UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Towards Biochemical Conversion of CO2 to Higher Value Chemicals Using Enzyme Design and Engineered Polyketide Synthases

Permalink

https://escholarship.org/uc/item/8vd4s91k

Author

Poust, Sean Kirk

Publication Date

2015

Peer reviewed|Thesis/dissertation

Towards Biochemical Conversion of CO$_2$ to Higher Value Chemicals Using Enzyme
Design and Engineered Polyketide Synthases

by

Sean Kirk Poust

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jay D. Keasling, Chair
Professor Wenjun Zhang
Professor David Savage

Spring 2015

Towards Biochemical Conversion of CO$_2$ to Higher Value Chemicals Using Enzyme
Design and Engineered Polyketide Synthases

Copyright 2015

by Sean Kirk Poust

Abstract

Towards Biochemical Conversion of CO$_2$ to Higher Value Chemicals Using Enzyme
Design and Engineered Polyketide Synthases

by

Sean Kirk Poust

Doctor of Philosophy in Chemical Engineering

University of California, Berkeley

Professor Jay D. Keasling, Chair

Polyketide synthases produce a remarkable number of diverse products. Many medicines
have been produced either directly from polyketide products found in nature, or with
limited modification via organic semisynthesis. However, we remain in the early days of
engineering the biosynthesis of polyketide synthases. Despite being an active area of
research for over 20 years, no commercial application of an engineered polyketide
synthase exists. Rapidly advancing technologies like next generation sequencing, DNA
synthesis, liquid handling automation and mass spectrometry give us tools to do things
faster, and in new ways. In the work presented in this doctoral dissertation I will describe
how I have applied some of these new tools to advance our understanding of polyketide
synthase biochemistry, as well develop catalysts for new small molecules using
engineered polyketide synthases.

Chapter 1 begins with a discussion of the current state of the art and challenges
associated with engineering these enzymes as well as offering opinions about routes
forward to narrow the considerable gap between the promise and reality of engineered
PKSs.

Chapter 2 presents my work applying mass spectrometry to study the formolase enzyme,
which is part of an engineered carbon fixation pathway. This work further characterized
the product profile of the formolase and describes alternative routes of carbon
assimilation from the pathway. Once this pathway is further evolved, it could supply
reduced carbon to engineered polyketide synthase proteins.

Chapter 3 explores the role of the histidine in the GHSxS active site motif of acyl
transferase domains. The role of this residue in the literature was unclear, with some
reports suggesting that this histidine could even serve as an alternative nucleophile. This
study showed that removal of this histidine increases hydrolysis, suggesting it is
important for the stabilization of acyl-enzyme intermediates.

Chapter 4 probes the mechanism for *gem*-dimethylation in complex polyketide biosynthesis. Specifically, this report shows that methylation can precede condensation in polyketide biosynthesis, contrary to the canonical understanding.

Chapter 5 describes the development of the yersiniabactin PKS as a *gem*-dimethylation catalyst. Specifically, the engineering of this enzyme towards production of several useful *gem*-dimethylated compounds, as well as the parallel engineering of a thioesterase domain to release *gem*-dimethylated intermediates is described.

The final chapter will summarize this work, as well as suggesting future efforts to further enable the engineering of polyketide assembly lines.

For my wife's support and love

Table of Contents

1. Chapter 1 – Introduction

Engineered modular polyketide synthases (PKSs) have the potential to be an extraordinarily effective retrosynthesis platform. Native PKSs assemble and tailor simple, readily available cellular acyl-CoAs into large, complex, chiral molecules (**Figure 1.1**). By successfully rearranging existing polyketide modules and domains, one could exquisitely control chemical structure from DNA sequence alone. As an example of the diverse biosynthetic potential of PKSs, we have concluded that approximately 20 of the roughly 150 commodity chemicals tracked by the petrochemical market information provider ICIS could be produced by mixing and matching naturally occurring PKS domains (ISIS Chemicals commodity and product finder; URL: http://www.icis.com/chemicals/channel-info-finder/). To form these chemicals, engineered PKSs would load acyl-CoAs accessible in *Escherichia coli* and other industrially relevant hosts, perform a programmed number of extension and then release products using previously published mechanisms. However, this potential has only just begun to be realized as the compounds that have been made using engineered PKSs represent a small fraction of the potentially accessible chemical space. We envision a future in which a single design algorithm, using a molecule of interest as input, successfully combines natural PKS sequences to produce the desired molecule.
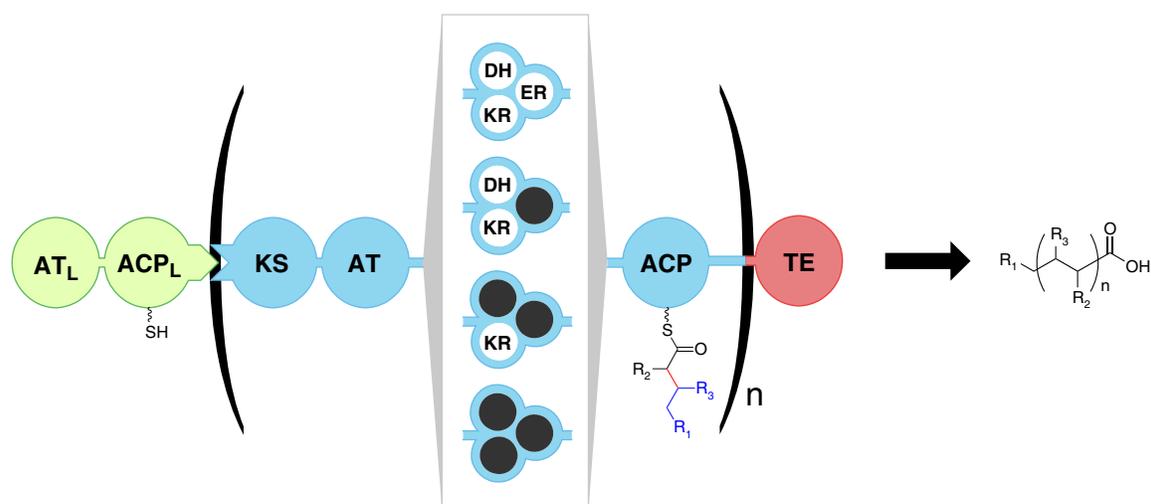
Figure 1.1. Symbolic representation of a general type I modular polyketide synthase and a general polyketide product. Biosynthesis begins with the selection of an acyl-CoA by the loading acyl transferase (AT_L) and subsequent transfer to the phosphopantetheine arm of the loading acyl carrier protein (ACP_L). A variable number of extension modules (represented as 'n') perform successive elongations of the enzyme-bound intermediates with downstream malonyl or methylmalonyl-ACPs (loaded by their cognate AT domains) via Claisen condensations catalyzed by ketosynthase (KS) domains. Accessory domains often present within a module such as ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) domains determine the final oxidation state and stereochemistry of the β-carbonyl attached to the ACP. The bond formed by the KS is represented on the ACP in red; the portion of the extended polyketide chain formed from the malonyl moiety is shown in black; and the portion from the upstream module(s) is shown in blue. In both the general polyketide product and in the intermediate attached to the ACP, R_1 represents the acyl chain from upstream module(s) or from the loading module and R_2 represents the side-chain on the extender unit, R_3 may be the oxygen of a ketone, an (*R*) or (*S*) hydroxyl or hydrogen, depending on the number of reducing domains in the particular module employed. Linker-mediated interactions promote chain transfer when modules are contained within different polypeptides. At the end of the synthesis, the final product is released as a free acid or as a lactone by a thioesterase. The bond entirely within the parentheses of the general product may be a double bond if a KR–DH pair is present within a module.

Assaying for activity of an engineered PKS can be difficult. These proteins can be difficult to express solubly in a heterologous host and often have slow kinetics (~0.01-1 min^{-1}) and low total turnover number (~1-100 molecules produced per molecule of enzyme). This often results in low concentrations of product, increasing the difficulty of detecting activity via an analytical method for the final product. PKS activity in this dissertation was often detected by a liquid chromatography / tandem mass spectrometry (LC-MS/MS) technique that detects a daughter ion from the phosphopantetheinylated active site on the acyl carrier protein of the PKS (**Figure 1.2**). The parent/daughter ion pair masses are dependent on the acyl intermediate covalently linked to the PKS. After an *in vitro* protein assay, the PKS protein is digested with trypsin to form smaller peptides. These peptides are separated by LC and injected into a tandem mass spectrometer. The first quadrupole targets the active site peptide with an acyl group attached to the phosphopantetheine arm. Collision-induced dissociation causes an elimination reaction to occur on the arm, forming a daughter ion, which is targeted by the next quadrupole (Dorrestein et al., 2006). This technique for detection of acyl intermediates is very sensitive due to the high ionization efficiency of peptides. This is useful for initial

experiments on suboptimal engineered PKSs, where absolute concentrations of products are low. Additionally, tandem mass spectrometry has low background and detection of many reaction products is feasible using this single assay.
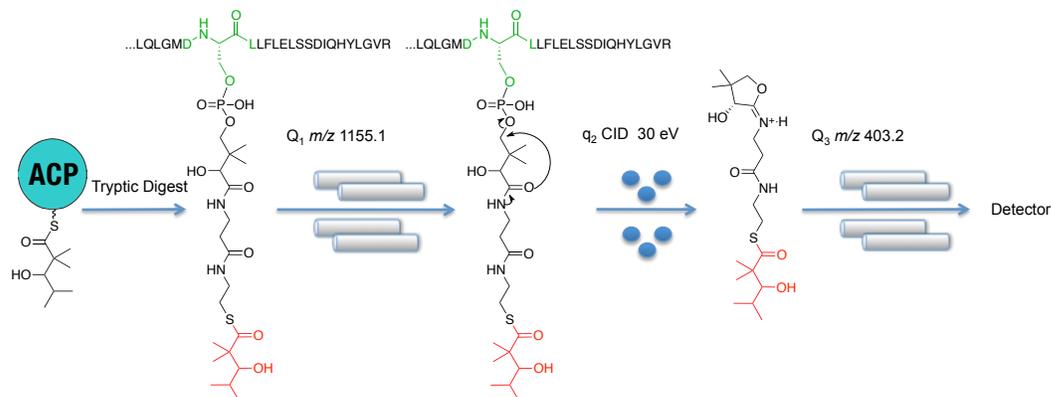
Figure 1.2. Schematic diagram of LC-MS/MS assay to observe acyl intermediates attached to PKSs. First the ACP is digested into peptides using trypsin, as shown on the far left. These peptides are separated using liquid chromatography. Tandem mass spectrometry targets both the parent phosphopantetheine-containing peptide and a daughter ion (with acyl intermediate attached, shown in red) produced by collision-induced dissociation in quadrupole (q2). Quadrupole 1 (Q1) targets the parent active site peptide with the acyl intermediate attached and quadrupole 3 (Q3) targets the daughter ion with the acyl intermediate attached. The DSL motif of the ACP is highlighted in green.

1.1. Narrowing the gap between the promise and reality of polyketide synthases as a synthetic biology platform

1.1.1. The state of the art

Many compounds have been produced using engineered PKSs. Examples include triketide lactones (McDaniel et al., 1997; Menzella et al., 2005; Pieper et al., 1995), pyrones (Hughes et al., 2012), linear branched carboxylic acids (Guo et al., 2010; Yuzawa et al., 2013), and advanced intermediates in the synthesis of epothilone (Menzella et al., 2010) (**Figure 1.3**). However, production titers from engineered PKSs are usually low in heterologous hosts such as *Streptomyces coelicolor* or *E. coli*, likely because the activities of the engineered enzymes are greatly reduced relative to their native counterparts, or the engineered enzymes are not solubly expressed in the host. For example, in the most extensive attempt to produce a wide variety of triketide lactones, Menzella and co-workers rearranged PKS domains from different clusters in a total of 154 bimodular combinations (Menzella et al., 2005). Remarkably, half produced detectable amounts of product, demonstrating the potential of PKSs as a platform to produce many different molecules. However, the great majority of functional constructs produced very little product and only a few produced compounds at titers above 10 mg/L. As a comparison, a 'wild type' strain of *Saccharopolyspora erythraea* can produce erythromycin at 660 mg/L (Hamedi et al., 2004). Menzella and coworkers subsequently used the knowledge of functional proteins gained from their earlier work to engineer trimodular PKS combinations (Menzella et al., 2007), where they were able to achieve a much higher proportion of functional proteins but, once again, the same problem

3

remained with most combinations producing only a small amount of product.
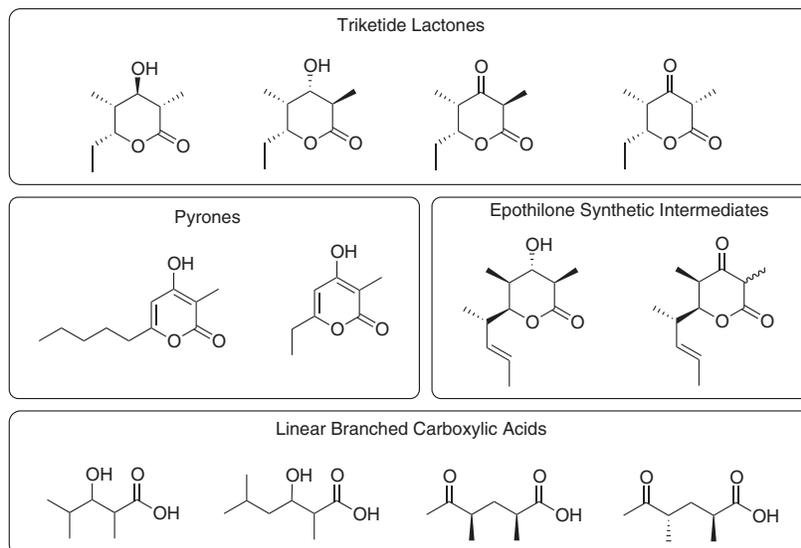
Figure 1.3. Example compounds produced by engineered polyketide synthases. These include triketide lactones, pyrones, linear branched carboxylic acids, and advanced intermediates in the synthesis of epothilone.

While this work demonstrates modularity and the chemistry available for engineered PKSs, the knowledge of how to universally combine modules without compromising titers or kinetics remains to be elucidated. Expanding upon existing PKS engineering efforts to produce analogs of natural compounds (recently reviewed (Winter and Tang, 2012)), we believe that designing PKSs to make compounds of interest from the bottom up, quickly, and at titers of at least 10 mg/L will drive the field forward. This titer is high enough to allow for NMR-based structural characterization and for activity assays. Little work in the area of reverse engineering of PKSs to produce compounds of interest with no natural analogs has been published, but the tools to produce such molecules are clearly beginning to become available.

1.1.2. How to develop polyketide synthase based retrosynthesis?

We propose two main thrusts toward improving PKS-based retrosynthesis. First is the development of a detailed scientific understanding of PKS domain specificity, reaction mechanisms, structures, and domain–domain protein interactions. Knowledge in this area has advanced considerably in the last 20 years. For more information on this area, we direct the reader to recent reviews (Keatinge-Clay, 2012; Khosla, 2009). As part of these efforts, engineering specificity of individual domains, especially acyl transferases, is an ongoing pursuit, recently reviewed by Dunn and Khosla (Dunn and Khosla, 2013).

A second thrust toward PKS-based retrosynthesis is improvement of the design–build–test–learn cycle for PKSs (**Figure 1.4**), which we believe has great, as yet under-developed potential. Here we address areas within each of the parts of this cycle:

i. **Design**: reliable methods for engineering chimeric PKS proteins.
ii. **Build**: fast DNA assembly and soluble protein production in PKS expression hosts.
iii. **Test**: analytical methods for product detection, finding kinetic bottlenecks and high throughput screening.
iv. **Learn**: incorporating lessons learned though iterations of this cycle toward development of a single retro-synthesis algorithm.
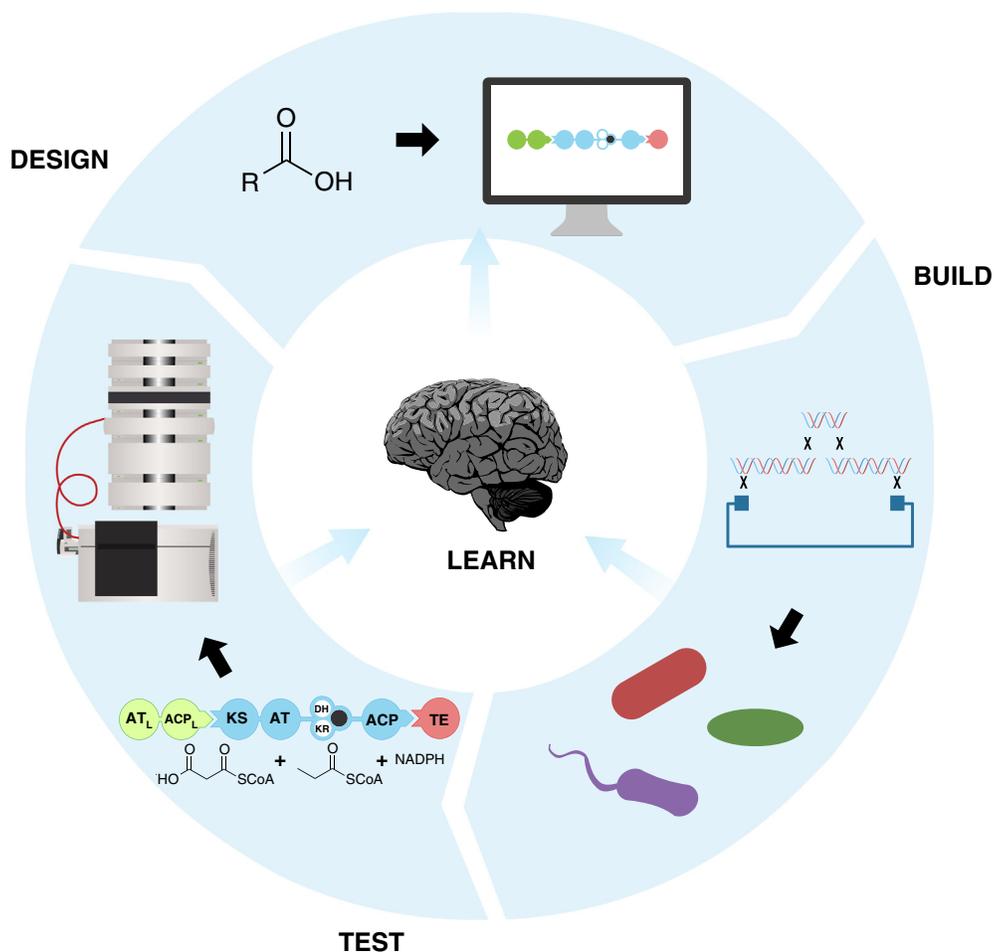
Figure 1.4. The iterative design-build-test cycle for engineering polyketide synthases. We define each portion of the cycle in the following ways. Design: sequence level design of chimeric PKS proteins to produce a desired molecule of interest. Build: DNA assembly of desired chimeric PKS sequences and soluble protein production in PKS expression hosts. Test: product detection via analytical methods and high throughput screening. Learn: incorporating lessons learned though iterations of this cycle toward development of a retrosynthesis algorithm.

1.1.3. Design

Developing PKS-based retrosynthesis will invariably involve the construction of chimeric proteins, and many different strategies have been successfully employed to produce them. Initial efforts relied on sequence conservation to determine the boundaries of domains, which were used to covalently fuse several different modules, as well as to swap KR domains, ACP domains and DH-KR didomains (McDaniel et al., 1997). A

conserved YRVXW sequence between KS and AT domains was identified using limited proteolysis and has been used to construct KS–AT chimeric proteins (Chen et al., 2006; Kim et al., 2004). Structure-based methods have also been used to determine functional domain boundaries. For example, an EEAPERE motif between the KS and AT domain was used by Chen and coworkers to dissociate these domains into functional stand-alone proteins (Chen et al., 2007).

Another strategy to facilitate nascent polyketide chain transfer between modules *in trans* is the incorporation of cognate interpolypeptide docking domains, often from heterologous clusters. Successful applications of this strategy include using class 1 docking domains from the erythromycin cluster to facilitate acyl-chain transfer between the ACP of module 1 and the KS domain of module 3 of the erythromycin cluster (Weissman, 2006), and incorporating linker domains from the phoslactomycin cluster between the ACP and KS of PikAI, the first polypeptide in the pikromycin cluster, creating two peptides that would be brought together post-translationally by the linkers to facilitate transfer between the dissected PikAI protein pairs (Yan et al., 2009). Most recently, cyanobacterial docking domains, which bind each other in a distinct manner and have a different structure from class 1 linker domains have also been used to facilitate transfer between the last two modules of the pikromycin PKS cluster *in trans* (Whicher et al., 2013). These constructions have facilitated transfer *in trans* with modules that either normally interact *in cis* (Weissman, 2006; Yan et al., 2009) or have their own native linker domains (Whicher et al., 2013); in the more general case of using docking domains to facilitate transfer between modules from different clusters, achieving titers above 10 mg/L remains difficult, as discussed above for the work of Menzella and coworkers (Menzella et al., 2005).

