions were done in triplicate. A) Extension and cyclization of **5** by EryMod6TE to form **11**. B) Extension and cyclization of **6** by EryMod6TE to form **12**. C) Extension and cyclization of **7** by EryMod6TE to form **13**.

ERYMOD6TE STEREOCHEMICAL SUBSTRATE SELECTIVITY

The ability to track relevant chemical species through the course of ketolactone formation and the ability to generate not only **5**, but also its three stereoisomers (Piasecki et al. 2011) was used to explore the stereoselectivity of EryMod6TE. An isolated KR from each class was chosen based on its activity against NAC-based substrates: AmpKR2 for A1-type product, RifKR7 for A2-type product, TylKR1 for B1-Type product, and EryKR1 for B2-type product (Figure 10). EryMod6TE demonstrated strict selectivity for the B2-type product. This is not unprecedented among ModTEs, **5** and its stereoisomers could all acylate the KS of EryMod2TE, but again only the B2-type **5** was extended (Wu et al. 2004). EryMod2 acts on a B2-type product. It is surprising that EryMod6TE only extends B2-type products because in the synthesis of erythromycin it extends an A1-type product, which possesses the opposite stereochemistry at the α- and β-carbons,. It is possible that the substrate selectivity of EryKS6 is altered outside the context of the synthase or against NAC-based substrates (rather than its natural ACP-based substrate). It is also possible that this stereoselectivity is enforced by EryTE.

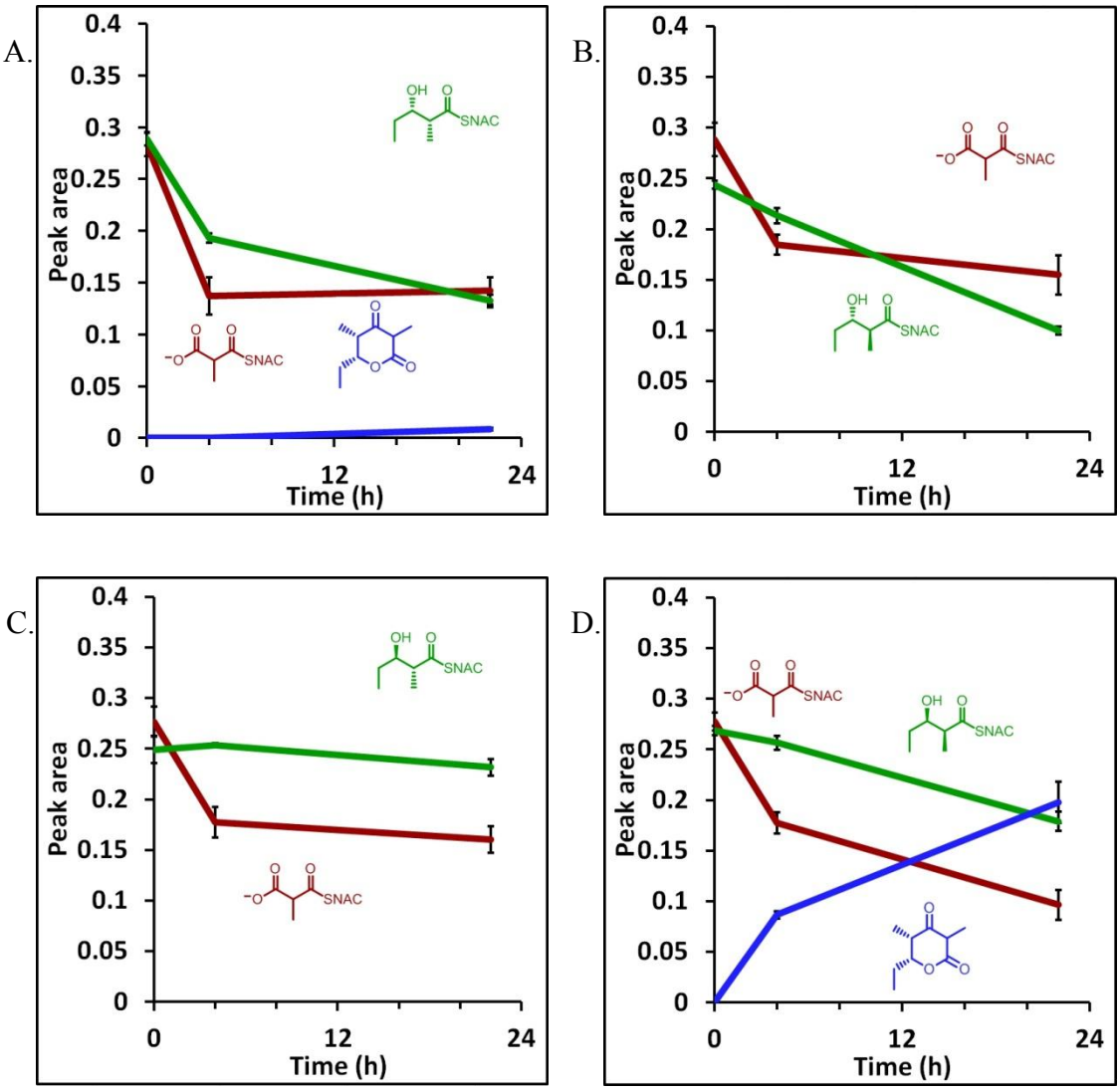

Figure 10. Triplicate HPLC time courses testing the stereocontrol of EryMod6TE.