The techniques above employ coarse-grained strategies: they rely on rationally combining a few DNA sequences and assaying for any resulting activity, which may be low (additional examples of PKS engineering have been reviewed by Chen and Khosla(Zhao et al., 2008)). These existing strategies also have no clear path toward improving an inactive or marginally active construct. A more refined strategy that could improve the design of chimeric PKSs is profiling the active and inactive members of large libraries of chimeric PKSs, recombined at different domain junctions (KS–AT, ACP–KS, etc.) using different library construction techniques (as described in the build section). Sequence and activity information derived from library screening experiments could be then incorporated into a retrosynthesis algorithm, as detailed in the learn section. Natural selection currently produces new PKSs by screening recombination libraries: computational modeling suggests that recombination is of central importance in generating new polyketide compounds(Callahan et al., 2009). This is borne out in sequence analysis of PKS gene clusters, which suggests that recombination both within clusters and between both inactive and active clusters is a source of new metabolite diversity (Jenke-Kodama et al., 2006; Jenke-Kodama et al., 2005). By learning from selection experiments in the laboratory, one could develop more refined PKS construction techniques.

These kinds of selection experiments are beginning to be performed in a laboratory setting. Kim and coworkers utilized DNA shuffling and screening using a bioassay to substitute portions of the DEBS loading domain in the pikromycin cluster (Kim et al., 2004). However, despite successes with mixing and matching strategies for iterative type II PKSs to make new metabolites (McDaniel et al., 1995), these experiments have yet to become mainstream in modular type I PKS engineering. Nonetheless, very recently Sugimoto and coworkers have utilized recombination within the type I aureothin cluster to produce the related compound luteoreticulin, for which there is no known naturally producing strain (Sugimoto et al., 2014). Random mutagenesis can also be used to improve chimeric proteins after they are artificially recombined in the laboratory. Fischbach and coworkers used this technique to quickly improve non-ribosomal peptide synthetase (NRPS) chimeric proteins by screening with a bioassay (Fischbach et al., 2007). NRPSs are thiotemplated megasynthases similar to PKSs, but use amino acids instead of acyl-CoAs as extender units. It remains to be seen whether similar results would be obtained using directed evolution with PKS domain chimeras. More PKS laboratory evolution studies could create the knowledge necessary to inform design of a retrosynthesis algorithm.

1.1.4. Build

The cost of gene synthesis has recently fallen at an exponential rate and is now low enough to allow for codon optimization of genes the size of PKSs (Cost Per Base of DNA Sequencing and Synthesis; URL: http://www.synthesis.cc/library/carlson_cost%20per_ba- se_oct_2012.png).

However, synthesis can still be expensive for construction of large combinatorial libraries of PKSs. Developments in DNA software like j5 (Hillson et al., 2011), which designs scarless combinatorial constructs that can be assembled using techniques such as Gibson (Gibson et al., 2009), or Golden Gate Assembly (Engler et al., 2009) will be helpful in construction of combinatorial libraries of chimeric PKS proteins. Technologies like MAGE (Wang and Church, 2011), could also be used to generate insertions or deletions up to 30 base pairs or more in a high-throughput fashion. Cross platform robotic automation technologies such as PR–PR (Linshiz et al., 2014; Linshiz et al., 2013), which can take the output from programs like j5 and then set up the required PCRs using many robotics platforms may facilitate large-scale projects and new approaches to PKS engineering.

We have found that initial proof-of-principle studies with engineered PKSs are best done *in vitro*, either in cell lysates containing overexpressed protein(s) or preferably with purified protein(s). *In vitro* systems allow all required substrates (acyl-CoAs, reducing power, etc.) to be supplied exogenously, obviating the need for initial host engineering to supply these precursors *in vivo*. Additionally, *in vitro* systems allow for higher concentrations of protein than *in vivo* systems and reduces background signal, which may be necessary if the initial engineered protein constructs are sub optimally active. Engineering PKSs for production of a wide array of molecules will ultimately necessitate the use of domains and modules from a variety of source organisms. Such constructed

chimeric PKSs will be inherently 'non-monophyletic,' complicating the choice of a suitable host for initial prototyping. Some examples of the use of chaperones to improve PKS solubility (Mutka et al., 2006) or achieve modest improvements in specific activity (Betancor et al., 2008) for *E. coli* expressed proteins have been reported. However, we feel that as the many factors that contribute to successful heterologous protein expression are still poorly understood, it is unlikely that a 'super host' (such as a highly engineered *E. coli* strain) capable of predictable success in soluble expression of all engineered PKSs will emerge in the near term; more probable is the need to prepare a 'suite of production hosts' (Baltz, 2010). As there is currently no *a priori* way to know which host would have the highest likelihood of soluble, functional expression, we focus on recent tools that allow for rapid sampling of optimized, genetically tractable hosts to find ones amenable to pilot studies.

Natural product discovery through the cloning and expression of metagenomic DNA libraries has spurred the development of several vectors with broad host ranges for *E. coli*, *Pseudomonas* and *Bacillus*, among others (Aakvik et al., 2009; Kakirde et al., 2011; Troeschel et al., 2012). Commensurate with the creation of such vectors have been developments of high-throughput transformation and conjugation techniques(Martinez et al., 2004), which would be crucial for adequately sampling large combinatorial libraries of enzyme variants. The recent discovery of a plasmid and the development of inducible promoter systems in *Myxococcus* (Iniesta et al., 2012; Zhao et al., 2008) may lead to increased utility of this genus, as perhaps will the development of a thermophilic, fast-growing myxobacterial strain that successfully produced the NRPS/PKS-based myxochromide at levels substantially higher than the native host (Perlova et al., 2009). Recent examples of host engineering in *Streptomyces* have focused on measures designed to eliminate background production of native secondary metabolites and otherwise reduce the genome in order to increase production of heterologous products and simplify their detection (Gomez-Escribano and Bibb, 2011; Komatsu et al., 2010). Codon-optimization of synthetic DNA to allow for sampling of many different hosts may be challenging and require production of additional tRNAs which could be chromosomally integrated under constitutive promoters. Additional host features that may promote solubility would be expression of native and heterologous chaperones which have been shown in some cases to increase production of non-native products (Betancor et al., 2008; Mutka et al., 2006). Upon demonstration of *in vitro* production of a target compound from soluble protein, *in vivo* production would be attempted in the appropriate host; additional pathway engineering would be needed to supply required acyl-CoAs and a suitable phosphopantetheinyl transferase, if necessary. Efforts in engineering *E. coli* for type I polyketide production were recently reviewed by Yuzawa and coworkers (Yuzawa et al., 2012).

1.1.5. Test

In our experience, many engineered PKSs are insoluble when expressed in heterologous hosts like *E. coli*. Fusing fragments of green fluorescent protein to libraries of candidate chimeric proteins and screening for solubility using a split GFP system (Cabantous and Waldo, 2006) or by utilizing a colony filtration blot (Cornvik et al., 2005) could be used

to prescreen libraries to remove insoluble members, further library screening could then be performed with lower-throughput analytical approaches such as GC/MS and LC/MS. This two-step approach would allow for larger libraries to be effectively screened. Once a soluble, active chimeric protein has been found, analytical techniques have been developed to find enzymatic bottlenecks through detection of covalent intermediates attached to acyl carrier protein peptides (Bumpus and Kelleher, 2008; Dorrestein et al., 2006). These techniques are accessible via a triple quadrupole mass spectrometer; many institutions have core proteomic facilities capable of performing these techniques, yet they appear underutilized in the PKS literature.

High-throughput screens for activities of individual domains have been exploited to engineer domains within modular synthases (recently exemplified in the work of Zhang and coworkers (Zhang et al., 2013)). However, engineering entire PKSs requires detection of the final product, and we believe mass spectrometry is the best technology for this application. Several new technologies have been developed that will aid in greatly improving mass spectrometry throughput. New platforms such as Agilent's RapidFire High-throughput MS Systems use solid phase extraction to achieve ten times the throughput of conventional mass spectrometry screening methods, with sample processing times between 6 and 10 s. Many solid phase chemistries are available and a variety of extraction/separation methods are in development. Nanostructure-Initiator Mass Spectrometry (NIMS) coupled with acoustic deposition has been used to screen enzymes acting on carbohydrates in a high-throughput format (Greving et al., 2012); building on NIMS demonstrations utilizing non-carbohydrate substrates (Rond et al., 2013), this technology has potential application to screening PKSs libraries. Other developments in mass spectrometry such as nano-desorption electrospray ionization, which has been used to obtain mass spectra *in situ* from single bacterial colonies, also have potential to dramatically reduce the laborious sample preparation of traditional approaches (Watrous et al., 2012). The additional throughput of the next generation of mass spectrometry technologies has great potential to enable screening libraries of chimeric PKSs—a new approach to the PKS design–build– test–learn cycle.

1.1.6. Learn

At present, functional chimeric proteins are usually reported, but we suspect that nonfunctional or insoluble chimeric proteins are generally not published. This likely publication bias does not allow the issues of insolubility or lack of function to be addressed systematically, as failure can be as important for generating knowledge as success. We propose the creation of an online database for reporting of unpublished non-functional proteins, perhaps through research gate (URL: http://www.researchgate.net/), DOE's KnowledgeBase (URL: http://kbase.science.energy.gov/) or a similar site. This database would allow for failure analysis by bioinformatics methods, and should bring a new perspective to PKS engineering. Bioinformatic efforts so far (reviewed recently (Boddy, 2014)), focus on identifying or classifying PKS modules and domains, not necessarily yielding information that is directly applicable to engineering PKSs or in failure analysis. Expanding this database to include documentation of failures and successes at every stage of the design–build–test–learn cycle could serve as the

foundation for a retrosynthesis algorithm. This kind of algorithm could allow non-experts to use PKSs for production of compounds of interest, expanding the role of engineered PKSs from niche science to a widespread chemical tool. As an example of the potential of developing algorithms based on an iterative design–build–test–learn PKS engineering approach, Chandran and coworkers developed an algorithm to activate several nonfunctional modules from their earlier Menzella work (Menzella et al., 2005) by swapping the nonfunctional KSs with other KSs that accepted the substrate of interest, observing 160-fold to 1300-fold increases in titers (Chandran et al., 2006).

1.1.7. Conclusions

The gap between the promise of type I PKSs as a retro-synthesis platform and the current state of the art is large. The promise continues to grow as new PKS modules and chemistries are discovered, expanding the potential range of accessible chemicals. Techniques and technologies have been developed that make every stage in the retro-synthesis design–build–test–learn cycle more tractable. Many strategies for the design of chimeric proteins as well as methods for library design offer several options when constructing novel PKSs. DNA synthesis, the increasing power of DNA design software and laboratory automation are enabling scientists to make cloning much less of an obstacle. Many genetically tractable hosts are available to facilitate soluble protein expression. Mass spectrometers are getting better every year, and new techniques are being developed which allow higher throughput and less sample preparation, allowing for library-based approaches in the design–build–test–learn cycle. By both building on the existing polyketide science and harnessing technologies developing in other areas, we hope to see engineered PKSs with commercial applications in medicine and industrial products within 10 years.

2. Chapter 2 – Mechanistic Analysis of an Engineered Enzyme that Catalyzes the Formose Reaction

Including material from submitted work: **Poust, S.,** Piety, J., Bar-Even, K., Louw, C., Baker, D., & Keasling, J. D. (2015). Mechanistic Analysis of an Engineered Enzyme that Catalyzes the Formose Reaction.

2.1. Abstract

An enzyme that catalyzes the formose reaction, termed "formolase," was recently engineered through a combination of computational protein design and directed evolution. We have investigated the kinetic role of the computationally designed residues, and further characterized the enzyme's product profile. Kinetic studies illustrated that the computationally designed mutations were synergistic in their contributions towards enhancing activity. Mass spectrometry revealed that the engineered enzyme produces two products of the formose reaction, dihydroxyacetone and glycolaldehyde, with the product profile dependent on the formaldehyde concentration. We further explored the effects of this product profile on the thermodynamics and yield of the overall carbon assimilation from the Formolase pathway to help guide future efforts to engineer this pathway.

Graphical Abstract:
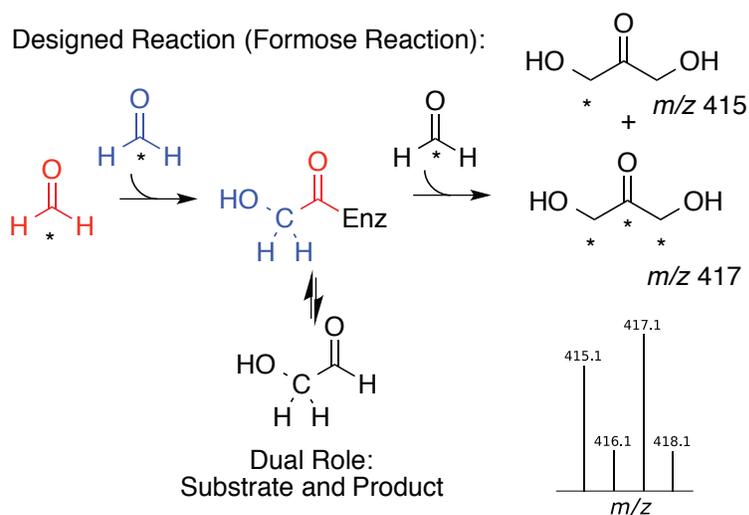
2.2. Introduction

The formose reaction, discovered in 1861 by Butlerow, couples together formaldehyde (FALD) molecules to form sugars in the presence of a base (Breslow, 1959). This reaction is believed to be a source of prebiotic sugars (Ruiz-Mirazo et al., 2014); however until recently (Siegel et al., 2015), no enzyme had been identified which catalyzes the formose reaction. To find a catalyst for the last step in a pathway designed to assimilate carbon from formate, a low level of carbon-coupling activity on FALD was identified in benzaldehyde lyase (BAL) (González and Vicuña, 1989; Siegel et al., 2015). The low level of coupling activity was subsequently improved through a combination of enzyme design and directed evolution, until no detectable level of the native BAL activity and only the formose reaction was observed, resulting in the "formolase" enzyme (Siegel et al., 2015).

BAL natively catalyzes the reversible conversion of (R)-benzoin to benzaldehyde (**Figure 2.1** A) using thiamine diphosphate (TPP) as a cofactor (Chakraborty et al., 2008). This enzyme has been utilized to synthesize of a variety of α-hydroxy ketones, including substrates without a phenyl ring (Müller et al., 2013), which suggested its tractability for protein design/evolution (Siegel et al., 2015). Furthermore, the base-catalyzed formose reaction in solution is known to form a heterogeneous mixture of products including the two-carbon product glycolaldehyde (GALD) (Breslow, 1959). However, when the reaction is catalyzed by thiazolium salts in organic solvents (which are similar to the thiamine diphosphate cofactor utilized by BAL), the formose reaction has been reported to be highly selective for the three carbon product dihydroxyacetone (DHA), with no observed production of the two-carbon product GALD (Matsumoto et al., 1984).

The design efforts to convert BAL into an enzyme specific for the formose reaction utilized a single 3-carbon reaction intermediate docked into BAL as input for the RosettaDesign program (Siegel et al., 2015), which yielded 4 mutations (A394G, A480W, A28S, G419N) that improved formose activity 13-fold over BAL (Des0, see below). In this work, we sought to understand the role of these computationally designed residues in improving the formolase kinetics by determining the kinetic parameters when each designed residue was reverted to the wildtype enzyme. Additionally, we further examined the full product profile utilizing GC-MS, particularly with regards to the role of GALD as the literature precedent for the role of this intermediate in the overall reaction is somewhat contradictory (Breslow, 1959; Matsumoto et al., 1984). Using the new reaction model developed in this study we assessed new potential routes of carbon assimilation by calculating thermodynamic driving force and maximum yields.

2.3. Results and Discussion

Des0 has four mutations within the active site: A394G, A480W, A28S and G419N. The mutations A394G and A480W are predicted to be concerted mutations in that a small

amount of room is introduced into the active site by the A394G mutation such that the larger amino acid W can cover the top of the transition state. A28S and G419N add additional hydrogen bonds to the hydroxyl groups being generated in the reaction. The G419N introduced a hydrogen bond to a hydroxyl that is new in the DHA reaction. It was unclear if the A28S mutation would actually be beneficial as it adds a hydrogen bond to a part of the intermediate that also occurs in the benzoin formation reaction (**Figure 2.1** B).

To determine the effects of designed mutations, each designed residue was reverted and assayed for activity (**Figure 2.1** C) using a spectrophotometric-coupled enzyme assay in which DHA or GALD is reduced by the NADH-dependent enzyme glycerol dehydrogenase (Yamada et al., 1982). As suspected, the A28S mutation did not have a significant effect on activity. Reverting G419N dropped activity ~4-fold, suggesting that new contacts with the intermediate are made. While A394G was not predicted to make new contacts it was predicted that removal of the methyl group via the A394G mutation was important for the optimal placement of A480W. This is reflected in that reverting the A394G mutation dropped activity over ~5-fold. Finally the mutation with the most significant increase in molecular contacts, A480W, appears to be critical for this reaction as reverting it decreased activity ~12-fold. While G419N and A394G both appear to be important for the increased catalytic efficiency, they are only important in the context of A480W. This suggests that A480W is reorienting the substrate or reaction intermediates such that the additional contacts with G419N are made. These data also illustrate the power of computational design in finding the synergistic mutations A394G/A480W that would have been difficult to find within a single round of evolution.
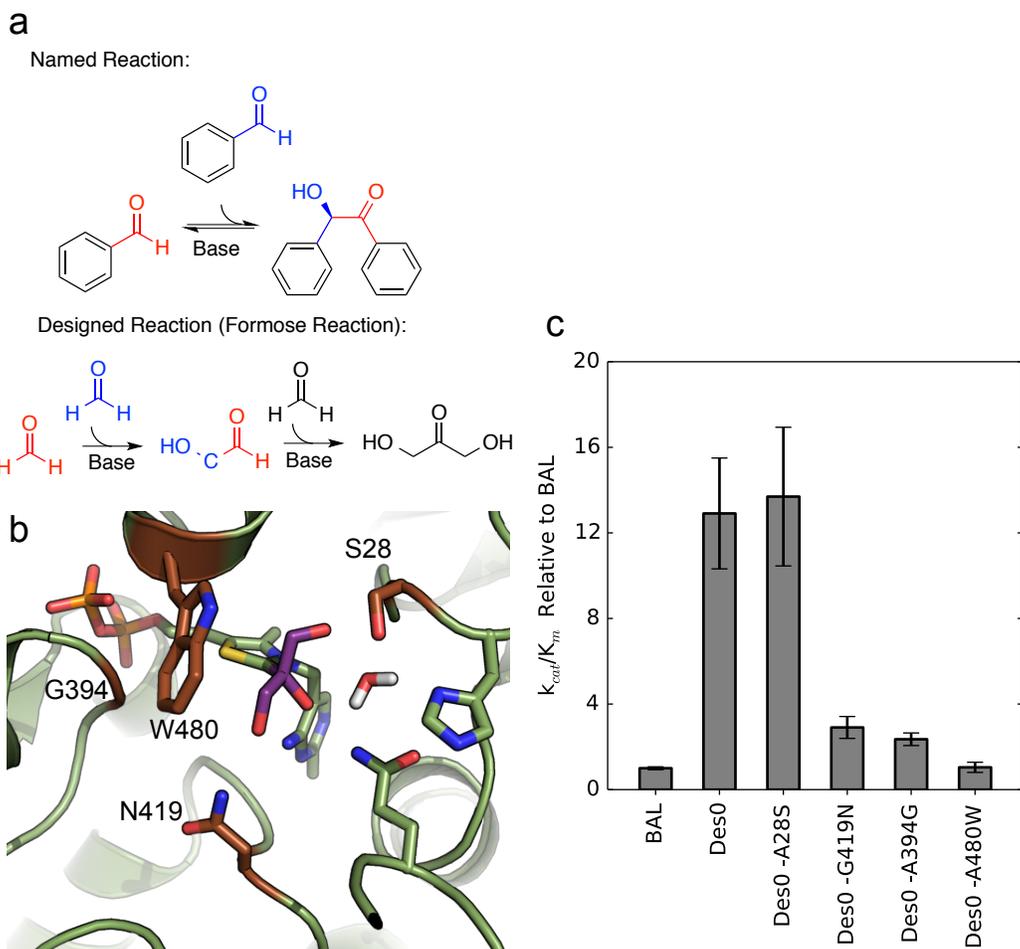
Figure 2.1. A) The native reaction catalyzed by benzaldehyde lyase (BAL): reversible ligation of (R)-benzoin to benzaldehyde. The Formolase couples formaldehyde (FALD) to form glycolaldehyde (GALD) and dihydroxyacetone (DHA). B) The Rosetta generated model of the Des0 variant active site, highlighting the computationally designed residues which improved the activity. C) Formose reaction k_{cat}/K_m parameters when the designed residues are individually reverted to the wildtype residues, relative to wildtype benzaldehyde lyase. Error bars represent the 95% confidence interval.

As previously mentioned, the continuous spectrophotometric assay utilized to determine the kinetic parameters couples the formose activity to the consumption of NADH via glycerol dehydrogenase. Glycerol dehydrogenase has activity towards both potential formose products (DHA and GALD) (Siegel et al., 2015; Yamada et al., 1982). While the previous studies to characterize the engineered enzyme utilized methods to specifically observe DHA formation, there were no significant efforts to determine if GALD was also formed. To fully characterize the product profile quantitatively, we adapted a GC-MS derivatization method which has been demonstrated for hydroxycarbonyl compounds, such as GALD and similar compounds (Spaulding and Charles, 2002; Spaulding et al., 1999). This method utilizes a hydroxylamine-containing, water-soluble derivatization agent, O-(2,3,4,5,6-pentafluorobenzyl) Hydroxylamine (PFBOA), to form oximes from carbonyl compounds in aqueous solution, facilitating their extraction into an organic layer and improving their chromatographic properties. Once extracted, compounds are silated using N,O-Bistrifluoroacetamide (BSTFA), to further improve their

14

chromatographic properties. Similar techniques have previously been utilized for enzymatic assays (Kobayashi and Kawai, 1982), and PFBOA has found a wide variety of applications (Cancilla and Que Hee, 1992). Using known standards, we verified that these techniques were linear for 7 point standard curves with GALD and DHA, with $R^2 > 0.99$ (**Figures 8.1.1** and **8.1.2**).
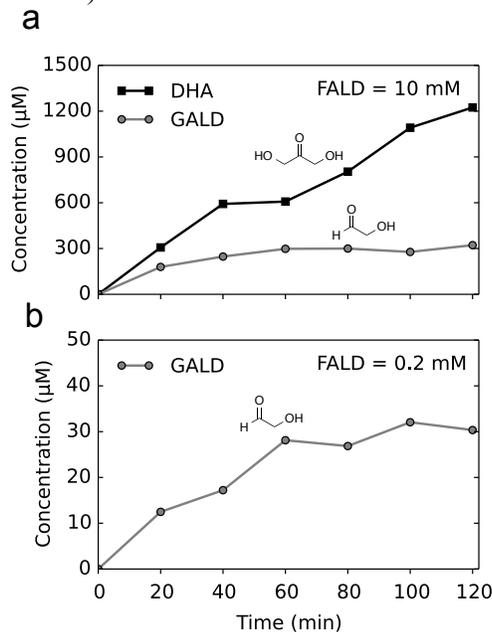
Figure 2.2. Product profiles of the Formolase at high and low concentrations of FALD. At high FALD concentrations DHA is the primary product, while at low FALD concentrations GALD is the primary product.

To probe the overall mechanism of the enzyme catalyzed reaction, we determined if the product profile was concentration dependent. Specifically, we monitored production of DHA and GALD ove time for the fully evolved "formolase" (Siegel et al., 2015) at both a high FALD concentration (10 mM) and 50-fold lower concentration (0.2 mM). Reactions were performed with 12.5 µM enzyme over a period of 120 minutes, with time points taken every 20 minutes, by heating samples to 95°C for 2 minutes (BAL has previously been shown not to be thermostable (Janzen et al., 2006)). As shown in **Figure 2.2** A, at high concentrations of FALD we observed that DHA was the primary product. At lower, more physiologically relevant, concentrations (**Figure 2.2** B) GALD was the primary product. However, at later time points (100 and 120 min), small (<10 µM, below the limit of reliable quantification, data not shown) amounts of DHA became visible. Thus, these data show that the product profile of the formolase enzyme is a function of the FALD concentration.
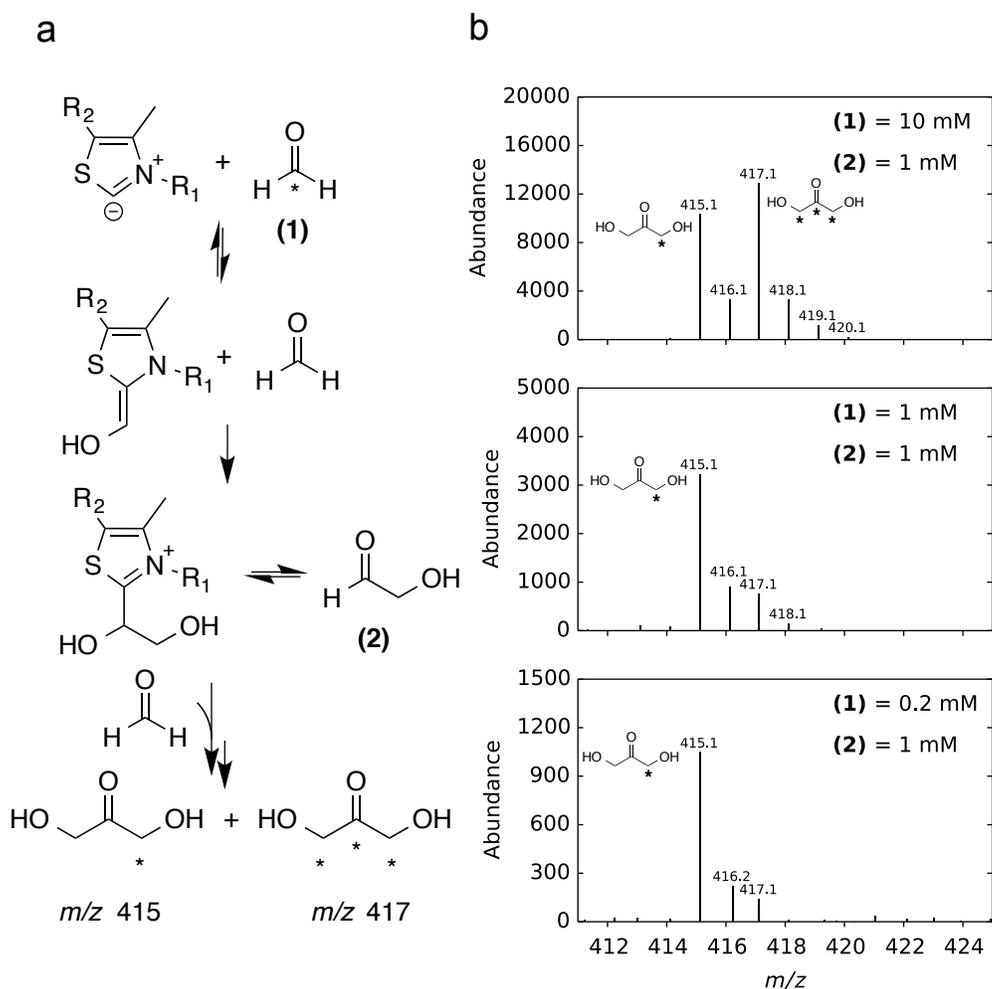
Figure 2.3. A) Proposed mechanism and competition experiment for the overall reaction catalyzed by the formolase. FALD can reversibly form an enamine intermediate with the ylide, which can then perform one round of ligation with another FALD molecule to form the GALD-Enzyme intermediate. GALD can leave and reenter the active site, or undergo a further ligation with FALD to form DHA. In the competition experiment, varying concentrations of labeled FALD **(1)** (indicated with a *) is competed against unlabeled GALD **(2)** to form various mixtures of singly and triply labeled DHA (derivatized *m/z* values shown, structures in **Figure 8.1.5**). B) MS spectra from the competition experiment. As the concentration of FALD increases, a higher percentage of triply labeled DHA is visible in the spectra.

Having conclusively shown that GALD is a product of the formolase enzyme, we further elucidated the role of GALD in the overall reaction. Specifically, we assessed if GALD was either a final product, or if it was able to leave and re-enter the active site. In the later case we asked whether it could function both as a product and a substrate for DHA production. To test this, we designed a competition experiment in which ^{13}C labeled FALD at various concentrations (0.2 mM, 1 mM and 10 mM) was competed against unlabeled GALD at 1 mM concentration. This experiment is outlined in the context of a proposed mechanism for the formolase in **Figure 2.3** A. The formation of GALD from two FALD molecules, as well as the formation of DHA from GALD and FALD, are both highly thermodynamically favorable Δ_rG'<-12 and <-6.0 kcal/mol, respectively, at

16

physiological pH 7, Ionic strength = 0.25 M, 1 mM reactant concentration) (Flamholz et al., 2012; Noor et al., 2013), so the reverse reactions are expected to be negligible.

As shown in **Figure 2.3** B, the production of singly labeled-derivatized DHA (*m/z* 415) at all three concentrations of FALD demonstrates that GALD can be used as a substrate for the enzyme-catalyzed formose reaction. To outcompete the entry of GALD and produce a majority of fully labeled DHA (*m/z* 417) from FALD, a ratio of 10:1 FALD:GALD is required. These results show that DHA can be formed both through a pathway in which the GALD-Enzyme intermediate never leaves the active site in addition to a pathway utilizing free GALD previously released into bulk solution, and subsequently consumed.

The mechanism proposed in **Figure 2.3** A is consistent with the product ratio being highly dependent on the FALD concentration, as observed in **Figure 2.2**. We propose that the product ratio (DHA/GALD) is determined by the partitioning of GALD-enzyme intermediate. The rate of GALD formation is first order in the concentration of GALD-enzyme intermediate, whereas the formation of DHA-enzyme intermediate depends on the concentration of both FALD and the GALD-enzyme intermediate. As the concentration of GALD-enzyme intermediate is the same for these competing reactions, at low FALD concentrations, GALD formation occurs at a faster rate, whereas at high FALD concentrations the formation of DHA-enzyme intermediate occurs at a faster rate.
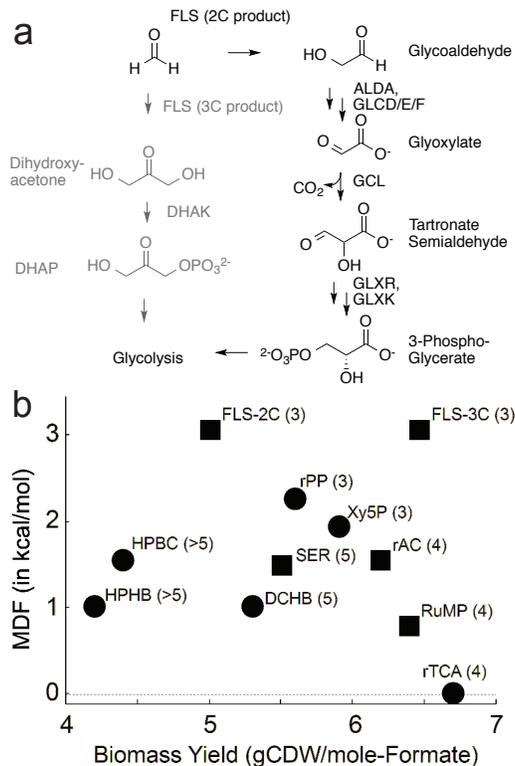
Figure 2.4. A) Potential carbon assimilation pathways for the formolase pathway, depending on the product profile. The DHA three-carbon pathway is indicated in grey, while the GALD-based pathway is indicated in black. B) Thermodynamics and carbon utilization efficiencies. Biomass yield, in gram cellular dry weight (gCDW) per mole of formate consumed, was calculated using flux balance analysis and the core metabolic model of *E. coli* supplemented with pathway enzymes and without considering ATP maintenance (Bar-Even et al., 2013). The Min-max Driving Force (MDF) is the lowest value of -ΔrG' in the pathway (i.e., the reaction(s) with the smallest chemical driving force) (Noor et al., 2014). Squares correspond to pathways that can directly assimilate formate, while circles mark carbon fixation pathways that can indirectly assimilate formate by its initial oxidation. Numbers in parentheses indicate the number of foreign enzymes that need to be expressed in *E. coli* to establish an active pathway. Pathway abbreviations are as follows: HPHB, 3-hydroxypropionate-4-hydroxybutyrate cycle; HPBC, 3-hydroxypropionate bicycle; DCHB, dicarboxylate-4-hydroxypropionate cycle; SER, serine cycle; rPP, reductive pentose phosphate cycle/Calvin-Benson-Bassham cycle; Xy5P, xylulose 5-phosphate cycle/dihydroxyacetone cycle); rAC, reductive acetyl-CoA pathway; RuMP, ribulose 4-phosphate cycle, FLS-3C, formolase pathway via dihydroxacetone; FLS-2C formolase pathway via GALD and rTCA, reductive TCA cycle.

We next sought to explore the potential consequences of the observed product profile on metabolic assimilation from the formolase pathway. Assimilation of DHA requires a single phosphorylation step to enter glycolysis. Conversely, the pathway for GALD is much longer, involving several oxidation steps, carbon-coupling and ATP-dependent activation (Baldomà and Aguilar, 1987; Chang et al., 1993; LeBlanc and Mortlock, 1971). These two possible pathways are shown in **Figure 2.4** A.

To quantitatively compare the DHA- and GALD-based pathways we predicted chemical driving force and yields for each of these two assimilation pathways as compared to other naturally occurring carbon fixing pathways (**Figure 2.4** B) (Noor et al., 2014). Both formolase assimilation pathways are predicted to have the smallest driving force at the same step (formyl-CoA synthetase and FALD dehydrogenase) but the value at this step is quite high (~3 kcal/mol), so backwards flux is not predicted to be an issue for either pathway. However, the yield for the 2C pathway is reduced as the decarboxylation at the level of glyoxylate condensation reduces the efficiency by increasing the ATP cost (a fourth of the ATPs utilized for formate reduction are "wasted"). Nevertheless, the 2C assimilation route compares favorably to other naturally occurring one-carbon pathways.

This mechanistic work suggests several avenues for future efforts engineering the formolase pathway. Since the GALD-based pathway is predicted to have a lower carbon yield than the DHA-based pathway, selectivity for DHA, particularly at low FALD concentrations is more desirable. This can likely be achieved through computational design and evolutionary optimization efforts of the engineered formolase enzyme. Alternatively, other thiamine diphosphate dependent enzymes with carbon-coupling activity toward smaller substrates may also prove to be promising starting points for the development of alternative, more specific formolases (Müller et al., 2013). Carbon assimilation pathways via tetrose or hexose sugars could also be envisioned through a combination of isomerase and formolase engineering.

As opposed to direct protein engineering, compartmentalization could be used to produce DHA selectively (Yeates et al., 2010). Since production of majority DHA requires highly toxic concentrations of FALD (10 mM), compartmentalizing this portion of the Formolase pathway may allow high concentrations of FALD to be used. In addition, since GALD is both a substrate and a product of the enzyme, compartmentalizing the formolase could ensure that any GALD produced would be further transformed to DHA rather than entering alternative assimilation pathways.

2.4. Methods

Protein Purification: For kinetic parameter determination, constructs were transformed into chemical competent *E. coli* BL21 (DE3). A single colony was picked and cultured overnight in 5 mL LB medium at 37˚C. This culture was decanted into 0.5 L of TB medium and incubated at 37˚C until mid-log phase. Expression of the gene of interest was induced with 1 mM IPTG. After 30 hours at 18˚C, the cells were pelleted by centrifugation, re-suspended in 25 mM phosphate buffer, pH 8, 1 mM $MgSO_4$, 0.1 mM thiamine pyrophosphate (TPP) and 20 mM imidazole and lysed by sonication in the presence of lysozyme. Lysates were cleared of cell debris by centrifugation and purified using Talon resin in a gravity fed column (Clontech). The resin was washed three times with 25 mM phosphate buffer, pH 8, 1 mM $MgSO_4$, 0.1 mM thiamine pyrophosphate (TPP) and 20 mM imidazole. Elution was done with 25 mM phosphate buffer, pH 8, 1 mM $MgSO_4$, 0.1 mM thiamine pyrophosphate (TPP) and 200 mM imidazole Eluate was concentrated to approximately 1 mL prior to dialysis against 25 mM phosphate buffer,

pH 8, 1 mM $MgSO_4$, 0.1 mM thiamine pyrophosphate (TPP). For GC-MS assays, HisPur Ni-NTA Beads (Thermo Scientific) were used to purify the final formolase variant.

Mutations of computationally designed resides: Kunkel mutagenesis was performed as described.

Kinetic Constant Determination: Kinetic constants for BAL and FLS were measured over a one-hour period in 100 mM KPO_4 buffer, pH 8.0, 1 mM $MgSO_4$, 0.1 mM TPP, 50 µg/mL glycerol dehydrogenase, and 0.8 mM NADH. Enzyme concentrations ranged from 1 to 40 µM, and FALD concentrations ranged from 1.3 mM to 20 mM. NADH formation was monitored at a wavelength of 340 nm. Kinetic constants for each enzyme using FALD as a substrate were calculated using regression fitting of rate of product production as a function of substrate concentration to a linear equation as saturation was not obtained at 20 mM.

GC-MS analysis for DHA and GALD: Enzymatic reactions were performed in 100 mM KPO_4 buffer, pH 8.0, 1 mM $MgSO_4$, and 0.5 mM TPP. For time course samples, the enzyme concentration was 12.5 µM, and either 10 mM or 0.2 mM FALD was used. To quench samples, 60 µL was aspirated from each reaction and quenched by heating at 95°C for 2 min in a PCR block. For competition assays 5 µM enzyme, 1 mM GALD and 0.2, 1 or 10 mM ^{13}C FALD were used (Cambridge Isotope Laboratories, Tewksbury, MA). Competition reactions were quenched after 20 min, and a total reaction volume of 500 µL was used. After quenching, samples were derivatized in a glass vial insert by the addition of 10 µL 200mM PFBOA to 50 µL of the quenched time course samples or 100 µL 200 mM PFBOA to the 500 µL competition samples. Samples were then incubated for 1 hr at room temperature. 100 µL hexanes was added to the samples using glass micropipettes and the samples were vortexed and phase separation was allowed for 5 min. 20 µL of the organic layer was aspirated, placed in a glass GC insert, and then 5 µL BSTFA was added to silylate PFBOA derivatives.

Electron ionization (EI) GC-MS analyses were performed with a model 7890A GC (Agilent) with a DB-5 fused silica capillary column (30-m length, 0.25-mm inner diameter, 0.25-m film thickness; J&W Scientific) coupled to an HP 5975C series mass selective detector; 1- µL injections were performed by a model 7683B autosampler. The GC oven was programmed from 50°C (held for 1 min) to 150°C at 15°C/min, to 300°C at 30°C/min and then held for 1 min; the injection port temperature was 250°C, and the transfer line temperature was 280°C. The carrier gas, ultra-high purity helium, flowed at a constant rate of 1 ml/min. Injections were splitless, with the split turned on after 0.5 min. For full-scan data acquisition, the MS scanned from 50 to 500 atomic mass units. For DHA and GALD quantification, authentic standards (Sigma) were used to generate calibration curves using the ion resulting from loss of a methyl group from the trimethylsilyl group during electron impact ionization ($[M - CH_3]^+$: *m/z* 312 for GALD and 414 for DHA) (Spaulding et al., 1999). See **Figure 8.1.5**.

3. Chapter 3 – Understanding the Role of Histidine in the GHSxG Acyltransferase Active Site Motif: Evidence for Histidine Stabilization of the Malonyl-Enzyme Intermediate

Including material from published work: **Poust, S.,** Yoon, I., Adams, P.D., Katz, L., Petzold, C.J., & Keasling, J. D. (2014). Understanding the Role of Histidine in the GHSxG Acyltransferase Active Site Motif: Evidence for Histidine Stabilization of the Malonyl-Enzyme Intermediate. *PloS One 9,* e109421.

3.1. Abstract

Acyltransferases determine which extender units are incorporated into polyketide and fatty acid products. The ping-pong acyltransferase mechanism utilizes a serine in a conserved GHSxG motif. However, the role of the conserved histidine in this motif is poorly understood. We observed that a histidine to alanine mutation (H640A) in the GHSxG motif of the malonyl-CoA specific yersiniabactin acyltransferase results in an approximately seven-fold higher hydrolysis rate over the wildtype enzyme, while retaining transacylation activity. We propose two possibilities for the reduction in hydrolysis rate: either H640 structurally stabilizes the protein by hydrogen bonding with a conserved asparagine in the ferredoxin-like subdomain of the protein, or a water-mediated hydrogen bond between H640 and the malonyl moiety stabilizes the malonyl-O-AT ester intermediate.