Plots of HPLC peak area (arbitrary units) over reaction time (hours) showing the reactants and desired product of ketolactone formation by EryMod6TE from **5** and its three stereoisomers. A) (2*S*,3*R*)-3-hydroxy-2-methylpentanoyl-*S*-NAC, generated by incubating 2 with AmpKR2 (A1-type), yielded trace quantities of ketolactone. This is consistent with the formation of the trace quantities of **5** as a side product of the reduction of **2** by AmpKR2 (Piasecki et al. 2011). B) (2*R*,3*R*)-3-hydroxy-2-methylpentanoyl-*S*-NAC, generated by incubating **2** with RifKR7 (A2-type), did not yield ketolactone. C) (2*R*,3*R*)-3-hydroxy-2-methylpentanoyl-*S*-NAC, generated by incubating **2** with TylKR1 (B1-type), did not yield ketolactone. D) (2*S*,3*R*)-3-hydroxy-2-methylpentanoyl-*S*-NAC (**5**), generated by incubating **2** with EryKR1 (B2-type), yielded ketolactone **8**.

ADDITIONAL TERMINAL MODTES

Module TE Selection, partial cloning, and expression of two constructs

In an effort to access more triketide lactone stereochemistries and determine the nature of EryMod6TE's stereocontrol, more ModTEs were selected for cloning and expression. To avoid possible disruptions in module activity resulting from non-natural TE fusions, terminal modules with genes that included TEs were chosen. These included the terminal modules from the amphotericin synthase (AmpMod18TE), the oleandomycin synthase (OleMod6TE), the pikromycin synthase (PikMod6TE), the spinosad synthase (SpnMod10TE), and the tylosin synthase (TylMod7TE). Primers (Table 4) were designed with the assistance of Dr. Jianting Zheng, who also provided genomic DNA.

All modules were amplified from their genomic DNA. PikMod6TE and SpnMod10TE were successfully ligated into an expression vector. After it was demonstrated by SDS-PAGE that these modules were expressed, preliminary chemical characterization took precedent over integrating the remaining modules in vectors.

Detection of TE-mediated glycerolysis activity by fluorescent HPLC

The Keatinge-Clay group has developed a fluorescent-HPLC method to measure *in vitro* PKS activity more sensitively and detect more chemical species. This is done by acylating the KS with a terminal-alkynyl-*S*-NAC, allowing the reaction to proceed, attaching an azide derivative of sulforhodamine using the azide-alkyne Huisgen cycloaddition, and then analysis using HPLC and a fluorometer (Hughes and Keatinge-Clay, 2013 manuscript in preparation).

During the course of this work it was demonstrated that EryTE uses a primary alcohol from a buffer glycerol molecule in place of water to result in glycerolysis, rather than hydrolysis, of terminal-alkynyl-*S*-NACs. Both Pik6TE and Spn10TE generated the

peak that Hughes and Keatinge-Clay assigned to glycerolysis product at approximately the same order of magnitude as EryMod6TE. This activity was observed with 4-pentynoyl-*S*-NAC and 6-heptynoyl-*S*-NAC. This suggests that both constructs are well folded and that the TEs of both are catalytically active. This glycerolysis product has not been confirmed by MS or NMR. PikMod6TE and SpnMod10TE have not been further explored using this method.

Stereochemical substrate selectivity of PikMod6TE and SpnMod10TE

The same technique used to map EryMod6TE's stereochemical requirement for diketide extension was applied to PikMod6TE and SpnMod10TE in hopes of accessing new ketolactones. SpnMod10TE was supplied with malonyl-*S*-NAC as it uses a malonyl-group (rather than a methylmalonyl group) in the context of its synthase. Both modules were supplied with **5** and its stereoisomers, as well as pantetheine-thioesters of the same diketides. Ketolactone formation by SpnMod10TE was not observed.

PikMod6TE was observed to extend A1-type diketides (as pantetheine and NAC thioesters) as well as B2-type diketides. Compared to the EryMod6TE control, PikMod6TE is more active against A1-type substrates, producing the A1- and B2-type ketolactones at ~40 % and ~20 % of the level at which EryMod6TE generates B2-type ketolactone. Because lysate was used, catalyst concentration was not controlled. This means that these measurements reflect catalyst concentration as well as catalytic rate. The production of A1- and B2-type ketolactone by PikMod6TE was confirmed by low-resolution MS.

FUTURE DIRECTIONS

The immediate next step should be NMR confirmation of A1-type triketide ketolactone and measurement of its specific rotation to compare with that of its enantiomer (**11**).

Now that the lab's fluorescent HPLC assay is out of preliminary stages, its use to further interrogate SpnMod10TE and PikMod6TE is appealing, especially since only TE activity has been observed from SpnMod10TE. It would also immediately allow testing of SpnMod10TE using a larger substrate, such as 11-undecynoyl-*S*-NAC.

The stereochemical substrate selectivity of EryMod6TE is tantalizing. PikMod6TE's ability to extend diketides of more than one stereoconfiguration suggests that combination of PikMod6TE with EryMod6TE into chimeric modules could test hypotheses about which domain is responsible for EryMod6TE's substrate selectivity: the KS or the TE. Because recombination of two active constructs does not necessarily produce active constructs, multiple chimeras should be designed and assembled in parallel. Any construct designs should be mindful of ACP/KS interactions demonstrated in the DEBS system (Kapur et al. 2010). Two constructs should be designed for each crossover point, one with EryMod6TE for the N-terminal portion and one with PikMod6TE for the N-terminal portion.