Graphical Abstract:
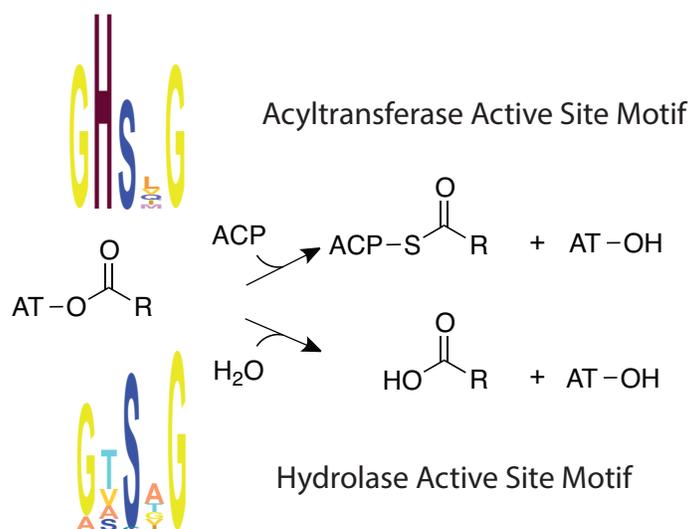
3.2. Introduction

During fatty acid and polyketide biosynthesis, acyltransferases (ATs) catalyze transfer of the acyl group from malonyl-, methylmalonyl-, or other short chain acyl-CoAs to acyl carrier proteins (ACPs) using a serine-histidine catalytic dyad. The OH-group of serine acts as a nucleophile to attack the thioester bond of the acyl donor, forming a covalent acyl-*O*-ester intermediate. The acyl group is subsequently transferred to the ACP via a bi bi ping pong mechanism (**Figure 3.1** A) (Keatinge-Clay et al., 2003).

For efficient transfer to the ACP, acyl-AT complexes must be stable in solution. However, ATs have been described as distant relatives of the α/β hydrolase superfamily (Keatinge-Clay et al., 2003; Serre et al., 1995), and hydrolysis is an unproductive side reaction, effectively wasting the activated acyl-CoA substrate. The active sites for α/β hydrolases carry a GxSxG motif, whereas ATs possess the highly conserved GHSxG motif in the active site, suggesting that the histidine is important for acyl transfer. These conserved GHSxG and GxSxG motifs are illustrated in sequence logos from pfam PF00698 (Acyl_transf_1*)* and an evolutionarily related family from the same pfam clan, PF01734 (Patatin), which have phospholipase activity (**Figure 3.1** B and C) (Bateman et al., 2004). The sequence logo for the AT family (**Figure 3.1** B) shows that that the histidine in the GHSxG motif is highly conserved.
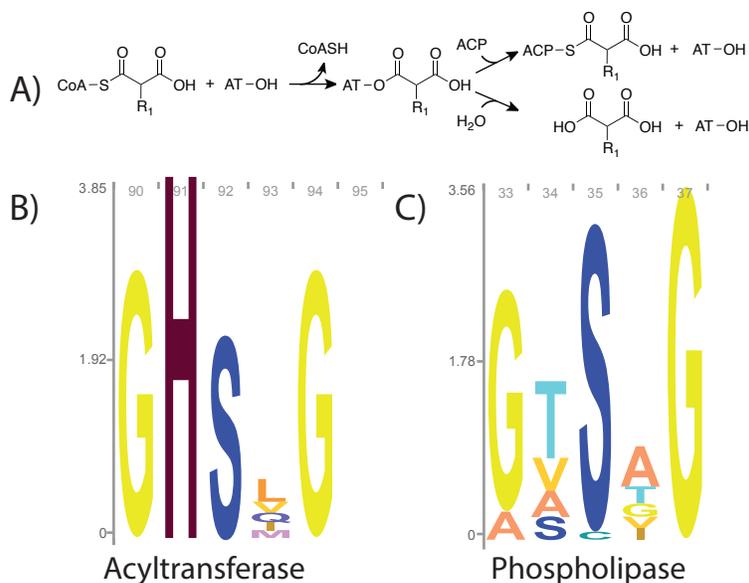
Figure 3.1. Acyltransferase reaction mechanism and sequence logos. A) AT reaction mechanism: ATs utilize a ping-pong mechanism to transfer acyl groups from CoA to ACP (for most ATs, R1=H or CH3). Hydrolysis, shown in the bottom branch is a competing, unproductive reaction. B) Sequence logo for the motif containing the active site serine of pfam PF00698, which encompasses ATs from fatty acid and polyketide biosynthesis. C) Sequence logo for pfam PF01734, which are evolutionarily related phospholipases. Logos created using Skyline (Wheeler et al., 2014).

Dreier and coworkers proposed that the conserved histidine in the GHSxG motif could act as an alternative catalytic nucleophile responsible for the observed transacylation activity of an AT mutant in which the active site serine was changed to alanine (Dreier et al., 2001). This hypothesis was refuted by Szafranska and coworkers, who found the same serine to alanine mutant had no acyl transfer activity upon additional purification of the AT (Szafranska et al., 2002). We sought to further investigate and clarify the role of the highly conserved histidine in acyl transfer reactions. Specifically, we examined three variants of the GHSxG motif in the malonyl-CoA specific yersiniabactin synthase AT domain: S641A, H640A, and the double mutant H640A+S641A in the context of the full PKS module.

3.3. Materials and Methods

His-tagged constructs were expressed in *Escherichia coli* BLR, and purified using Ni-NTA chromatography followed by anion exchange chromatography using a HiTRAP Q column (GE Healthcare). We observed that Ni-NTA chromatography alone was insufficient to remove background acyltransfer activity from negative controls (e.g. the double mutant S641A+H640A). For each variant, we measured rates of malonyl-CoA hydrolysis using a steady-state coupled fluorometric assay, (Dunn et al., 2013; Molnos et al., 2003) at physiological concentrations of malonyl-CoA (Bennett et al., 2009). We also investigated the formation of malonyl-O-AT and the subsequent intraprotein transfer to form malonyl-S-ACP. We accomplished this by examining peptide acylation using high-resolution mass spectrometry after trypsin digestion of the protein following the *in vitro* loading/transfer reactions (Bumpus and Kelleher, 2008). Detailed information on protocols is provided in the Supplementary Information in Appendix 8.2.

3.4. Results and Discussion

Mutating H640 to alanine (mutant H640A) resulted in an approximately 7-fold increase in the malonyl-CoA hydrolysis rate over the wildtype enzyme (WT) rate (**Table 3.1**). Hydrolysis rates were measured at a concentration of 35 μM malonyl-CoA, which is the concentration of malonyl-CoA in exponentially growing *E. coli* as measured by Bennett and coworkers (Bennett et al., 2009). These values likely represent k_{cat} values, as the K_m for cognate acyl-CoA hydrolysis by other acyltransferases has been previously measured to be in the low uM range (Dunn et al., 2013). As expected, when the active site serine was mutated to an alanine, hydrolysis was not detectable (Table 1, S641A and S641A+H640A). Despite the increased rate of malonyl-CoA hydrolysis, the H640A mutant was still capable of transferring the malonyl moiety to the ACP when incubated with 35 μM malonyl-CoA. Loading/transfer reactions were quenched with 50% acetonitrile after 20 seconds and we detected similar amounts of malonyl-S-ACP (monoisotopic peptide m/z = 1141.0562, z = 4) for both wildtype and H640A (**Figure 3.2** A). As with the hydrolysis assay, the S641A and S641A+H640A mutants had no detectable transacylation ability. Additionally, we observed malonyl-O-AT complex formation for both the wildtype (monoisotopic peptide m/z = 1000.9903, z = 4) and H640A (monoisotopic peptide m/z = 984.4849, z = 4) PKSs (**Figure 3.2** B).

Table 3.1. Hydrolysis rates for yersiniabactin AT mutants at a concentration of 35 μM malonyl-CoA.

Variant	$v/[E]_o$ (min^{-1})	Fold increase over wildtype
Wildtype	0.39 +/- 0.09	1
H640A	3.01 +/- 0.32	7.7
S641A	n.d.	
H640A+S641A	n.d.	

n.d. = below limit of detection; error bars are the standard deviation of 3 replicates
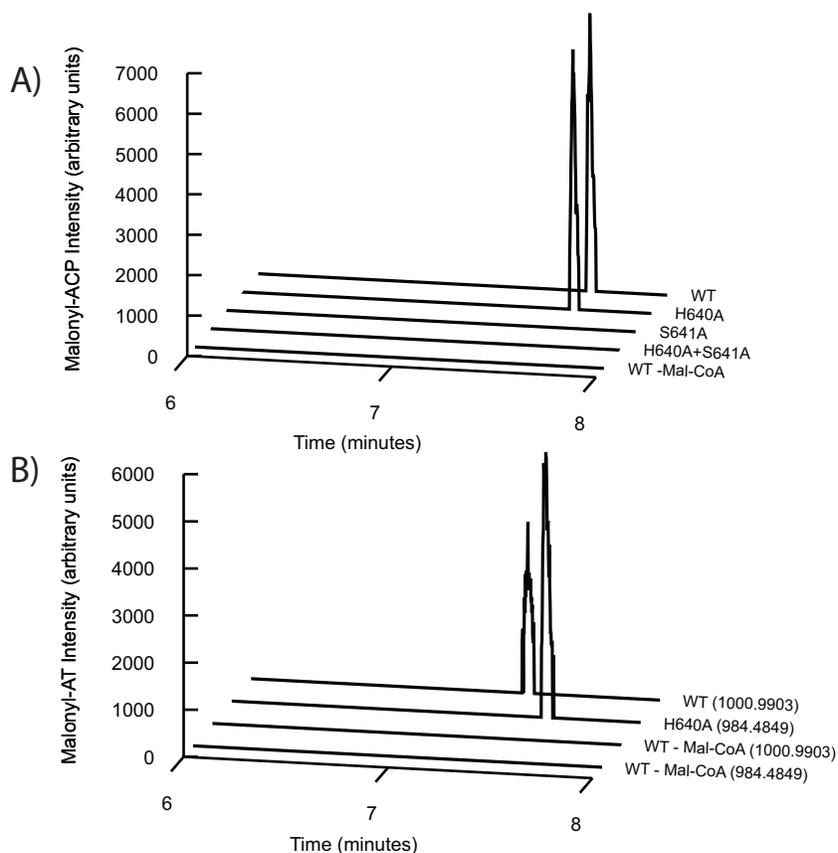
Figure 3.2. Acylation and transacylation activities of WT, H640A, S641A, and H640A+S641A. A) Transacylation activity as observed by high-resolution LC/QTOFMS. Data for the variants (WT, H640A, S641A, H640A+S641A) incubated with malonyl-CoA and a wildtype negative control without malonyl-CoA are shown (WT – Mal-CoA). B) Formation of malonyl-AT complex for wildtype (monoisotopic peptide m/z = 1000.9903, z = 4) and H640A (monoisotopic peptide m/z = 984.4849, z = 4) as observed by high-resolution LC/QTOFMS. Data for wildtype and H640A (WT, H640A) incubated with malonyl-CoA as well as a wildtype negative control without malonyl-CoA (WT – Mal-CoA) are shown. The mass for each chromatogram is shown in parenthesis to the right. Addi-tional details on chromatogram preparation in the Supplementary Information in Appendix 8.1.

The question immediately arises as to why the removal of H640 increases the rate of hydrolysis. To address this issue, we examined the structure of the AT from DynE8, which has a covalently bound malonate (**Figure 3.3**) (Liew et al., 2012). DynE8 is an iterative PKS involved in enediyne biosynthesis, and the DynE8 AT has 28% amino acid

24

identity to the yersiniabactin AT. In this structure, the histidine in the GHSxG motif forms a hydrogen bond with a conserved asparagine in the ferredoxin-like subdomain (**Figure 3.3**). Through this hydrogen bond, the histidine may structurally stabilize the protein, reducing the hydrolysis rate. A water-mediated hydrogen bond between histidine in the GHSxG motif and the carbonyl oxygen in the AT ester bond is also present in the DynE8 structure (**Figure 3.3**), potentially stabilizing the malonyl-O-AT complex and slowing the rate of hydrolysis. We propose two possible mechanisms for this stabilization: either the water-mediated hydrogen bond favors the sp^2 hybridization of the carbonyl oxygen in the malonyl ester bond over the sp^3 hybridization of the tetrahedral transition state for hydrolysis; or the ordering of water provided by the hydrogen bonding network prevents water molecules from attacking the nearby carbonyl. We speculate that as the functional oxyanion hole forms upon the AT binding the ACP to facilitate transacylation (as proposed by Keatinge-Clay and coworkers) (Keatinge-Clay et al., 2003), the stabilizing hydrogen bonding interaction would be disrupted. Future crystallographic studies of AT active site mutants may elucidate the exact mechanism of stabilization.
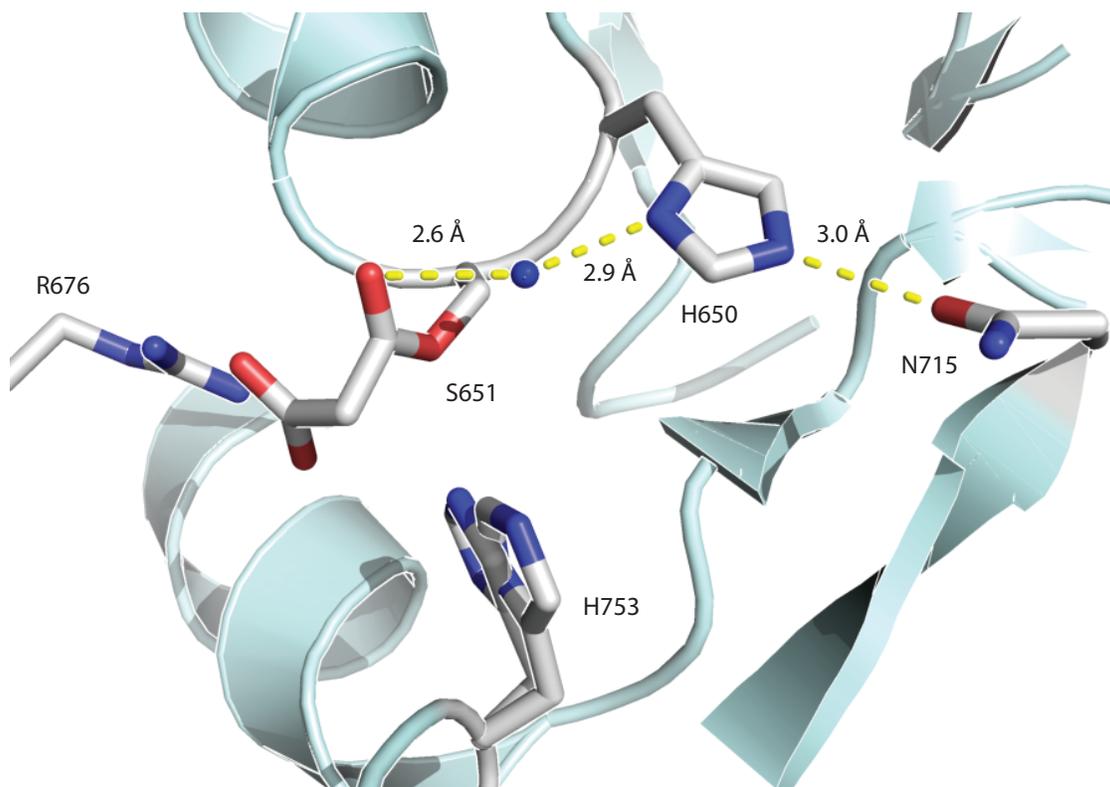
Figure 3.3. Crystal structure with malonate of the acyltransfserase from DynE8, an iterative type I PKS. Residues involved in catalysis are labeled and shown as sticks. The hydrogen bonding water highlighted in the text is shown as a sphere. **Figure 3.3** was prepared using Pymol from the PDB entry 4AMP (Liew et al., 2012).

In summary, we have proposed a stabilizing role of the conserved histidine in the GHSxG active site motif of the yersiniabactin synthase AT domain. To generalize, this suggests that ATs have evolved to protect acyl intermediates, functionally diverging from their α/β hydrolase relatives. Future work examining analogous histidine to alanine mutations in

other AT domains would further support the role of the conserved histidine in stabilizing the malonyl-O-AT complex.

4. Chapter 4 – Divergent Mechanistic Routes for the Formation of gem-Dimethyl Groups in the Biosynthesis of Complex Polyketides

4.1. Abstract

While the presence of *gem*-dimethyl groups in polyketide-derived natural products add steric bulk and, accordingly, lend increased stability to medicinal compounds, our ability to rationally incorporate this functional group in modified natural products is limited. To characterize the mechanism of *gem*-dimethyl group formation, with a goal toward engineering of novel compounds containing this moiety, we characterized the *gem*-dimethyl producing polyketide synthase (PKS) modules of yersiniabactin and epothilone using mass spectrometry. Our work demonstrated, contrary to the canonical understanding of reaction order in PKSs, that methylation can precede condensation in *gem*-dimethyl producing PKS modules. Experiments showed that both PKSs are able to use dimethylmalonyl-ACP as an extender unit. Interestingly, for epothilone module 8, dimethylmalonyl-ACP appeared to be the sole route to form a *gem*-dimethylated product while the yersiniabactin PKS could methylate pre- or post-ketosynthase condensation.

Graphical Abstract:
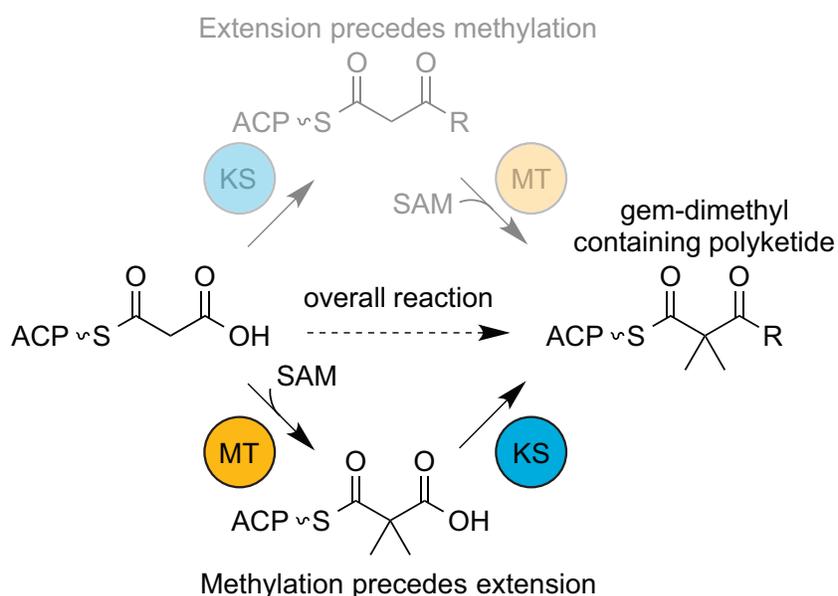
4.2. Introduction

Approximately 10% of all approved drugs contain a geminal dimethyl group. The introduction of this group into compounds can decrease their rates of chemical and metabolic degradation, thus improving the efficacy of drugs (Burkhard et al., 2010). Consequently, there is great potential utility for the regiospecific incorporation of *gem*-dimethyl groups in engineered polyketide products. Regiospecific methylation of polyketide products using organic semi-synthesis is challenging. To enable bio-based approaches to form new *gem*-dimethyl containing pharmaceutical agents, and to clarify discrepancies with the current mechanistic understanding, we sought to determine the reaction mechanism of PKS-based *C*-methylation (Keatinge-Clay, 2012).

Type I polyketide synthases catalyze Claisen condensations and tailoring reactions in an assembly line fashion to elaborate a remarkably diverse collection of secondary metabolites, many of which have medicinal and industrial applications. Polyketide chains are extended by sequential homologation reactions between ketosynthase (KS)-bound thioesters and α-carboxy building blocks (traditionally malonyl- or methylmalonyl-acyl carrier protein (ACP)) to form β-keto-polyketides. Tailoring reactions, such as reduction of the β-keto group on the growing polyketide chain, and subsequent sulfonation or *O*-methylation of the corresponding β-hydroxyl, customarily follow condensation. Incorporation of methyl or *gem*-dimethyl groups by methyltransferase (MT) containing PKS modules is thought to follow this biosynthetic paradigm (Fischbach and Walsh, 2006; Hertweck, 2009; Keatinge-Clay, 2012).

The currently accepted view of PKS-based reactions is that KS-mediated condensation precedes all subsequent events that take place within the module, including, when an MT domain is present, mono- or dimethylation of the β-ketoacyl-ACP (**Scheme 4.1** B, Route 1) (Keatinge-Clay, 2012). However, another route is possible in which methylation precedes condensation (**Scheme 4.1** B, Route 2). Both pathways require the formation of an enolate intermediate at some point in the reaction sequence. A pK_a-based argument supports Route 1, as the pK_a of a β-ketoacyl-ACP intermediate of Route 1 is ca. 2 units lower than the analogous malonyl-ACP of Route 2 and therefore more easily deprotonated (Evans, 2005). In further support of proposed Route 1, enolate attack on the upstream acyl-KS should occur more readily with the less sterically hindered acetyl enolate, than the more sterically hindered isobutyryl enolate of Route 2. On the other hand, if methylation precedes condensation as in Route 2, the observed production of isobutyryl-ACP in the yersiniabactin PKS (*vide infra*) is rationalized by a similar argument that malonyl-ACP is significantly easier to deprotonate than acetyl-ACP (approximate 10 unit difference in pK_a) and decarboxylation of dimethylmalonyl-CoA would be facilitated by the presence of electron donating methyl groups.

There is precedent for the involvement of dimethylmalonyl moieties in biochemical reactions. When provided *in vitro* to carboxymethylproline synthase (CarB), a member of the crotonase superfamily, a dimethylmalonyl moiety has been shown to be a source of a nucleophilic enolate for C-C bond forming reactions, despite its steric bulk (Batchelar et

al., 2008). Dimethylmalonyl groups are also naturally added to sugar residues in cervimycin biosynthesis (Bretschneider et al., 2012), and have been proposed as a source of isobutyryl groups in polyketide synthesis, but the latter has not been conclusively demonstrated (Young et al., 2013).

Examples of polyketides containing *gem*-dimethyl groups include the potential anticancer agents epothilone, pederin and bryostatin as well as the siderophore yersiniabactin (Scheme 1 A) (Anand et al., 2010). Currently, biochemical characterization of gem-dimethyl producing PKSs is limited to the yersiniabactin PKS. Using an *in vitro* system, Miller and coworkers (Miller et al., 2002) showed that the yersiniabactin PKS used malonyl-CoA and two *S*-adenosyl methionine (SAM) molecules to produce its *gem*-dimethyl containing extended product, 3-(hydroxyphenylthiazolinylthiazolinyl) [HPTT]-β-keto-2,2-dimethyl-ACP. They also established that the yersiniabactin AT domain has a 500 fold kinetic preference for malonyl-CoA over methylmalonyl-CoA. Additionally, Mazur and coworkers (Mazur et al., 2003) observed that the condensation reaction was dependent on SAM and that isobutyryl-ACP was a major side product of the reaction - approximately 75% of the ACP domains of the PKS were occupied by an isobutyryl moiety following the reaction. The observed dependence of the condensation reaction on SAM (Mazur et al., 2003) could be explained by the MT domain generating dimethylmalonyl-ACP which participates in the decarboxylative homologation reaction. Epothilone PKS module 8 (EpoM8) forms a similar *gem*-dimethyl group based on the final structure of epothilone, but this module has not been previously studied biochemically. The AT domain of EpoM8 is bioinformatically predicted to prefer methylmalonyl-CoA over malonyl-CoA (Julien et al., 2000). This was verified biochemically in a competition assay, though some loading of malonyl-CoA loading was observed in the absence of methylmalonyl-CoA (**Figure 8.3.1**). Thus, the MT domain in the module is predicted to naturally carry out only a single C-methylation. Despite these initial studies the exact mechanism and biosynthetic timing that PKSs employ to produce *gem*-dimethyl groups remains unclear.

Scheme 4.1. A) Representative *gem*-dimethyl containing compounds. B) Potential mechanistic routes for methylation and condensation in *gem*-dimethyl producing polyketide synthase domains. Route 1: KS catalyzed condensation precedes methylation. Route 2: Methylation precedes KS condensation. Acetyl-ACP and isobutyryl-ACP represent potential side products of unproductive decarboxylation (shown in gray). (AT) Acyltransferase; (KS) ketosynthase; (MT) methyltransferase. R represents the acyl chain transferred from the previous module.

4.3. Results and Discussion

To test if Route 2 is used, we performed *in vitro* chain elongation reactions with dimethylmalonyl-ACP, generated through directed acylation of the apo-ACP domain using two *gem*-dimethyl producing PKS modules: yersiniabactin and EpoM8. In addition, we sought to understand if the KS domains of these PKSs demonstrated a preference for the methylation state of the ACP-bound substrate, as reported by Mazur and coworkers (Mazur et al., 2003). Determination of these parameters would provide a clearer picture of which path(s) are operative in the production of this unusual group. Accordingly, we characterized these *gem*-dimethyl producing PKS clusters using a well-established tandem mass spectrometry method (Dorrestein et al., 2006) that can detect intermediates covalently attached to ACP domains in a single-turnover *in vitro* assay. The yersiniabactin and EpoM8 PKSs were expressed in *Escherichia coli* strains BLR and BAP1 (Pfeifer et al., 2001), respectively. These proteins were purified using Ni-NTA chromatography. The purified yersiniabactin PKS was approximately 70% holo-form, while the EpoM8 was only the holo-form . Since both malonyl-CoA and SAM are present in the expression strains, we analyzed the acylation state of each PKS immediately after purification to determine if acylation or methylation occurred during expression. Examination of the freshly purified yersiniabactin PKS and EpoM8 ACPs

revealed isobutyryl-ACP (a decarboxylation product of dimethylmalonyl-ACP) on EpoM8 (**Figure 4.1** A) and both dimethylmalonyl-ACP and isobutyryl-ACP appended to the yersiniabactin PKS (**Figure 4.1** B). This shows that at least the first half of Route 2 is viable for the yersiniabactin PKS. Interestingly, these results suggest that EpoM8 MT is able to methylate malonyl-ACP loaded during *in vivo* expression in *E. coli*, even though the preferred substrate for the EpoM8 AT is methylmalonyl-ACP (which is not natively present in *E. coli*). Furthermore, *in vitro* studies of the yersiniabactin PKS revealed that acetyl-ACP was not methylated in the presence of SAM (**Figure 8.3.2**), contrary to previous proposals which suggested the methylation of acetyl-ACP accounted for the production of isobutyryl-ACP (Mazur et al., 2003). These facts strongly suggest that isobutyryl-ACP is derived from dimethylmalonyl-ACP and implies that Route 2 contributes in the overall formation of a *gem*-dimethyl group.
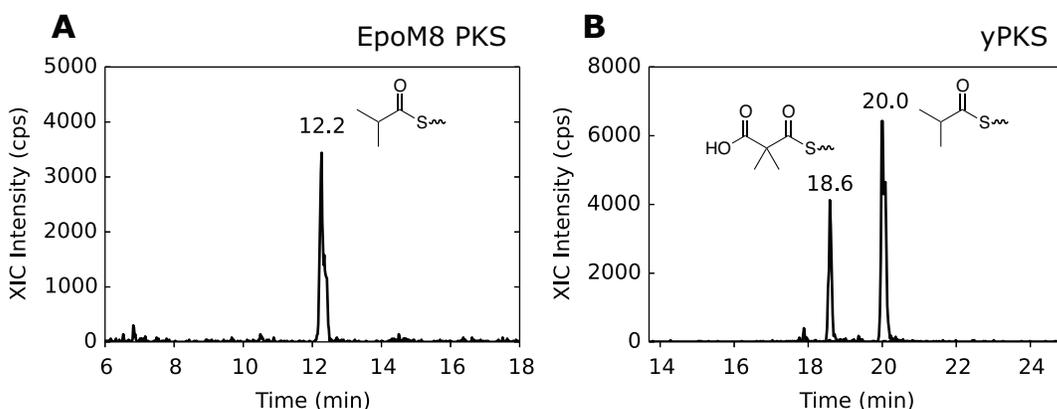
Figure 4.1. Analysis of the acylation state of epothilone module 8 PKS (EpoM8 PKS) and yersiniabactin PKS post Ni-NTA purification. Retention times are shown on top of peaks. A) EpoM8 isobutyryl-ACP chromatogram. B) Yersiniabactin PKS dimethylmalonyl-ACP chromatogram and isobutyryl-ACP chromatogram. All XIC m/z transitions monitored in this study are summarized in the supplementary information in Appendix 8.3. (XIC) Extracted ion chromatogram; (cps) counts per second.

To unambiguously determine if the dimethylmalonyl moiety loaded on the yersiniabactin and EpoM8 PKSs can be utilized by the KS domain in an *in vitro* extension reaction, we expressed apo-PKS (expressed in *E. coli* strain BLR, this resulted in approximately 100% apo form of the EpoM8). We then synthesized authentic dimethylmalonyl-CoA (Batchelar et al., 2008; Bretschneider et al., 2012) and loaded the ACP domains with the dimethylmalonyl moiety using the phosphopantetheinyl transferase Sfp (Quadri et al., 1998). Loading of dimethylmalonyl-ACP was verified by LC-MS/MS (**Figure 8.3.3**). The acyl group to be extended was provided to the appropriate KS domain by either a diketide-*S*-*N*-acetyl-cysteamine substrate (Sharma and Boddy, 2007), the acyl group of which mimics the structure of the native intermediate in the case of EpoM8, or by reconstitution of the upstream non ribosomal peptide synthetase (NRPS) enzymes in the case of the yersiniabactin PKS. As shown in **Figure 4.2**, both EpoM8 and yersiniabactin extended dimethylmalonyl-ACP in the absence of SAM to produce the correct, fully extended and methylated product with retention times matching those observed during fully reconstituted *in vitro* reactions (**Figure 8.3.4**). Additionally, the side product isobutyryl-ACP was produced from dimethylmalonyl-ACP by both enzymes as shown in

Figure 8.3.5. A large amount of isobutyryl-ACP was produced by both enzymes relative to the extended methylated product suggesting that Sfp loading experiments may not fully capture the precisely orchestrated catalytic cycle of these enzymes. The formation of acyl-KS also seems to have activated the KS for decarboxylation of dimethylmalonyl-ACP to isobutyryl-ACP (**Figure 8.3.5**), potentially indicating a conformational shift in the enzyme upon KS acylation.
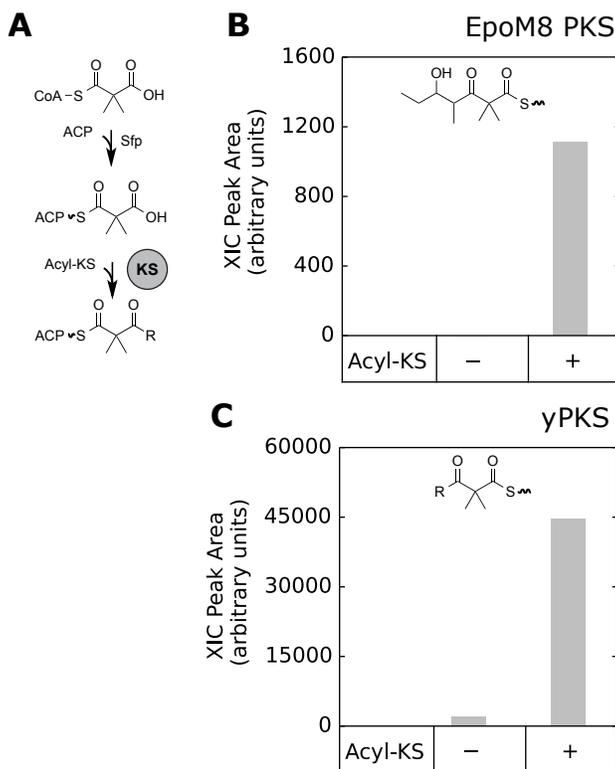
Figure 4.2. Dimethylmalonyl-ACP is extended by the EpoM8 and yersiniabactin PKS. A) Experiment overview: dimethylmalonyl-ACP was formed using Sfp, dimethylmalonyl-CoA and apo-PKS. Then substrates to form acyl-KS were added to generate extended products (experimental details in the Supplementary information in Appendix 8.3). Extracted ion chromatogram (XIC) peak areas for dimethylmalonyl-ACP extension experiments for B) EpoM8 and C) yersiniabactin PKS. For the yersiniabactin case, R represents the acyl chain transferred from the previous NRPS module: a HPTT moiety. All XIC *m/z* transitions monitored in this study are summarized in **Tables 8.3.1** and **8.3.2**.

To determine if Route 1 is also operative for each PKS, we attempted to form the β-keto unmethylated (yersiniabactin PKS) or monomethylated (EpoM8 PKS) products by incubating malonyl-CoA with acyl-KS while omitting SAM and subsequently adding SAM (**Figure 4.3** A). We then examined products before and after SAM addition using LC-MS/MS. Contrary to the results of Mazur et al (Mazur et al., 2003), the yersiniabactin KS domain was found to use malonyl-ACP as a substrate for condensation in the absence of SAM, and the MT domain methylated the resulting β-keto acyl-ACP moiety after SAM was added (**Figure 4.3** B). Small amounts of the unmethylated and singly methylated yersiniabactin-derived PKS product were also observed during reactions with the native malonyl-CoA substrate in the presence of SAM (**Figure 8.3.6**). EpoM8 was unable to extend methylmalonyl-ACP or malonyl-ACP in the absence of SAM, but it was

able to form the dimethyated final product for both extender units in the presence of SAM (**Figures 4.4** and **8.3.7**). This suggests that EpoM8 methylates methylmalonyl-ACP and then uses dimethylmalonyl-ACP as the substrate in the condensation reaction. Additionally, EpoM8 did not form the singly methylated intermediate (i.e. an intermediate resulting solely from the extension of methylmalonyl-ACP with no action by the MT) when full reconstitution reactions were performed in the presence of methylmalonyl-CoA and SAM. Together, these findings demonstrate that, *in vitro*, the yersiniabactin PKS is able to employ both pathways in Scheme 1, while EpoM8 is restricted to using solely Route 2.
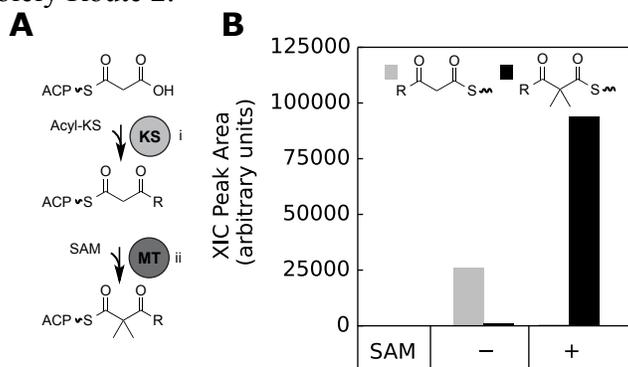
Figure 4.3. Characterization of Route 1 (KS catalyzed condensation precedes methylation) for the yersiniabactin PKS. A) Experiment overview: in part i SAM was omitted, while malonyl-CoA and acyl-KS were included to allow for unmethylated extended-ACP to form. In part ii SAM was added (experimental details in the Supplementary information in Appendix 8.3). B) Unmethylated extended-ACP extracted ion chromatogram (XIC) peak areas (gray) and methylated extended-ACP peak areas (black). R is a HPTT moiety. All XIC *m/z* values monitored in this study are summarized in **Tables 8.3.1** and **8.3.2**.
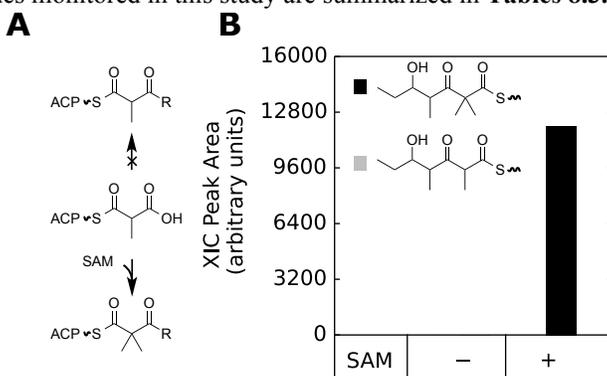
Figure 4.4. Characterization of Route 1 (KS catalyzed condensation precedes methylation) for the epothilone PKS. A) Experiment overview: using methylmalonyl-CoA as an extender unit SAM was omitted from the overall reaction (no product formed), or SAM was included, allowing the complete reaction to take place. B) Unmethylated extended-ACP extracted ion chromatogram (XIC) peak areas (gray) and methylated extended-ACP peak areas (black). All XIC *m/z* values monitored in this study are summarized in **Tables 8.3.1** and **8.3.2**.

We have clearly demonstrated that the yersiniabactin and epothilone PKSs are capable of forming dimethylmalonyl-ACP and isobutyryl-ACP, its decarboxylated derivative, and, contrary to the traditional understanding of methylation in type I PKSs, that both enzymes can use dimethylmalonyl-ACP as an extender unit in acyl chain elongation. For EpoM8, formation of dimethylmalonyl-ACP via Route 2 appeared to be the only means

to form a *gem*-dimethyl group (at least *in vitro*), while the yersiniabactin PKS could methylate pre- or post-condensation. The increased decarboxylation of dimethylmalonyl-ACP we observed upon acylation of the KS suggests that conformational changes in the enzyme are important during the catalytic cycle of these *gem*-dimethyl producing PKSs, as recently demonstrated for pikromycin module 5 using cryo electron microscopy (Whicher et al., 2014). Future studies will employ cryo electron microscopy to examine the spatial relationship between MT with the ACP during the catalytic cycle to determine the partition between Routes 1 and 2 for the yersiniabactin PKS. Based on our results, we would suggest that PKS engineering strategies to incorporate this group using combinatorial biosynthesis should swap whole modules, instead of swapping individual MT domains into modules unaccustomed to a particular order of reactions. Ultimately, improved understanding of this remarkable reaction will allow easier integration of this monomer unit in novel, useful PKS-based compounds.

5. Chapter 5 – Incorporation of gem-Dimethyl Groups into Engineered Polyketide Products

5.1. Introduction

Polyketide synthases (PKSs) are modular proteins capable of performing defined chemical transformations on a growing acyl chain. While much of the chemistry such as the specificity of acyl transfer and the level of reduction performed by a polyketide synthase module have been modified through engineering efforts, there are several chemistries that have not been incorporated into engineered polyketide products such as *gem*-dimethylation. *Gem*-dimethyl groups are useful for several reasons: 1) the increased branching increases the octane rating of aliphatic hydrocarbons, which could be useful for biofuels applications; 2) they act to reduce the rate of hydrolysis/metabolic degradation through steric hinderance (Drossman et al., 1988); and 3) the addition of steric bulk forces molecules into certain conformations (commonly known as the Thorpe-Ingold effect). The conformational effect can increase the rates of certain reactions, sometimes by multiple orders of magnitude (Levine and Raines, 2012) (**Figure 5.1**). Gem-dimethylation is known to occur in several polyketide products and is a useful moiety to incorporate into engineered product via combinatorial biosynthesis, as regiospecific C-methylation via semisynthesis of polyketide backbones would be challenging. Additionally, ~10% of approved drugs have at least 1 gem-dimethyl group, (Burkhard et al., 2010) suggesting that biocatalytic routes for methylation of PKS products would find pharmaceutical applications.
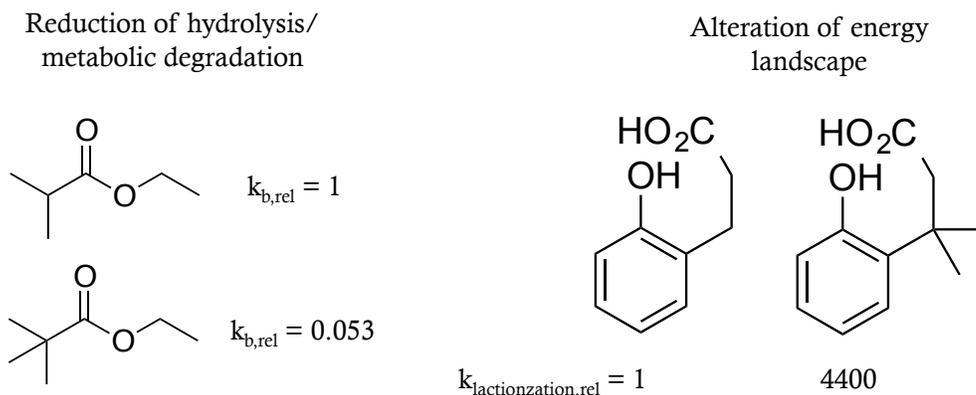
Figure 5.1. Gem-dimethyl group mechanism of action: gem-dimethyl groups can act by blocking the rate of hydrolysis through steric hindrance (Drossman et al., 1988), or by forcing molecules into reactive/stiffer conformations (Levine and Raines, 2012).