Although activity of a bimodular system has been measured in-lab using CoA-based substrates and radio-TLC (Andrew Harper, unpublished results) integration of truncated extender units has not yet been achieved with a multimodular system. Activity of a multimodular system or even an entire synthase in a cell-free context would be a logical future goal of this line of work.

Methods

GENERAL CONSIDERATIONS

N-acetylcysteamine (NAC) (Hughes and Keatinge-Clay 2011) and all β-ketoacyl-*S*-NAC substrates (**2-4**) (Piasecki et al. 2011) were synthesized according to literature procedures. Methylmalonic acid was purchased from TCI America, ATP was purchased from Meiya Pharmaceuticals, and $NADP^+$ was purchased from CalBioChem. IPTG was purchased from either CarboSynth or Anatrace, and Ni-NTA agarose was purchased from Amintra. For purified proteins, final concentrations were determined using a Thermo Scientific Nanodrop 1000. Acrylamide for SDS-PAGE was purchased as a 30% acrylamide and bis-acrylamide solution (19:1) from BioRad and gels were poured according to the manufacturer's instruction. Sonication was done using a Misonix S-4000 ultrasonic processor set to 10 % amplitude with 0.5 inch horn and tip. All biocatalytic reactions were performed at ambient temperature (~23 °C). Thin layer chromatography (TLC) was conducted with EMD gel 60 F_{254} pre-coated plates (0.25 mM). TLCs were visualized with vanillin stain made by adding 1.5 mL 95-98% H_2SO_4 dropwise to 6 g vanillin dissolved in 95 mL ethanol. Fisher scientific silica gel 60 (particle size 230-400 μm) was used for flash column chromatography. All HPLC monitoring was performed on a Waters 1525 binary HPLC pump connected to a Waters 2998 photodiode array detector using a Varian Microsorb-MV C_{18} column (250 x 4.6 mm, 5 μm particle size, 100 Å pore size) with a matching Metaguard column and mobile phases consisting of water with 0.1% TFA (solvent A) and methanol with 0.1% TFA (solvent B) at a flow rate of 1 mL/min. ^1H NMR data were acquired on a Varian Mercury 400 MHz instrument at ambient temperature and are reported in terms of chemical shift (δ ppm), multiplicity, coupling constant, and integration and are referenced downfield from $(CH_3)_4Si$ to the residual solvent peak at 7.26 ppm for $CDCl_3$ as an internal standard. ^{13}C NMR data were

acquired on either a Varian Mercury 400 MHz instrument or a Varian Oxford 600 MHz instrument at ambient temperature and are reported in terms of chemical shift and referenced to the residual solvent peak at 77.16 ppm for $CDCl_3$ as an internal standard. LC-MS analysis was performed on an Agilent Technologies 1200 Series HPLC with a Gemini C_{18} column (5 μm, 2 x 50 mm, Phenomenex) coupled to an Agilent Technologies 6130 quadrupole mass spectrometer system equipped with an electrospray-ionization source. A 5-95% B gradient over 12 minutes at a flow rate of 0.7 mL/min was run in which the mobile phases were water with 0.1% formic acid (solvent A) and acetonitrile with 0.1% formic acid (solvent B).

CLONING

Step number	Temperature (°C)	Duration (s)	
1.	98	60	
2.	98	5	
3.	Primer specific	10	
4.	72	Template specific	Go to step 5 (39 x)
5.	72	300	
6.	4	Hold	

Table 2. Standard PCR thermal protocol.

Step number	Temperature (°C)	Duration (s)	
1.	98	60	
2.	98	5	
3.	70	10	Decrease 2 °C each round
4.	72	Template Specific	Go to step 2 (9 x)
5.	98	5	
6.	Primer specific	10	
7.	72	Template specific	Go to step 5 (29 x)
8.	72	300	
9.	4	Hold	

Table 3. Touchdown PCR thermal protocol.

Unless otherwise noted, PCR reactions were 10ul and consisted of 1x Phusion GC buffer, 0.1 units of Phusion polymerase, 100 μM ea. dNTPs, 1ng/ul template DNA, and 1 μM ea. forward and reverse primers. Thermocycling was done according to a standard (Table 2) or touchdown protocol (Table 3) with tailored annealing temperatures an extension times. Primers are detailed in Table 4 and were ordered from Sigma. All restriction enzymes and T4 ligase were purchased from New England Biolabs and used according to the manufacturer's suggested protocols. Agarose gel purification was done using Qiagen's QiaQuick kit. Plasmid DNA was purified using Qiagen's QiaPrep kit. Sequencing was done at the University of Texas Institute for Cellular and Molecular Biology Core Facilities.

Template		Sequence (5’ → 3’)
Amp1	Forward	5’ CAAAAAGCATACAATCAACTATCAACTATTAAC TATATCGTAATACCATATGCCGGACGAAAGTAAGC TCGTCGA 3’
	Reverse	5’ CGGCCGGCAGGTCCCGCAT 3’
Amp2	Forward	5’ ACTGTCCCTGGAGGACGCCTGCA 3’
	Reverse	5’ GTCGAGGTCGACGAGCAGCAGGCT 3’
Amp3	Forward	5’ CCTCGACGACGAGGACGTCCTGGA 3’
	Reverse	5’ CGCACAAATTTGTCATTTAAATTAGTGATGGTGA TGGTGATGCACGTGAAGCTTGGCCTCGCCGGCCGG AAGTTC 3’
OleMod6TE	Forward	5’ GTGACGTGTCATATGCTCGCCGCCTCCCGTGAAG CGATC 3’
	Reverse	5’ GAGACTTAGCTCGAGGGGGGTGAGCCCCTTCAG CCAGACGT 3’
PikMod6TE	Forward	5’ AAAGGGAGCCATATGACGAGTTCCAACGAACAG TTGGTGGAC 3’
	Reverse	5’ GTCCGGATTCTCGAGCTTGCCCGCCCCCTCGATG CCCTCGA 3’
SpnMod10TE	Forward	5’ GAGTAGATACATATGCCGGTCGCCGAAGACGAT CTCGTCGCGA 3’
	Reverse	5’ TAAAGCCAACTCGAGGCGTTGCCGAGCGGTGCG CTGGTTGAGTTTGTCA 3’
TylMod7TE	Forward	5’ AATCGCATGCATATGATGGCCATGTCCGCCGA 3’
	Reverse	5’ GCTGTTAGAAAGCTTGCGGCCTCTCCTCTCTCCC C 3’

Table 4. Primer sequences.

AmpMod18TE

A condition to amplify AmpMod18TE as a single fragment was not found. Instead it was amplified in three smaller fragments: Amp1, Amp2, and Amp3. Each fragment has ~50-100bp of sequence overlap with each other and the vector XW55 (Darren Gay, unpublished data). This will allow for homologous recombination in *Saccharomyces cerevisiae* of all three gel-purified fragments and XW55 that has been digested with NdeI and PmlI. All three fragments were amplified from *Streptomyces*

nodosus genomic DNA using the standard thermal profile (Table 2) with a 74.5 °C annealing temperature, a 40s extension time, and 5% v/v DMSO. Amplification of Amp2 was somewhat weak and may benefit from increased DMSO concentration. Once assembled in *S. cerevisiae*, AmpMod18TE may be moved into pET28b by the use of NdeI and HindIII cut sites.

OleMod6TE

OleMod6TE was amplified from *Streptomyces antibioticus* genomic DNA using the standard thermal profile (Table 2) with a 74.5 °C annealing temperature, a 120 s extension time, and 6% DMSO. Primers included NdeI and XhoI cut sites for insertion into pET28b.

PikMod6TE

PikMod6TE was amplified from *Streptomyces venezuelae* genomic DNA using the touchdown thermal profile (Table 3) with a 64 °C annealing temperature, a 120 s extension time, and 5% DMSO. Gel purified amplicon and pET28b were digested with NdeI and XhoI and the vector was subsequently treated with alkaline phosphatase. Ligation and successful transformation were confirmed with Sanger sequencing.

SpnMod10TE

SpnMod10TE was amplified from *Saccharopolyspora spinosa* genomic DNA using the standard thermal profile (Table 2) with a 71.4 °C annealing temperature, a 120 s extension time, and 4% DMSO. Gel purified amplicon and pET28b were digested with NdeI and XhoI and the vector was subsequently treated with alkaline phosphatase. Ligation and successful transformation were confirmed with Sanger sequencing.

TylMod7TE

TylMod7TE was amplified from ***Streptomyces fradiae*** genomic DNA using the standard thermal profile (Table 2) with a 74.5°C annealing temperature, a 120 s extension time, and 6% DMSO. Primers included NdeI and HindIII cut sites for insertion into pET28b.

PROTEIN EXPRESSION AND PURIFICATION

Streptomyces coelicolor MatB, (Hughes and Keatinge-Clay 2011) TylKR1, EryKR1, AmpKR, MycKR5, GDH (Piasecki et al. 2011), and RifKR7 (Valenzeno et al. 2010) were expressed in *E. coli* BL21 (DE3). All PKS constructs were expressed in *E. coli* K207-3 (Murli et al. 2003). Starter cultures (50 mL) were grown to inoculate pre-warmed Luria broth supplemented with 25 mg/mL kanamycin. When OD_{600} reached 0.4, the media was cooled to 15 °C and then induced with 0.5 mM IPTG. After 16 h, the cells were harvested by centrifugation (3,000 x *g* for 20 minutes), and the pellets of 6 L of cell growth were resuspended in 50 mL lysis buffer (30 mM HEPES pH 7.5, 250 mM NaCl, 10 % v/v glycerol). In lysis buffer for SpnMod10TE, NaCl concentration was raised to 500 mM and glycerol concentration was raised to 20% v/v. The cells were then lysed by sonication on ice with two rounds of 15 separate 2 s pulses with 10 s between pulses and 60 s between rounds. Cellular debris was then removed by centrifugation (30,000 x *g* for 45 minutes). For proteins that were purified (MatB and GDH), the crude lysate was passed over a nickel-NTA column equilibrated with lysis buffer. The column was washed with lysis buffer containing 15 mM imidazole, and the protein was eluted with lysis buffer containing 150 mM imidazole. For isolated ketoreductases (EryKR1, TylKR1, AmpKR2, and RifKR7) and PKSs, lysate was used. To generate lysate, cells from 6 L of cell growth were pelleted after expression (3,000 x *g* for 20 minutes), and the pellets of were resuspended in 50 mL lysis buffer before sonication (using the protocol above) and

centrifugation (30,000 x *g* for 45 minutes). The crude lysate was then twice dialyzed for 8 h at 4 °C in 10 kDa MWCO cellophane dialysis tubing against 1 L of lysis buffer.