This project seeks to incorporate this group in several engineered compounds (**Figure 5.2**), ranging in complexity from small, branched acids, to large antibiotic analogs. We have made progress towards an isooctane precursor (biofuel) which upon catalytic reduction would have a high octane rating, an advanced intermediate in the synthesis of the drug epothilone (Menzella et al., 2010), and a gem-dimethylated form of the

antibiotic erythromycin which would be more resistant to hydrolysis from erythromycin esterases (Barthelemy et al., 1984) (**Figure 5.3**).
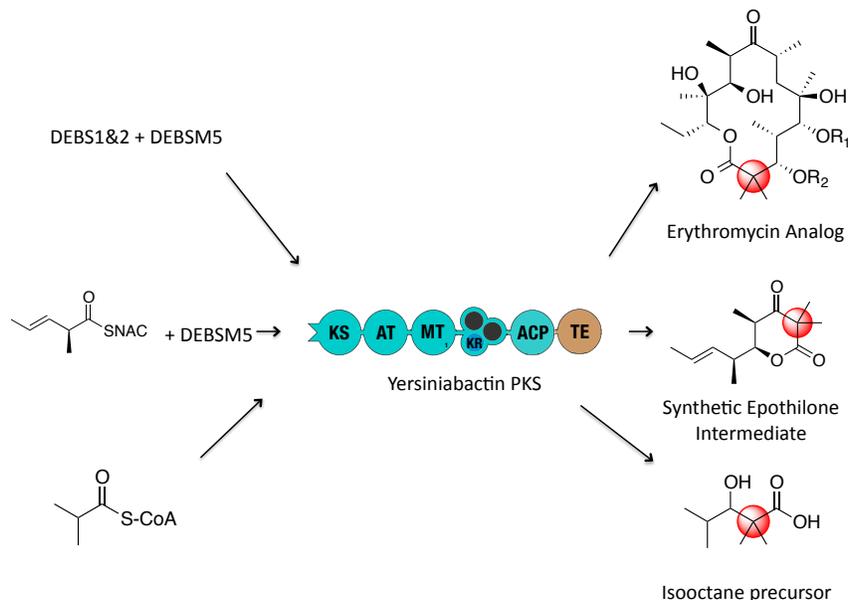
Figure 5.2. Target molecules in this study: a gem-diemthylated form of erythromycin that would be more resistant to erythromyin esterases (Barthelemy et al., 1984), an advanced intermediate in the synthesis of the drug epothilone (Menzella et al., 2010), and a compound that could be reduced to isooctane for use in gasoline. Proteins from the erythromycin cluster would also be needed to produce an erythromyin analog: the first two polykeitde synthase polypeptides (DEBS1&2) as well as the 5th module of the erythromycin cluster (DEBSM5) would provide an acyl intermediate to an engineered yersiniabactin protein. This engineered protein would then add two more carbons to the backbone and a gem-dimethyl group and cyclyze the product. Additional glycosylations and hydroxylations would also need to take place to form the complete erythromycin analog. A single module, such as DEBSM5 loaded using a SNAC substrate could provide the acyl intermediate for an engineered yersiniabactin protein to form the synthetic epothilone intermediate shown. A suitable loading didomain, such as the lipomycin domain, could provide an isobutyryl moiety to the yersiniabactin PKS, which could then be processed to the short, highly branched acid shown.
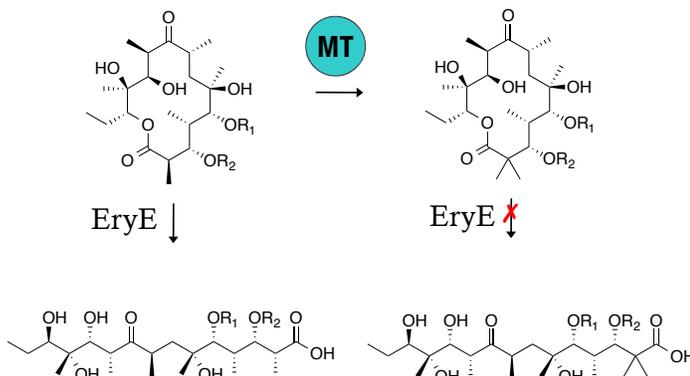
Figure 5.3. Potential reduction of erythromycin hydrolysis from erythromycin esterase by the addition of a methyl group via a methyltransferase domain (MT). Fluorine has also been substituted at this position, suggesting modifications at the alpha position are possible without greatly affecting the antibiotic activity (Liang et al., 2005) .

36

We have chosen to utilize the yersiniabactin PKS from the yersiniabactin hybrid nonribosomal peptide synthase (NRPS) polyketide synthase gene cluster (**Figure 5.4**) as a source of a gem-dimethyl group for several reasons. First, it produces a gem-dimethyl group from malonyl-CoA, which is available in the native metabolism of genetically tractable hosts like *E. coli* and industrially relevant hosts like yeast. Other gem-dimethyl producing PKSs utilize methyl-malonyl-CoA which is not present in *E. coli* or yeast metabolism. Second, the yersiniabactin gene cluster is found in some pathogenic strains of *E. coli*. This suggests that there may not be codon and folding issues in adapting this PKS to a genetically tractable, industrially-relevant host. Finally, activity of this cluster has already been shown *in vitro* with enzymes purified from *E. coli* (Miller et al., 2002), suggesting a higher likelihood of success. The dimethylation yersiniabactin synthase performs is quite unusual biologically and chemically difficult; thus exploring it is both scientifically interesting while leading towards useful compounds.
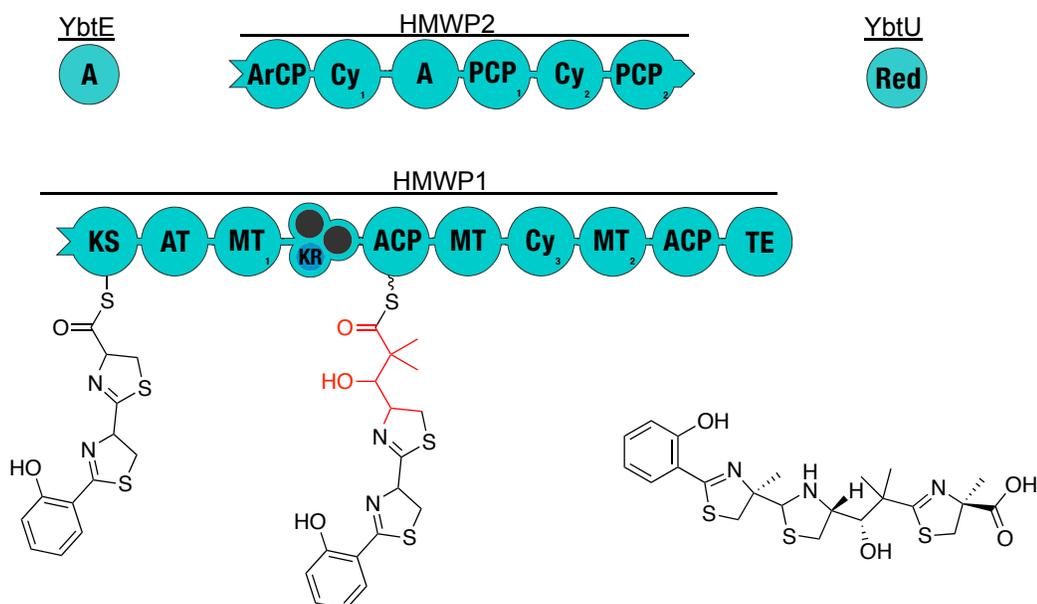
Figure 5.4. The molecule yersiniabactin produced by the yersiniabactin nonribosomal peptide synthase (NRPS) / polyketide synthase (PKS) hybrid gene cluster. The domain organization of the yersiniabactin gene cluster and acyl intermediates are illustrated. The portion of the product elaborated by the PKS is illustrated in red. The full yersiniabactin molecule is elaborated by 3 cysteine molecules, 3 S-adenosyl methionine molecules (SAM), 1 malonyl-CoA molecule and a salicylic acid molecule. Abbreviations: ACP, acyl carrier protein domain; PCP, peptidyl carrier protein domain; ArCP, aryl carrier protein domain; CY, cyclization domain; A, cysteine-specific adenylation domain; AT, malonyl CoA-specific acyltransferase domain; KR, ketoreductase domain; KS, ketosynthase domain; MT, methyltransferase domain.

5.2. Results and Discussion

To determine if production of the isooctane precursor in **Figure 5.2** was feasible, we performed a single turnover assay *in vitro*. A truncated protein containing the peptidyl carrier protein (PCP, construct pSP5), which naturally interacts with the KS of yersiniabactin was charged *in vitro* using Sfp (a broad specificity phosphopantetheinyl transferase) and isobutyryl-CoA, yielding isobutyryl-PCP. This truncated protein has been used successfully in the literature in the past as an acyl group donor for a similar experiment (Miller et al., 2002). Isobutyryl-PCP was then incubated with the yersiniabactin PKS (construct pSP8) fragment and the requisite substrates for extension (malonyl-CoA), dimethylation (SAM) and reduction (NADPH). The *in vitro* reaction was then digested with trypsin to yield loaded peptides that could be detected using LC-MS/MS.

The LC-MS/MS trace for the dimethylated 2,2,4-trimethyl-3-oxopentanoyl-ACP product is shown in **Figure 5.5**. These results show that the yersiniabactin synthase is capable of producing the isooctane carbon backbone; extension and dimethylation occur on an unnatural starter product.

To release the product from the ACP, a thioesterase needs to be fused with the yersiniabactin PKS to release the alpha-dimethyl product. The initial choice for thioesterase was the erythromycin TE as it has been successfully used with engineered type I PKSs in the past (Guo et al., 2010; Yuzawa et al., 2013). However, our results indicate that the erythromycin TE appended to the yersiniabactin product does not hydrolyze the alpha-dimethyl product (Construct pSP2). The alpha-dimethyl product was still observed attached to the ACP for constructs including the thioesterase (**Figure 5.6**). Alpha-dimethyl hydrolysis could be slower due to steric hindrance as has been reported for the hydrolysis of oxygen esters (Drossman et al., 1988). A TE specific to an alpha-dimethyl thioester has not been shown to exist in nature, and thus it is possible the TE specificity does not include alpha-dimethyl substrates.
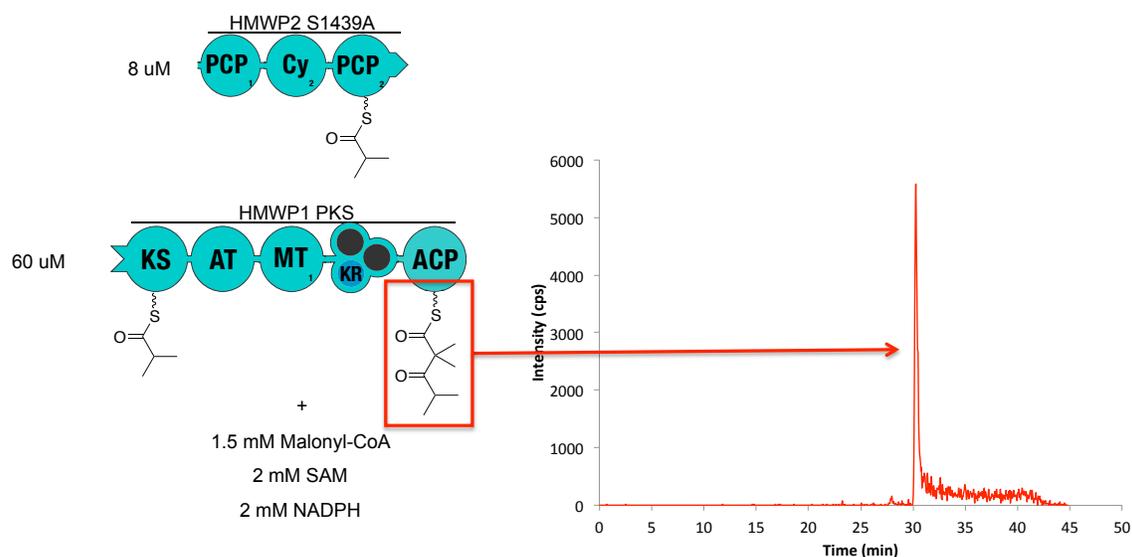
Figure 5.5. Initial experiment showing that the yersiniabactin PKS could accept isobutyryl-PCP as a substrate and extend and dimethylated it to an isooctane precursor ACP.
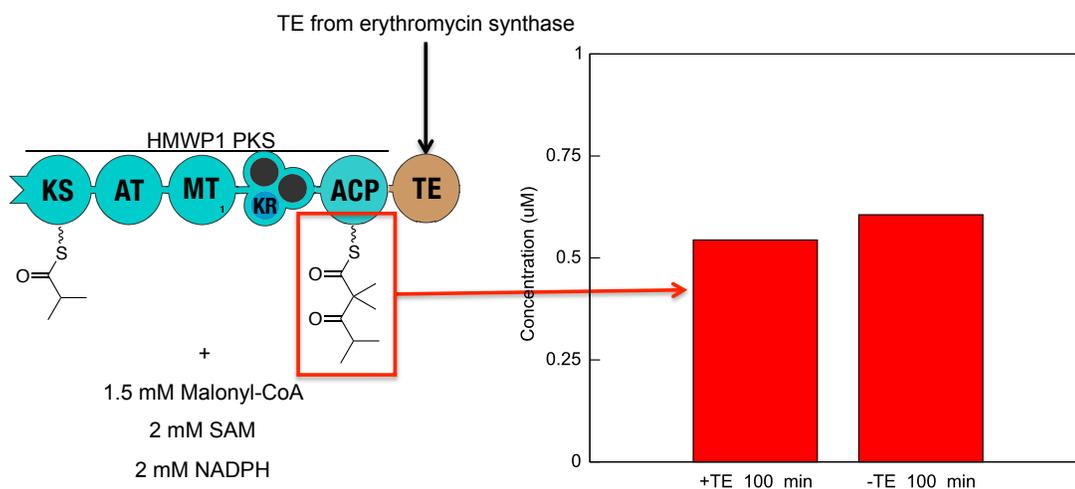
Figure 5.6. ACP loading in the presence and absence of a thioesterase. The thioesterase from the erythromycin cluster does not seem to be able to cleave the alpha dimethyl substrate.

As the TE from erythromycin appeared to be the rate-limiting step, we next sought to further investigate alterative TEs. There are a limited number of thioesterases from PKS clusters characterized in the literature, and fewer with crystal structures. One promising possibility was the TE from the pikromycin cluster, which shows a larger active site when compared to the erythromycin TE (**Figure 5.7**), suggesting it might have more activity towards sterically bulky substrates such as an alpha gem-dimethylated compound.
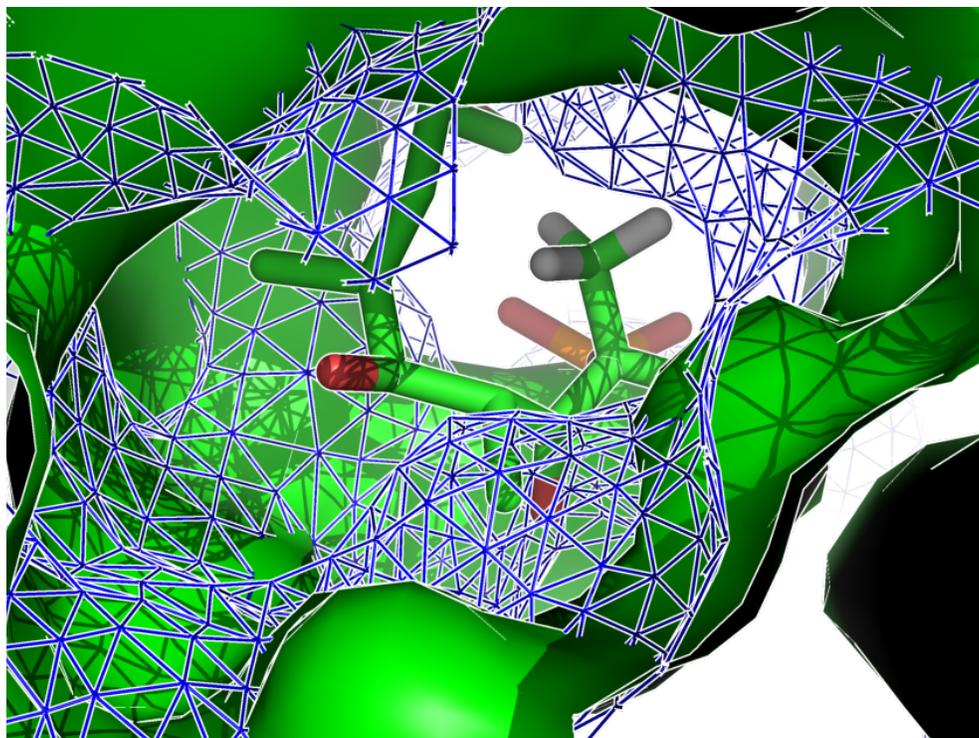
Figure 5.7. Comparison of pikromycin TE and erythromyinc TE active sites. Pik TE Shown as green surface (PDB: 2HFK (Akey et al., 2006)), DEBS TE shown as blue mesh (PDB: 1KEZ (Tsai et al., 2001))

Next, we characterized the k_{cat}/K_m of both the erythromycin TE (construct pSY48) and pikromycin TEs (construct pSP37) on the unbranched propionyl-S-N-acetyl-cysteamine (prop-SNAC) and the alpha-dimethyl pivaloyl-S-N-acetyl-cysteamine (piv-SNAC) (**Table 5.1**). On unbranched substrates these two TEs have only a ~3 fold difference in k_{cat}, however on highly branched substrates such as pivaloyl-SNAC, the pikromycin TE has an almost ~20-fold higher k_{cat} than the erythromycin TE. This suggested that the Pik TE should be used for future constructs producing alpha-dimethyl compounds.

We next performed saturation mutagenesis in which the 96 amino acids nearest the active site were mutated individually using NNK primers and Kunkel mutagenesis. From this library, 16 plates were screened for increased activity on Piv-SNAC in lysate using an Ellman's reagent based assay. Ellman's reagent produces a yellow color from free thiols in a concentration dependent manner, meaning hydrolysis activity can be coupled to an increase in absorbance. This screening yielded two variants which greatly increased the yellow color in the Ellman's assay: A218T and G150A (**Fig 5.8**). The G150A mutation is immediately next to the active site serine of the thioesterase. We obtained k_{cat}/K_m values for these mutants (**Table 5.1**), and found that they had modest gains in k_{cat} as well as some reduction in K_m on piv-SNAC. However, we noticed that both had increased in k_{cat} for prop-SNAC, suggesting that our screen may have enriched for better hydrolases.

40

Table 5.1. k_{cat}/K_m values for thioesterase variants
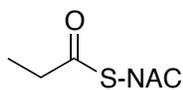
prop-SNAC
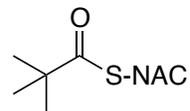
piv-SNAC

Enzyme	k_{cat} (1/min)	K_m (mM)	k_{cat}/K_m	k_{cat} (1/min)	K_m (mM)	k_{cat}/K_m
DEBS TE	2.1	35.2	0.059	0.14	50.9	0.003
Pik TE, WT	6.7	7.9	0.85	2.5	15.7	0.16
Pik TE, A218T	8.8	6.6	1.34	2.9	7.4	0.39
Pik TE, G150A	11.9	7.9	1.51	2.7	14.6	0.19
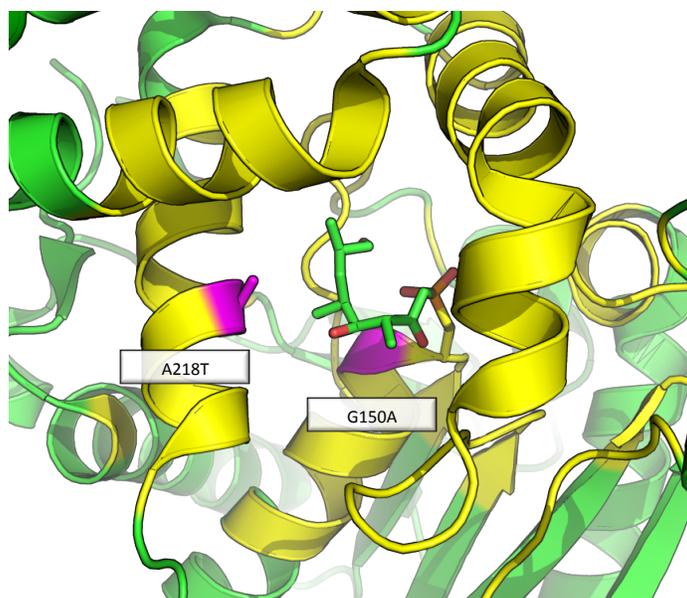
Figure 5.8. Location of mutants found via Ellman's reagent based screening the 96 amino acids mutated where the library was targeted (highlighted in yellow). (PDB: 2HFK (Akey et al., 2006)),

Thioesterases from most polyketide clusters should be more aptly named cyclases, as they generally regiospecifically utilize a hydroxyl in the polyketide chain as a nucleophile to cyclize the chain to form a lactone ring. Chemically, this is a fairly challenging reaction. As our screen seemed to enrich for hydrolase activity, we hypothesized that this may have broken the native cyclase activity. In collaboration with David Sherman's lab at the University of Michigan, we assayed the abilities of these variants to form the two native products of the Pikromycin PKS: 10-deoxymethynolide (12-membered ring) and narbonolide (14-membered ring). In this assay, the TE is appended to either PikAIII

(Module 5) or PikAIV (Module 6) of the Pikromycin cluster and a thiophenol thioester mimicking the upstream intermediate is used to start the biosynthesis. For the production of 10-deoxymethynolide, only the PikAIII+TE and the thiophenol are used, whereas native PikAIII, the PikAIV+TE construct and the thiophenol are used for the production of narbonolide. Products are then monitored by HPLC (Hansen et al., 2013).