BIOCATALYTIC SYNTHESES

Extender units (MatB reaction)

Reactions generating methylmalonyl-*S*-NAC (**1**) contained 500 mM HEPES (pH 8.0), 100 mM NaCl, 10 % v/v glycerol, 100 mM sodium methylmalonate, 100 mM ATP, 200 mM $MgCl_2$, 100 mM NAC, and 0.5 mg/mL MatB and were allowed to proceed at for 48 h before immediate use or storage at 4 °C. Reaction progress was monitored via HPLC at 235 nm (linear gradient of 0-50% B over 15 minutes) (Hughes and Keatinge-Clay 2011). To generate malonyl-*S*-NAC, sodium malonate was substituted for sodium methylmalonate.

β-hydroxyacyl-*S*-NACs (KR reactions)

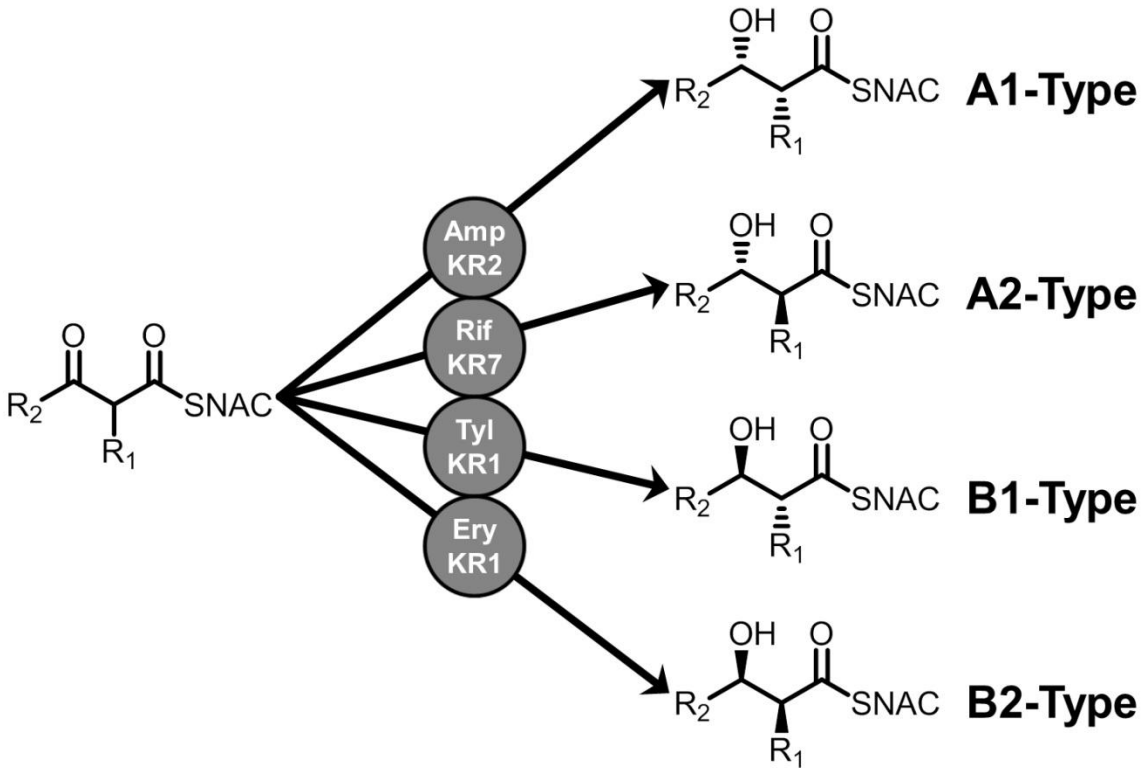

Figure 11. KR types used to generate β-hydroxy-diketide-*S*-NACs.

Generation of reduced diketides was performed in 500 mM HEPES (pH 7.5), 100 mM NaCl, 15% v/v glycerol, 5% v/v DMSO, 300 mM D-glucose, 1 mM $NADP^+$, 100 mM β-ketoacyl-*S*-NAC **2-4**, 0.05 mg/mL GDH, and 20% v/v KR-containing lysate and went to completion after 16 h. Use of TylKR1 generated **6** and **7** from **3** and **4**, respectively. To generate **5** and its three stereoisomers, a KR of each type (Keatinge-Clay 2007) was used to generate its cognate reduced diketide-*S*-NAC product from **2** (Piasecki et al. 2011): The A1-type AmpKR2 for the (2*R*,3*S*)-diketide, A2-type RifKR7 (Valenzeno et al. 2010) for the (2*S*,3*S*)-diketide, the B1-type TylKR1 for the (2*R*,3*R*)-diketide, and the B2-type EryKR1 for the (2*S*,3*R*)-diketide (**5**) (Figure 11). Reaction

progress was monitored by TLC (15% MeOH:$CHCl_3$) and HPLC absorbance at 235 nm (linear gradient of 15-35% B over 20 minutes).

Triketide lactones from EryMod6TE

The progress of reactions yielding triketide lactones **8-13** was monitored via TLC (50% EtOAc:hexanes) and vanillin staining. For 0.5 mmol reactions yielding β-hydroxy triketide lactones **8**, **9**, or **10**, 5 mL of MatB reaction yielding methylmalonyl-*S*-NAC **1** and 5 mL of KR reaction yielding **5**, **6**, or **7** (respectively) were combined with 20 mL of EryMod6TE lysate. After 24 h, the reactions were extracted with EtOAc (3 x 150 mL), dried with $MgSO_4$, filtered, and concentrated under vacuum. This crude extract was then extracted with 100 mL of chloroform to remove remaining glycerol, concentrated under vacuum, and purified by flash chromatography (70% EtOAc:hexanes).

To produce ketolactones **11**, **12**, or **13**, the same procedure was used with the following amendment: 5 mL of the same KR reaction was first extracted with EtOAc (3 x 50 mL), dried with $MgSO_4$, filtered, and concentrated under vacuum, and resuspended in 5 mL of 500mM HEPES pH 7.5. It was then combined with 5 mL of MatB reaction and 20 mL of EryMod6TE lysate as before.

The preparative (7 mmol) reaction of **10** was done by combining 70 mL of MatB reaction with 70 mL of KR reaction containing **7** and 190 mL of EryMod6TE lysate. After two days the reaction was extracted with EtOAc (6 x 250 mL). The remaining emulsion was centrifuged (1000 x g for 1 minute) and the organic phase was removed. The remaining aqueous phase was then extracted with EtOAc (2 x 250 mL), and the remaining emulsion was centrifuged again (1000 x g for 1 minute) and the organic phase was again removed. The combined organic phases were dried with $MgSO_4$, filtered, and concentrated under vacuum. The crude extract was then extracted with chloroform (2 x

200 mL) to remove remaining glycerol, concentrated under vacuum, and purified via flash column chromatography (silica, 50% EtOAc:hexanes).

NMR characterization of triketide lactones

(2*R*,3*S*,4*S*,5*R*)-3,5-Dihydroxy-2,4-dimethylheptanoic acid δ-lactone (**8**). ^{1}H NMR (400 MHz, CDCl3) δ= 4.13 (m, 1H), 3.82 (dd, J=4.5 Hz, J=10 Hz, 1H), 2.47 (dq, J=4.5 Hz, 7 Hz, 1H), 2.17 (m, 1H), 1.83 (m, 1H), 1.55 (m, 1H), 1.41 (d, J=7 Hz, 3H), 1.00 (t, J=7 Hz, 3H), 0.97 (d, J=7 Hz,3H). ^{13}C NMR (600 MHz, CDCl3) δ=173.51, 81.31, 73.96, 39.84, 36.75, 25.26, 14.26, 9.86, 4.35. This characterization is in agreement with literature reported data (Kao et al. 1994).

(2*R*, 3*S*, 5*R*)-3,5-Dihydroxy-2-methylheptanoic acid δ-lactone (**9**). ^{1}H NMR (400 MHz, CDCl3) δ= 4.16 (m, 1H), 3.76 (td, 10.5, 4.5 1H), 2.35 (m, 1H), 2.20 (m, 1H), 1.82-1.62 (m, 3H), 1.41 (d, J=7 Hz, 3H), 1.00 (t, J=7 Hz, 3H). ^{13}C NMR (400 MHz, CDCl3) 173.53, 77.89, 70.48, 45.24, 37.84, 28.94, 13.63, 9.29.

(2*R*, 3*S*, 5*R*)-3,5-Dihydroxy-2-methyloctanoic acid δ-lactone (**10**). ^{1}H NMR (400 MHz, CDCl3) δ= 4.24 (m, 1H), 3.76 (m, 1H), 2.35 (m, 1H), 2.20 (m, 1H), 1.78-1.46 (m, 5H), 1.40 (d, J=7 Hz, 3H), 0.92 (t, J=7 Hz, 3H). ^{13}C NMR (400 MHz, CDCl3). δ= 174.06, 76.70, 70.15, 45.15, 38.26,38.01, 18.12, 13.87, 13.62.

(4*S*, 5*R*)-2,4-Dimethyl-3-oxoheptanoic acid δ-lactone (**11**). ^{1}H NMR (400 MHz, CDCl3) δ= 4.65 (m, 1H), 3.61 (q, J=7 Hz, 1H), 2.62 (dq, J=7 Hz, J=3.2 Hz, 1H), 1.85 (m, 1H), 1.65 (m, 1H), 1.36 (d, J=7 Hz, 3H), 1.12 (d, J=7 Hz, 3H), 1.07 (t, J=7 Hz, 3H). ^{13}C NMR (400 MHz, CDCl3) δ= 205.59, 170.23, 78.68, 50.55, 44.52, 24.18, 10.07, 9.88, 8.39. NMR shifts suggest that **11** is a keto tautomer in solution, although it crystalized as the enol tautomer (Figure 5).