Production of narbonolide and 10-deoxymethynolide for this assay are shown in **Figure 5.9** A and B, respectively. Interestingly, the G150A mutation seems to disfavor the formation of 10-deoxymethynolide (**Figure 5.9** B), while not affecting the production of narbonolide (**Figure 5.9** A). The pikromycin cluster is unique in its ability to produce two products and is also unique in the glycine next to the active site serine—other characterized TEs have an alanine at this position. It is intriguing that reverting this glycine to the more conserved alanine reduces the ability of the TE to produce two products, perhaps by limiting the flexibly of the serine next to it. Future crystallographic studies will further investigate the mechanism of this phenomenon.
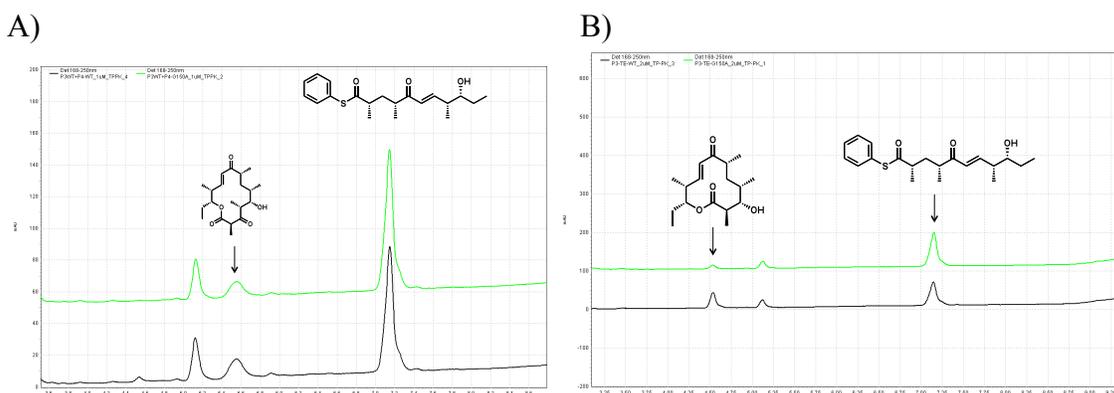
Figure 5.9. Production of narbonolide (A) and 10-dml (B) from G150A pikTE variant (green) and wildtype (black) protein. Assays similar to those performed previously (Hansen et al., 2013), data from Aaron Koch of the Sherman Lab at the University of Michigan.

To mediate the acyl transfer between DEBS5 and the yersiniabactin PKS, as required to produce both a gem-dimethylated form of erythromycin and a epothilone synthetic intermediate (**Figure 5.2**), we constructed variants with cognate linker domains on the C terminus of DEBS M5 and on the N terminus of the yersiniabactin PKS. The cognate linker domains utilized have been previously used in the literature and were from the phoslactomycin cluster (Yan et al., 2009) the curacin cluster (Whicher et al., 2013) the erythromycin cluster (Weissman, 2006), as well as the native linker from the yersinaibactin cluster (Miller et al., 2002) (DEBS M5 constructs: pBS31-36 and yersiniabactin constructs: pBS41-45). Constructs were designed using j5 (Hillson et al., 2011), and assembled using Gibson.

After all constructs were cloned, they were overexpressed in *E. coli* BAP1, lysed using sonication in 10% glycerol, 1 mM TCEP, 100 mM phosphate buffer pH 7.2 and centrifuged. All constructs expressed solubly except for BS43 (phoslactomycin C-terminal linker) and BS44 (erythromycin M3 C-terminal linker). Additionally, the upstream ACP4 was cloned (Construct BS48), overexpressed in *E. coli* BLR and loaded

with isobutyryl-CoA using the phsophopantethenyl transferase sfp. Cognate clarified supernatants were combined with ACP4 in lysate; as well as methyl malonyl-CoA and NADPH (substrates for DEBSM5); and malonyl-CoA, NADPH and SAM (substrates for the yersinaibactin PKS). After a 30 min reaction at 30°C, lysates were trypsinized, cleaned up using a C18 tip and injected onto an Agilent 6460 LC-MS. Results show the correct intermediate attached to DEBSM5 for several variants with the highest activity observed from construct BS35, which includes the linker from erythromycin module 4 (data shown in **Figure 5.10**). Unfortunately, no intermediates were observed on any of the engineered yersiniabactin constructs (experiment outline and results in **Figure 5.10**).
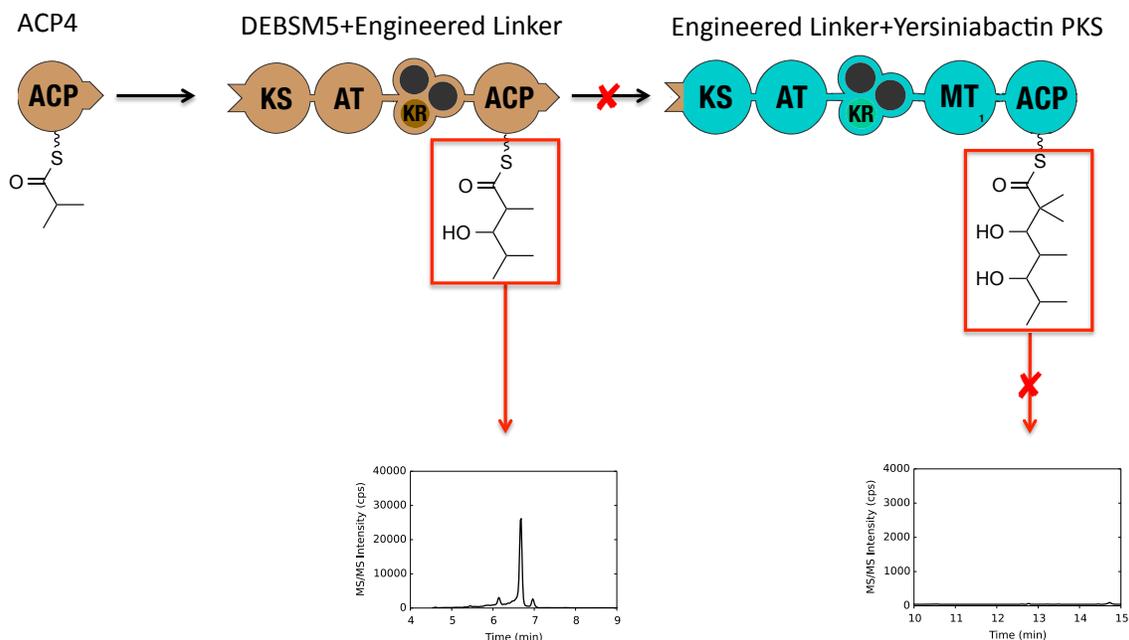
Figure 5.10. Engineering of DEBS M5 and the yersiniabactin PKS to produce engineered gem-dimethylated triketides. ACP4 was primed with isobutyryl-CoA to load DEBSM5 variants with engineered C terminal linker domains. These engineered DEBS5 variants were incubated with yersiniabactin variants with cognate N terminal linker domains. LC-MS/MS showed that the DEBSM5 variants were capable forming intermediates (data shown for construct BS35 with the linker from DEBS module 4), but these intermediates were not extended by the engineered yersiniabactin proteins.

These experiments introduced both an unnatural substrate and an unnatural protein-protein interaction into the yersiniabactin system. Either of these could be potential failure points. To determine which was the case, instead of using the unnatural acyl-ACP of engineered DEBS M5 proteins, we used the natural PCP upstream from the yersiniabactin protein, loaded with different acyl groups (similar to the experiments described in **Figure 5.5**). We formed both isobutyryl-PCP and 3-hydroxy butyryl-PCP, to determine if the introduction of the hydroxyl group prevented extension by the yersiniabactin PKS. As shown in **Figure 5.11**, the yersiniabactin PKS is capable of extending an isobutyryl acyl group, but the introduction of the hydroxyl group seems to prevent extension. Thus this suggests that the KS specificity prevents formation of the triketide intermediates for the experiments performed in **Figure 5.10**.
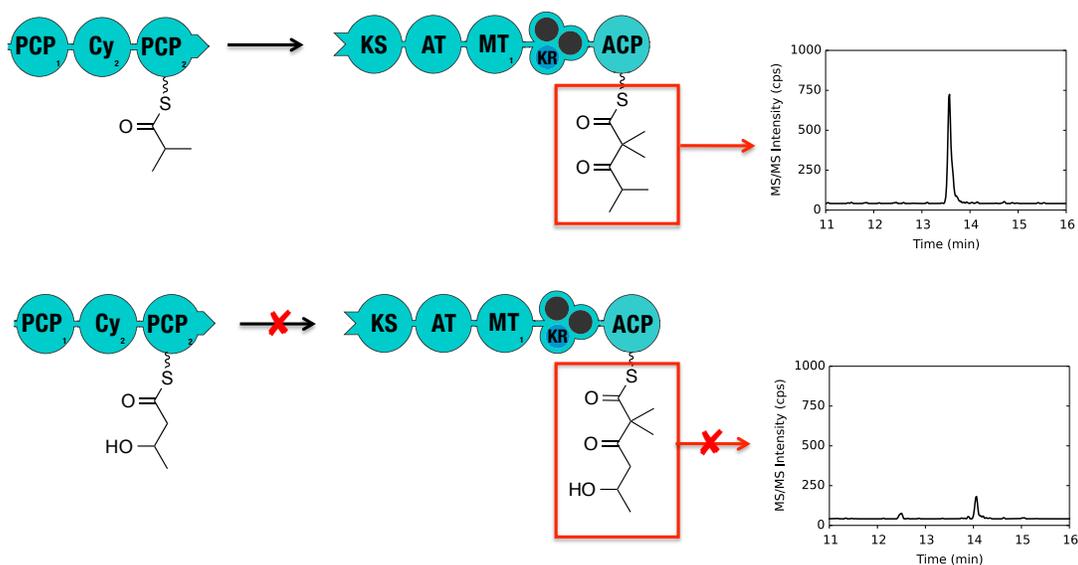
Figure 5.11. Isobutyryl-PCP vs 3-hydroxy-butyryl-PCP extension. The yersiniabactin PKS can extend isobutyryl-PCP, but it was unable to extend 3-hydroxy-butyryl-PCP. These results suggests that the KS domain discriminates against hydroxyl containing acyl groups.

5.3. Conclusions and Future Work

It is not surprising that the individual domains within PKS assembly lines show inherent specificity that must be modulated in order to produce new compounds. Here, we have identified two bottlenecks for the production of engineered gem-dimethylated polyketide compounds, both at the thioesterase level and the ketosynthase level. After identification of the thioesterase bottleneck, the pikromycin thioesterase was shown to be a good candidate to release gem-dimethylated ACPs. The yersiniabactin ketosynthase domain was shown to discriminate against hydroxyl groups, which are part of the incoming acyl chain to produce a gem-dimethylated form of erythromycin and the desired epothilone synthetic intermediate (**Figure 5.2**). Future work will mutagenize the KS active site, similar to the work done for the thioesterase domain and then adapt the lysate based screening method used to generate the data in **Figure 5.10** to determine which KS variants are able to process hydroxyl-containing acyl intermediates.

6. Summary, Conclusions and Future Directions

The work presented in this dissertation explored several aspects of polyketide biochemistry in Chapters 3 and 4, examining both acyl transferase and methyl transferase domains in polyketide synthases in more detail. As part of the acyl transferase work, we observed that a histidine to alanine mutation (H640A) in the GHSxG motif of the malonyl-CoA specific yersiniabactin acyl transferase resulted in an approximately seven-fold higher hydrolysis rate over the wildtype enzyme, while retaining transacylation activity. This led us to conclude that this histidine is involved in stabilizing the malonyl-O-AT ester intermediate required for acyl transfer. To explore the mechanism of *gem*-dimethyl group formation, we characterized the epothilone module 8 and yersiniabactin PKSs. Contrary to the canonical understanding of PKSs, experiments showed that both PKSs are able to use dimethylmalonyl-ACP as an extender unit, showing that methylation could precede condensation. Interestingly, for epothilone module 8, dimethylmalonyl-ACP appeared to be the sole route to form a gem-dimethylated product while the yersiniabactin PKS could methylate pre- or post-ketosynthase condensation. Additionally, engineering work towards the production of useful gem-dimethylated small molecules was presented in Chapter 5. Bottlenecks in both the ketosynthase and thioesterase domains were identified and screening strategies to address both of these bottlenecks were developed.

Much of the work presented in this dissertation was facilitated by the use of mass spectrometry to examine acyl intermediates attached to polyketide synthase acyl carrier proteins. Not only has this been useful for the biochemical characterization in Chapters 3-4 and determining that the thioesterase and ketosynthase were the failure modes for constructs in Chapter 5, but this tool has also been useful for "debugging" other polyketide synthase efforts that I have been part of, either to produce diacids (Hagen et al., 2015) or in combinatorial library evaluation (Phelan et al., 2015). Throughout the course of this my work, I moved from examining acyl intermediates *in vitro* with purified proteins to reactions in lysates. This greatly improved the scalability of the technique and meant many more constructs could be debugged in parallel, as demonstrated at the end of Chapter 5. Going forward, this will allow for engineering more constructs in parallel and will hopefully allow access to even more chemical space using engineered PKSs. The PKS field has a longstanding unrealized goal of rational retrosynthesis. My work here demonstrates the utility of acyl intermediate mass spectrometry in debugging engineered synthases, and I hope my work will encourage the field to adopt this technique for PKS-based retrosynthesis efforts.

Nevertheless, we have observed that the kinetics of even the native PKS enzymes expressed and purified from *E. coli* are very slow ($\sim$0.01 min^{-1}, (Hagen et al., 2014)). As the acyl intermediate technique examines peptides, it is very sensitive and can detect even very low activities from PKSs, as peptides ionize very efficiently during electrospray ionization. At the same time, this allows us to hide behind our cleverness with assay development and the power of modern mass spectrometers. Making any sort of engineered polyketide synthase application economical with kinetics of $\sim$0.01 min^{-1} will be difficult, if not impossible. While the appeal of reaching new chemical space with

engineered polyketide synthases is very enticing, a priority of at least equal importance must be determination of the kinetics of these enzymes in their native host, and then development of a tractable host with can produce properly folded, kinetically-competent PKSs.

The native PKS kinetics can be estimated in a two-step process: first, absolute quantification of the concentration of polyketide proteins *in vivo* is determined with bottom-up proteomics and peptide standards. Second, the product productivity is determined from monitoring the concentration of a final polyketide molecule over a time course. With these two numbers, one can estimate the *in vivo* k_{cat} for these enzymes by dividing the product productivity by the polyketide protein concentration. The fact that industrial fermentations can reach very high titers of polyketide products in mutagenized native hosts (~100 g/L) suggests that the native kinetics must be much better than what we have observed so far *in vitro*.

Next, a broad range of hosts can be tested through the introduction of a small polyketide synthase, using the same technique as above to determine which host is able to produce the most kinetically competent synthase. These test hosts can range from highly genetically tractable, but phylogenetically distant from the native host, like *E. coli*, to other somewhat genetically tractable *Streptoymces* strains, like *Streptomyces venezuelae*. Something in between, like *Bacillus subtilis* (which has its own polyketide synthase clusters) would also be a good test host. Once k_{cat} values are estimated for several hosts and compared to the native kinetics determined as above, we will have a much better idea of how much phylogenic distance and cellular environment matters for producing kinetically competent polyketide synthases. We can then focus our efforts on developing genetic tools for an appropriate host.

We have made great strides in engineering PKSs to move from no product to detectable product with these acyl intermediate techniques, but we must next focus on moving titers from detectable to economical. It is my hope that by carefully studying the kinetics of these enzymes in their native hosts as compared to several heterologous hosts, we will be able to move engineered PKSs towards economical titers.

7. References

Aakvik, T., Degnes, K.F., Dahlsrud, R., Schmidt, F., Dam, R., Yu, L., Volker, U., Ellingsen, T.E., and Valla, S. (2009). A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species. FEMS microbiology letters *296*, 149-158.

Akey, D.L., Kittendorf, J.D., Giraldes, J.W., Fecik, R.A., Sherman, D.H., and Smith, J.L. (2006). Structural basis for macrolactonization by the pikromycin thioesterase. Nat Chem Biol *2*, 537-542.

Anand, S., Prasad, M.V.R., Yadav, G., Kumar, N., Shehara, J., Ansari, M.Z., and Mohanty, D. (2010). SBSPKS: structure based sequence analysis of polyketide synthases. Nucleic Acids Res *38*, W487-W496.

Baldomà, L., and Aguilar, J. (1987). Involvement of lactaldehyde dehydrogenase in several metabolic pathways of Escherichia coli K12. J Biol Chem *262*, 13991-13996.

Baltz, R.H. (2010). Streptomyces and Saccharopolyspora hosts for heterologous expression of secondary metabolite gene clusters. J Ind Microbiol *37*, 759-772.

Bar-Even, A., Noor, E., Flamholz, A., and Milo, R. (2013). Design and analysis of metabolic pathways supporting formatotrophic growth for electricity-dependent cultivation of microbes. Biochimica et biophysica acta *1827*, 1039-1047.

Barthelemy, P., Autissier, D., Gerbaud, G., and Courvalin, P. (1984). Enzymic hydrolysis of erythromycin by a strain of Escherichia coli. A new mechanism of resistance. The Journal of antibiotics *37*, 1692-1696.

Batchelar, E.T., Hamed, R.B., Ducho, C., Claridge, T.D.W., Edelmann, M.J., Kessler, B., and Schofield, C.J. (2008). Thioester Hydrolysis and C-C Bond Formation by Carboxymethylproline Synthase from the Crotonase Superfamily. Angew Chem Int Ed Engl *47*, 9322-9325.

Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths‐Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L.L.*, et al.* (2004). The Pfam protein families database. Nucleic Acids Res *32*, D138-D141.

Bennett, B.D., Kimball, E.H., Gao, M., Osterhout, R., Van Dien, S.J., and Rabinowitz, J.D. (2009). Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. Nat Chem Biol *5*, 593-599.

Betancor, L., Fernandez, M.J., Weissman, K.J., and Leadlay, P.F. (2008). Improved catalytic activity of a purified multienzyme from a modular polyketide synthase after

coexpression with Streptomyces chaperonins in Escherichia coli. Chembiochem *9*, 2962-2966.

Boddy, C. (2014). Bioinformatics tools for genome mining of polyketide and non-ribosomal peptides. J Ind Microbiol *41*, 443-450.

Breslow, R. (1959). On the mechanism of the formose reaction. Tetrahedron Lett *1*, 22-26.

Bretschneider, T., Zocher, G., Unger, M., Scherlach, K., Stehle, T., and Hertweck, C. (2012). A ketosynthase homolog uses malonyl units to form esters in cervimycin biosynthesis. Nat Chem Biol *8*, 154-161.

Bumpus, S.B., and Kelleher, N.L. (2008). Accessing natural product biosynthetic processes by mass spectrometry. Curr Opin Chem Biol *12*, 475-482.

Burkhard, J.A., Wuitschik, G., Rogers-Evans, M., Müller, K., and Carreira, E.M. (2010). Oxetanes as Versatile Elements in Drug Discovery and Synthesis. Angew Chem Int Ed Engl *49*, 9052-9067.

Cabantous, S., and Waldo, G.S. (2006). In vivo and in vitro protein solubility assays using split GFP. Nat Meth *3*, 845-854.

Callahan, B., Thattai, M., and Shraiman, B.I. (2009). Emergent gene order in a model of modular polyketide synthases. Proceedings of the National Academy of Sciences of the United States of America *106*, 19410-19415.

Cancilla, D.A., and Que Hee, S.S. (1992). O-(2,3,4,5,6-Pentafluorophenyl)methylhydroxylamine hydrochloride: a versatile reagent for the determination of carbonyl-containing compounds. J Chromatogr A *627*, 1-16.

Chakraborty, S., Nemeria, N., Yep, A., McLeish, M.J., Kenyon, G.L., and Jordan, F. (2008). Mechanism of Benzaldehyde Lyase Studied via Thiamin Diphosphate-Bound Intermediates and Kinetic Isotope Effects. Biochemistry *47*, 3800-3809.

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

Chang, Y.Y., Wang, A.Y., and Cronan, J.E. (1993). Molecular cloning, DNA sequencing, and biochemical analyses of Escherichia coli glyoxylate carboligase. An enzyme of the acetohydroxy acid synthase-pyruvate oxidase family. J Biol Chem *268*, 3911-3919.