(5*R*)-2-Methyl-3-oxoheptanoic acid δ-lactone (**12**). ^{1}H NMR (400 MHz, CDCl3) δ= 4.64 (m, 1H), 3.57 (q, J=6.5 Hz, 1H), 2.74 (dd, J=2.8 Hz, J=19 Hz, 1H), 2.44 (dd, J=12 Hz, J=19 Hz, 1H), 1.90-1.75 (m, 2H), 1.37 (d, J=6.5 Hz, 3H), 1.08 (t, J=7 Hz, 3H). ^{13}C NMR (400 mHz, CDCl3) δ= 201.71, 170.07, 75.46, 51.80, 42.83, 27.65, 9.33, 7.93.

(5*R*)-2-Methyl-3-oxooctanoic acid δ-lactone (**13**). ^{1}H NMR (400 MHz, CDCl3) δ= 4.70 (m, 1H), 3.56 (q, J=6.5 Hz, 1H), 2.73 (dd, J=2.7 Hz, J=19 Hz, 1H), 2.47 (dd, J=12 Hz, J=18 Hz, 1H), 1.6-1.4 (m, 3H), 1.36 (d, J=6.5 Hz, 3H), 0.99 (t, J=7 Hz, 3H). ^{13}C NMR (400 MHz, CDCl3) δ=201.76, 170.13, 74.04, 51.82, 43.22, 36.47, 18.16, 13.79, 7.83.

MONITORING OF KETOLACTONE FORMATION BY HPLC

EryMod6TE time courses with B-type and B2-type products

MatB reaction and extracted, reduced diketide were prepared and HPLC-analyzed as indicated above. Ketolactone reactions consisted of 250 µL MatB reaction, 1 mL PKS lysate, an amount of extracted β-hydroxyacyl-*S*-NAC equivalent to that of methylmalonyl-*S*-NAC (as estimated by HPLC peak area), and 500 mM HEPES pH 7.5 added to produce a final volume of 2.1 mL. Timepoints were taken at 0, 1, 2, 4, 8, 12, 22, and 24 h by quenching 200 µL of the reaction with an equal volume of 2 M acetic acid in methanol and storing at -20 °C until analyzed. Before analysis, timepoint samples were centrifuged at 21,000 x g for 1 min (to pellet precipitate) and the supernatant was filtered through a 0.45 µm nylon membrane. For reactions yielding **11-13**, 20-minute linear gradients (5-100% solvent B) were performed. For reactions yielding **12**, an additional isocratic run (30% B) was necessary to separate **6** from **12**. 25 µL of the quenched reaction was injected for each timepoint and absorbance was monitored at 242 nm (Figures 6-9). Ketolactone formation was further confirmed via LC-MS (Table 5).

Compound	Formula	*m/z* (expected)	*m/z* (observed)
11	$C_9H_{14}O_3$	171.09	171.4
12	$C_8H_{12}O_3$	157.08	157.2
13	$C_9H_{14}O_3$	171.09	171.2

Table 5. LC/ESI-MS confirming ketolactone formation.

Stereochemical substrate selectivity of EryMod6TE and PikMod6TE

The reactions to form **5** and its stereoisomers were performed, extracted, and analyzed using the procedure detailed above. Generation of **1** was also performed using the conditions described above. Ketolactone reactions consisted of 250 μL MatB reaction, 1 mL PKS lysate, an amount of extracted β-hydroxyacyl-*S*-NAC equivalent to that of methylmalonyl-*S*-NAC (as estimated by HPLC peak area), and 500 mM HEPES pH 7.5 added to produce a final volume of 2.1 mL. Timepoints taken at 0, 4, and 22 h were quenched, filtered, and assayed by HPLC as described in previous section.

References

Cane DE. 2010. Programming of erythromycin biosynthesis by a modular polyketide synthase. J Biol Chem 285(36):27517-27523.

Castonguay R, He W, Chen AY, Khosla C, Cane DE. 2007. Stereospecificity of ketoreductase domains of the 6-deoxyerythronolide B synthase. J Am Chem Soc 129(44):13758-13769.

Chen AY, Cane DE, Khosla C. 2007. Structure-based dissociation of a type I polyketide synthase module. Chem Biol 14(7):784-792.

Clemons PA, Bodycombe NE, Carrinski HA, Wilson JA, Shamji AF, Wagner BK, Koehler AN, Schreiber SL. 2010. Small molecules of different origins have distinct distributions of structural complexity that correlate with protein-binding profiles. Proc Natl Acad Sci U S A 107(44):18787-18792.

Fischbach MA, Walsh CT, Clardy J. 2008. The evolution of gene collectives: How natural selection drives chemical innovation. Proc Natl Acad Sci U S A 105(12):4601-4608.

Harper AD, Bailey CB, Edwards AD, Detelich JF, Keatinge-Clay AT. 2012. Preparative, *in vitro* biocatalysis of triketide lactone chiral building blocks. Chembiochem 13(15):2200-2203.

Hertweck C. 2009. The biosynthetic logic of polyketide diversity. Angew Chem Int Ed Engl 48(26):4688-4716.

Hilterhaus L, Liese A. 2007. Building blocks. Adv Biochem Eng Biotechnol 105:133-173.

Hughes AJ, Detelich JF, Keatinge-Clay AT. 2012. Employing a polyketide synthase module and thioesterase in the semipreparative biocatalysis of diverse triketide pyrones. Med Chem Commun 3:956-959.

Hughes AJ, Keatinge-Clay A. 2011. Enzymatic extender unit generation for *in vitro* polyketide synthase reactions: structural and functional showcasing of *Streptomyces coelicolor* MatB. Chem Biol 18(2):165-176.

Hughes AJ, Keatinge-Clay AT. 2013. Manuscript in preparation.

Kao CM, Luo G, Katz L, Cane DE, Khosla C. 1994. Engineered biosynthesis of a triketide lactone from an incomplete modular polyketide synthase. J Am Chem Soc 116(25)11612-11613.

Kapur S, Chen AY, Cane DE, Khosla C. 2010. Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase. Proc Natl Acad Sci U S A 107(51):22066-22071.

Kapur S, Lowry B, Yuzawa S, Kenthirapalan S, Chen AY, Cane DE, Khosla C. 2012. Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation. Proc Natl Acad Sci U S A 109(11):4110-4115.

Keatinge-Clay AT. 2007. A tylosin ketoreductase reveals how chirality is determined in polyketides. Chem Biol. 2007 Aug;14(8):898-908.

Khosla C, Tang Y, Chen AY, Schnarr NA, Cane DE. 2007. Structure and mechanism of the 6-deoxyerythronolide B synthase. Annu Rev Biochem 76:195-221.

Kinoshita K, Pfeifer BA, Khosla C, Cane DE. 2003. Precursor-directed polyketide biosynthesis in *Escherichia coli*. Bioorg Med Chem Lett 13(21):3701-3704.

Lee HY, Harvey CJ, Cane DE, Khosla C. 2011. Improved precursor-directed biosynthesis in *E. coli* via directed evolution. J Antibiot (Tokyo) 64(1):59-64.

Lum AM, Huang J, Hutchinson CR, Kao CM. 2004. Reverse engineering of industrial pharmaceutical-producing actinomycete strains using DNA microarrays. Metab Eng 6(3):186-196.

Mickel SJ, Sedelmeier GH, Niederer D, Daeffler R, Osmani A, Schreiner K, Seeger-Weibel M, Bérod B, Schaer K, Gamboni R, Chen S, Chen W, Jagoe CT, Kinder FR Jr., Loo M, Prasad K, Repič O, Shieh W, Wang R, Waykole L, Xu DD, Xue S. 2011. Large-scale synthesis of the anti-cancer marine natural product (+)-discodermolide. Part 1: synthetic strategy and preparation of a common precursor. Org Proc Res Dev 8(1):92-100.

Murli S, Kennedy J, Dayem LC, Carney JR, Kealey JT. 2003. Metabolic engineering of *Escherichia coli* for improved 6-deoxyerythronolide B production. J Ind Microbiol Biotechnol 30(8):500-509.

Oliynyk M, Brown MJ, Cortés J, Staunton J, Leadlay PF. 1996. A hybrid modular polyketide synthase obtained by domain swapping. Chem Biol 3(10):833-839.

Peano C, Talà A, Corti G, Pasanisi D, Durante M, Mita G, Bicciato S, De Bellis G, Alifano P. 2012. Comparative genomics and transcriptional profiles of *Saccharopolyspora erythraea* NRRL 2338 and a classically improved erythromycin over-producing strain. Microb Cell Fact 11:32.

Piasecki SK, Taylor CA, Detelich JF, Liu J, Zheng J, Komsoukaniants A, Siegel DR, Keatinge-Clay AT. 2011. Employing modular polyketide synthase ketoreductases as biocatalysts in the preparative chemoenzymatic syntheses of diketide chiral building blocks. Chem Biol 18(10):1331-1340.

Pohl NL, Gokhale RS, Cane DE, Chaitan Khosla. 1998. Synthesis and incorporation of an N-acetylcysteamine analogue of methylmalonyl-CoA by a modular polyketide synthase. J Am Chem Soc 120(43):11206-11207.

Siskos AP, Baerga-Ortiz A, Bali S, Stein V, Mamdani H, Spiteller D, Popovic B, Spencer JB, Staunton J, Weissman KJ, Leadlay PF. 2005. Molecular basis of Celmer's

rules: stereochemistry of catalysis by isolated ketoreductase domains from modular polyketide synthases. Chem Biol 12(10):1145-1153.

Valenzano CR, You YO, Garg A, Keatinge-Clay A, Khosla C, Cane DE. 2010. Stereospecificity of the dehydratase domain of the erythromycin polyketide synthase. J Am Chem Soc 132(42):14697-14699.

Weissman KJ. 2009. Introduction to polyketide biosynthesis. Methods Enzymol 459:3-16.

Weissman KJ, Leadlay PF. 2005. Combinatorial biosynthesis of reduced polyketides. Nat Rev Microbiol 3(12):925-936.

Wiesmann KE, Cortés J, Brown MJ, Cutter AL, Staunton J, Leadlay PF. 1995. Polyketide synthesis *in vitro* on a modular polyketide synthase. Chem Biol 2(9):583-589.

Wu J, Kinoshita K, Khosla C, Cane DE. 2004. Biochemical analysis of the substrate specificity of the beta-ketoacyl-acyl carrier protein synthase domain of module 2 of the erythromycin polyketide synthase. Biochemistry 43(51):16301-16310.

Zhang H, Wang Y, Wu J, Skalina K, Pfeifer BA. 2010. Complete biosynthesis of erythromycin A and designed analogs using *E. coli* as a heterologous host. Chem Biol 17(11):1232-1240.