Chen, A.Y., Cane, D.E., and Khosla, C. (2007). Structure-Based Dissociation of a Type I Polyketide Synthase Module. Chem Biol *14*, 784-792.

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

Cornvik, T., Dahlroth, S.-L., Magnusdottir, A., Herman, M.D., Knaust, R., Ekberg, M., and Nordlund, P. (2005). Colony filtration blot: a new screening method for soluble protein expression in Escherichia coli. Nat Meth *2*, 507-509.

Dorrestein, P.C., Bumpus, S.B., Calderone, C.T., Garneau-Tsodikova, S., Aron, Z.D., Straight, P.D., Kolter, R., Walsh, C.T., and Kelleher, N.L. (2006). Facile Detection of Acyl and Peptidyl Intermediates on Thiotemplate Carrier Domains via Phosphopantetheinyl Elimination Reactions during Tandem Mass Spectrometry. Biochemistry *45*, 12756-12766.

Dreier, J., Li, Q., and Khosla, C. (2001). Malonyl-CoA:ACP Transacylase from Streptomyces coelicolor Has Two Alternative Catalytically Active Nucleophiles. Biochemistry *40*, 12407-12411.

Drossman, H., Johnson, H., and Mill, T. (1988). Structure activity relationships for environmental processes 1: Hydrolysis of esters and carbamates. Chemosphere *17*, 1509-1530.

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

Dunn, B.J., and Khosla, C. (2013). Engineering the acyltransferase substrate specificity of assembly line polyketide synthases. J R Soc Interface *10*.

Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009). Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. PLoS One *4*, e5553.

Evans, D. (2005). "pKa's of CH bonds in Hydrocarbons and Carbonyl Compounds" http://evans.rc.fas.harvard.edu/pdf/evans_pKa_table.pdf.

Firnberg, E., and Ostermeier, M. (2012). PFunkel: Efficient, Expansive, User-Defined Mutagenesis. PLoS ONE *7*, e52031.

Fischbach, M.A., Lai, J.R., Roche, E.D., Walsh, C.T., and Liu, D.R. (2007). Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes. Proceedings of the National Academy of Sciences of the United States of America *104*, 11951-11956.

Fischbach, M.A., and Walsh, C.T. (2006). Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. Chem Rev *37*, 3468-3496.

Flamholz, A., Noor, E., Bar-Even, A., and Milo, R. (2012). eQuilibrator—the biochemical thermodynamics calculator. Nucleic Acids Res *40*, D770-D775.

Gehring, A.M., Mori, I., Perry, R.D., and Walsh, C.T. (1998). The Nonribosomal Peptide Synthetase HMWP2 Forms a Thiazoline Ring during Biogenesis of Yersiniabactin, an Iron-Chelating Virulence Factor of Yersinia pestis. Biochemistry *37*, 11637-11650.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Meth *6*, 343-345.

Gomez-Escribano, J.P., and Bibb, M.J. (2011). Engineering Streptomyces coelicolor for heterologous expression of secondary metabolite gene clusters. Microbial biotechnology *4*, 207-215.

González, B., and Vicuña, R. (1989). Benzaldehyde lyase, a novel thiamine PPi-requiring enzyme, from Pseudomonas fluorescens biovar I. J Bacteriol *171*, 2401-2405.

Greving, M., Cheng, X., Reindl, W., Bowen, B., Deng, K., Louie, K., Nyman, M., Cohen, J., Singh, A., Simmons, B.*, et al.* (2012). Acoustic deposition with NIMS as a high-throughput enzyme activity assay. Anal Bioanal Chem *403*, 707-711.

Guo, X., Liu, T., Valenzano, C.R., Deng, Z., and Cane, D.E. (2010). Mechanism and Stereospecificity of a Fully Saturating Polyketide Synthase Module: Nanchangmycin Synthase Module 2 and Its Dehydratase Domain. J Am Chem Soc *132*, 14694-14696.

Hagen, A., Poust, S., de Rond, T., Katz, L., Petzold, C.J., and Keasling, J.D. (2015). An engineered, substrate tolerant PKS extension module enables the production of adipic acid, *in preparation*.

Hagen, A., Poust, S., de Rond, T., Yuzawa, S., Katz, L., Adams, P.D., Petzold, C.J., and Keasling, J.D. (2014). In Vitro Analysis of Carboxyacyl Substrate Tolerance in the Loading and First Extension Modules of Borrelidin Polyketide Synthase. Biochemistry *53*, 5975-5977.

Hamedi, J., Malekzadeh, F., and Saghafi-nia, A.E. (2004). Enhancing of erythromycin production by Saccharopolyspora erythraea with common and uncommon oils. J Ind Microbiology and Biotechnology *31*, 447-456.

Hansen, D.A., Rath, C.M., Eisman, E.B., Narayan, A.R.H., 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 *135*, 11232-11238.

Hertweck, C. (2009). The Biosynthetic Logic of Polyketide Diversity. Angew Chem Int Ed Engl *48*, 4688-4716.

Hillson, N.J., Rosengarten, R.D., and Keasling, J.D. (2011). j5 DNA Assembly Design Automation Software. ACS Synth Biol *1*, 14-21.

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.

Iniesta, A.A., Garcia-Heras, F., Abellon-Ruiz, J., Gallego-Garcia, A., and Elias-Arnanz, M. (2012). Two systems for conditional gene expression in Myxococcus xanthus inducible by isopropyl-beta-D-thiogalactopyranoside or vanillate. Journal of bacteriology *194*, 5875-5885.

Janzen, E., Müller, M., Kolter-Jung, D., Kneen, M.M., McLeish, M.J., and Pohl, M. (2006). Characterization of benzaldehyde lyase from Pseudomonas fluorescens: A versatile enzyme for asymmetric C–C bond formation. Bioorg Chem *34*, 345-361.

Jenke-Kodama, H., Borner, T., and Dittmann, E. (2006). Natural Biocombinatorics in the Polyketide Synthase Genes of the Actinobacterium Streptomyces avermitilis. PLoS Comput Biol *2*, e132.

Jenke-Kodama, H., Sandmann, A., Müller, R., and Dittmann, E. (2005). Evolutionary Implications of Bacterial Polyketide Synthases. Mol Biol Evol *22*, 2027-2039.

Julien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L., and Khosla, C. (2000). Isolation and characterization of the epothilone biosynthetic gene cluster from Sorangium cellulosum. Gene *249*, 153-160.

Kakirde, K.S., Wild, J., Godiska, R., Mead, D.A., Wiggins, A.G., Goodman, R.M., Szybalski, W., and Liles, M.R. (2011). Gram negative shuttle BAC vector for heterologous expression of metagenomic libraries. Gene *475*, 57-62.

Keating, T.A., Miller, D.A., and Walsh, C.T. (2000). Expression, Purification, and Characterization of HMWP2, a 229 kDa, Six Domain Protein Subunit of Yersiniabactin Synthetase. Biochemistry *39*, 4729-4739.

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

Keatinge-Clay, A.T., Shelat, A.A., Savage, D.F., Tsai, S.-C., Miercke, L.J.W., O'Connell III, J.D., Khosla, C., and Stroud, R.M. (2003). Catalysis, Specificity, and ACP Docking Site of Streptomyces coelicolor Malonyl-CoA:ACP Transacylase. Structure *11*, 147-154.

Khosla, C. (2009). Structures and Mechanisms of Polyketide Synthases. J Org Chem *74*, 6416-6420.

Kim, B.S., Sherman, D.H., and Reynolds, K.A. (2004). An efficient method for creation and functional analysis of libraries of hybrid type I polyketide synthases. Protein Eng Des Sel *17*, 277-284.

Kobayashi, K., and Kawai, S. (1982). Enzymatic determination of hydrogen peroxide using gas chromatography. J Chromatogr A *245*, 339-345.

Komatsu, M., Uchiyama, T., Omura, S., Cane, D.E., and Ikeda, H. (2010). Genome-minimized Streptomyces host for the heterologous expression of secondary metabolism. Proceedings of the National Academy of Sciences of the United States of America *107*, 2646-2651.

LeBlanc, D.J., and Mortlock, R.P. (1971). Metabolism of d-Arabinose: a New Pathway in Escherichia coli. J Bacteriol *106*, 90-96.

Levine, M.N., and Raines, R.T. (2012). Trimethyl lock: a trigger for molecular release in chemistry, biology, and pharmacology. Chemical Science *3*, 2412-2420.

Liang, C.-H., Yao, S., Chiu, Y.-H., Leung, P.Y., Robert, N., Seddon, J., Sears, P., Hwang, C.-K., Ichikawa, Y., and Romero, A. (2005). Synthesis and biological activity of new 5-O-sugar modified ketolide and 2-fluoro-ketolide antibiotics. Bioorg Med Chem Lett *15*, 1307-1310.

Liew, C.W., Nilsson, M., Chen, M.W., Sun, H., Cornvik, T., Liang, Z.-X., and Lescar, J. (2012). Crystal Structure of the Acyltransferase Domain of the Iterative Polyketide Synthase in Enediyne Biosynthesis. J Biol Chem *287*, 23203-23215.

Linshiz, G., Stawski, N., Goyal, G., Bi, C., Poust, S., Sharma, M., Mutalik, V., Keasling, J.D., and Hillson, N.J. (2014). PR-PR: Cross-Platform Laboratory Automation System. ACS Synth Biol *3*, 515-524.

Linshiz, G., Stawski, N., Poust, S., Bi, C., Keasling, J.D., and Hillson, N.J. (2013). PaR-PaR Laboratory Automation platform. ACS Synth Biol *2*, 216-222.

MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., and MacCoss, M.J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics *26*, 966-968.

Martinez, A., Kolvek, S.J., Yip, C.L.T., Hopke, J., Brown, K.A., MacNeil, I.A., and Osburne, M.S. (2004). Genetically Modified Bacterial Strains and Novel Bacterial Artificial Chromosome Shuttle Vectors for Constructing Environmental Libraries and Detecting Heterologous Natural Products in Multiple Expression Hosts. Appl Environ Microbiol *70*, 2452-2463.

Matsumoto, T., Yamamoto, H., and Inoue, S. (1984). Selective formation of triose from formaldehyde catalyzed by thiazolium salt. J Am Chem Soc *106*, 4829-4832.

Mazur, M.T., Walsh, C.T., and Kelleher, N.L. (2003). Site-Specific Observation of Acyl Intermediate Processing in Thiotemplate Biosynthesis by Fourier Transform Mass Spectrometry: The Polyketide Module of Yersiniabactin Synthetase. Biochemistry *42*, 13393-13400.

McDaniel, R., Ebert-Khosla, S., Hopwood, D.A., and Khosla, C. (1995). Rational design of aromatic polyketide natural products by recombinant assembly of enzymatic subunits. Nature *375*, 549-554.

McDaniel, R., Kao, C.M., Hwang, S.J., and Khosla, C. (1997). Engineered intermodular and intramodular polyketide synthase fusions. Chem Biol *4*, 667-674.

McLoughlin, S.M., and Kelleher, N.L. (2005). Monitoring Multiple Active Sites on Thiotemplate Enzymes in Parallel: A Molecular Movie of Yersiniabactin Bioassembly. J Am Chem Soc *127*, 14984-14985.

Menzella, H.G., Carney, J.R., Li, Y., and Santi, D.V. (2010). Using Chemobiosynthesis and Synthetic Mini-Polyketide Synthases To Produce Pharmaceutical Intermediates in Escherichia coli. Appl Environ Microbiol *76*, 5221-5227.

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

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. Nature Biotechnol *23*, 1171-1176.

Miller, D.A., Luo, L., Hillson, N., Keating, T.A., and Walsh, C.T. (2002). Yersiniabactin Synthetase: A Four-Protein Assembly Line Producing the Nonribosomal Peptide/Polyketide Hybrid Siderophore of Yersinia pestis. Chem Biol *9*, 333-344.

Molnos, J., Gardiner, R., Dale, G.E., and Lange, R. (2003). A continuous coupled enzyme assay for bacterial malonyl–CoA:acyl carrier protein transacylase (FabD). Anal Biochem *319*, 171-176.

Müller, M., Sprenger, G.A., and Pohl, M. (2013). CC bond formation using ThDP-dependent lyases. Curr Opin Chem Biol *17*, 261-270.

Mutka, S.C., Carney, J.R., Liu, Y., and Kennedy, J. (2006). Heterologous Production of Epothilone C and D in Escherichia coli. Biochemistry *45*, 1321-1330.

Noor, E., Bar-Even, A., Flamholz, A., Reznik, E., Liebermeister, W., and Milo, R. (2014). Pathway Thermodynamics Highlights Kinetic Obstacles in Central Metabolism. PLoS Comput Biol *10*, e1003483.

Noor, E., Haraldsdóttir, H.S., Milo, R., and Fleming, R.M.T. (2013). Consistent Estimation of Gibbs Energy Using Component Contributions. PLoS Comput Biol *9*, e1003098.

Perlova, O., Gerth, K., Kuhlmann, S., Zhang, Y., and Muller, R. (2009). Novel expression hosts for complex secondary metabolite megasynthetases: Production of

myxochromide in the thermopilic isolate Corallococcus macrosporus GT-2. Microbial cell factories *8*, 1.

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

Phelan, R., Kelly, P., Poust, S., Katz, L., Petzold, C.J., and Keasling, J.D. (2015). Evaluation of Combinatorial PKS Libraries using Mass Spectrometry, *in preparation*.

Pieper, R., Luo, G., Cane, D.E., and Khosla, C. (1995). Cell-free synthesis of polyketides by recombinant erythromycin polyketide synthases. Nature *378*, 263-266.

Quadri, L.E.N., Weinreb, P.H., Lei, M., Nakano, M.M., Zuber, P., and Walsh, C.T. (1998). Characterization of Sfp, a Bacillus subtilis Phosphopantetheinyl Transferase for Peptidyl Carrier Protein Domains in Peptide Synthetases. Biochemistry *37*, 1585-1595.

Rond, T., Peralta-Yahya, P., Cheng, X., Northen, T., and Keasling, J. (2013). Versatile synthesis of probes for high-throughput enzyme activity screening. Anal Bioanal Chem *405*, 4969-4973.

Ruiz-Mirazo, K., Briones, C., and de la Escosura, A. (2014). Prebiotic Systems Chemistry: New Perspectives for the Origins of Life. Chem Rev *114*, 285-366.

Serre, L., Verbree, E.C., Dauter, Z., Stuitje, A.R., and Derewenda, Z.S. (1995). The Escherichia coli Malonyl-CoA:Acyl Carrier Protein Transacylase at 1.5-Å Resolution.: Crystal Structure of a Fatty Acid Synthase Component. J Biol Chem *270*, 12961-12964.

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.

Siegel, J.B., Smith, A.L., Poust, S., Wargacki, A.J., Bar-Even, A., Louw, C., Shen, B.W., Eiben, C.B., Tran, H.M., Noor, E.*, et al.* (2015). Computational protein design enables a novel one-carbon assimilation pathway. Proc Natl Acad Sci U S A.

Spaulding, R., and Charles, M. (2002). Comparison of methods for extraction, storage, and silylation of pentafluorobenzyl derivatives of carbonyl compounds and multi-functional carbonyl compounds. Anal Bioanal Chem *372*, 808-816.

Spaulding, R.S., Frazey, P., Rao, X., and Charles, M.J. (1999). Measurement of Hydroxy Carbonyls and Other Carbonyls in Ambient Air Using Pentafluorobenzyl Alcohol as a Chemical Ionization Reagent. Anal Chem *71*, 3420-3427.

Sugimoto, Y., Ding, L., Ishida, K., and Hertweck, C. (2014). Rational Design of Modular Polyketide Synthases: Morphing the Aureothin Pathway into a Luteoreticulin Assembly Line. Angew Chem Int Ed Engl *53*.

Suo, Z., Chen, H., and Walsh, C.T. (2000). Acyl-CoA hydrolysis by the high molecular weight protein 1 subunit of yersiniabactin synthetase: Mutational evidence for a cascade of four acyl-enzyme intermediates during hydrolytic editing. Proc Natl Acad Sci U S A *97*, 14188-14193.

Szafranska, A.E., Hitchman, T.S., Cox, R.J., Crosby, J., and Simpson, T.J. (2002). Kinetic and Mechanistic Analysis of the Malonyl CoA:ACP Transacylase from Streptomyces coelicolor Indicates a Single Catalytically Competent Serine Nucleophile at the Active Site. Biochemistry *41*, 1421-1427.

Troeschel, S.C., Thies, S., Link, O., Real, C.I., Knops, K., Wilhelm, S., Rosenau, F., and Jaeger, K.E. (2012). Novel broad host range shuttle vectors for expression in Escherichia coli, Bacillus subtilis and Pseudomonas putida. Journal of biotechnology *161*, 71-79.

Tsai, S.-C., Miercke, L.J.W., Krucinski, J., Gokhale, R., Chen, J.C.-H., Foster, P.G., Cane, D.E., Khosla, C., and Stroud, R.M. (2001). Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: Versatility from a unique substrate channel. Proc Natl Acad Sci U S A *98*, 14808-14813.

Wang, H.H., and Church, G.M. (2011). Multiplexed Genome Engineering and Genotyping Methods: Applications for Synthetic Biology and Metabolic Engineering. In Methods in Enzymol, C.A. Voigt, ed. (Academic Press), pp. 409-426.

Watrous, J., Roach, P., Alexandrov, T., Heath, B.S., Yang, J.Y., Kersten, R.D., van der Voort, M., Pogliano, K., Gross, H., Raaijmakers, J.M.*, et al.* (2012). Mass spectral molecular networking of living microbial colonies. Proceedings of the National Academy of Sciences of the United States of America *109*, E1743–E1752.

Weissman, K.J. (2006). The Structural Basis for Docking in Modular Polyketide Biosynthesis. Chembiochem *7*, 485-494.

Wheeler, T.J., Clements, J., and Finn, R.D. (2014). Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. BMC Bioinf *15*, 7.

Whicher, J., Smaga, S., Hansen, D., Brown, W., Gerwick, W., Sherman, D., and Smith, J. (2013). Cyanobacterial Polyketide Synthase Docking Domains: A Tool for Engineering Natural Product Biosynthesis. Chem Biol *20*, 1340-1351.

Whicher, J.R., Dutta, S., Hansen, D.A., Hale, W.A., Chemler, J.A., Dosey, A.M., Narayan, A.R.H., Hakansson, K., Sherman, D.H., Smith, J.L.*, et al.* (2014). Structural rearrangements of a polyketide synthase module during its catalytic cycle. Nature *510*, 560-564.

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

Yamada, H., Nagao, A., Nishise, H., and Tani, Y. (1982). Glycerol Dehydrogenase from Cellulomonas sp. NT3060: Purification and Characterization. Agric Biol Chem *46*, 2333-2339.

Yan, J., Gupta, S., Sherman, D.H., and Reynolds, K.A. (2009). Functional Dissection of a Multimodular Polypeptide of the Pikromycin Polyketide Synthase into Monomodules by Using a Matched Pair of Heterologous Docking Domains. Chembiochem *10*, 1537-1543.

Yeates, T.O., Crowley, C.S., and Tanaka, S. (2010). Bacterial Microcompartment Organelles: Protein Shell Structure and Evolution. Annu Rev Biophys *39*, 185-205.

Young, J., Stevens, D.C., Carmichael, R., Tan, J., Rachid, S., Boddy, C.N., Müller, R., and Taylor, R.E. (2013). Elucidation of Gephyronic Acid Biosynthetic Pathway Revealed Unexpected SAM-Dependent Methylations. J Nat Prod *76*, 2269-2276.

Yuzawa, S., Eng, C.H., Katz, L., and Keasling, J.D. (2013). Broad Substrate Specificity of the Loading Didomain of the Lipomycin Polyketide Synthase. Biochemistry *52*, 3791-3793.

Yuzawa, S., Kim, W., Katz, L., and Keasling, J.D. (2012). Heterologous production of polyketides by modular type I polyketide synthases in Escherichia coli. Curr Opin Biotechnol *23*, 727-735.

Zhang, K., Nelson, Kathryn M., Bhuripanyo, K., Grimes, Kimberly D., Zhao, B., Aldrich, Courtney C., and Yin, J. (2013). Engineering the Substrate Specificity of the DhbE Adenylation Domain by Yeast Cell Surface Display. Chem Biol *20*, 92-101.

Zhao, J.Y., Zhong, L., Shen, M.J., Xia, Z.J., Cheng, Q.X., Sun, X., Zhao, G.P., Li, Y.Z., and Qin, Z.J. (2008). Discovery of the autonomously replicating plasmid pMF1 from Myxococcus fulvus and development of a gene cloning system in Myxococcus xanthus. Appl Environ Microbiol *74*, 1980-1987.

8. Appendices

8.1. Supplementary information for Chapter 2 – Mechanistic Analysis of a Computationally Engineered Enzyme that Catalyzes the Formose Reaction

Stds GALD (312)

y = 64135x
R² = 0.99881
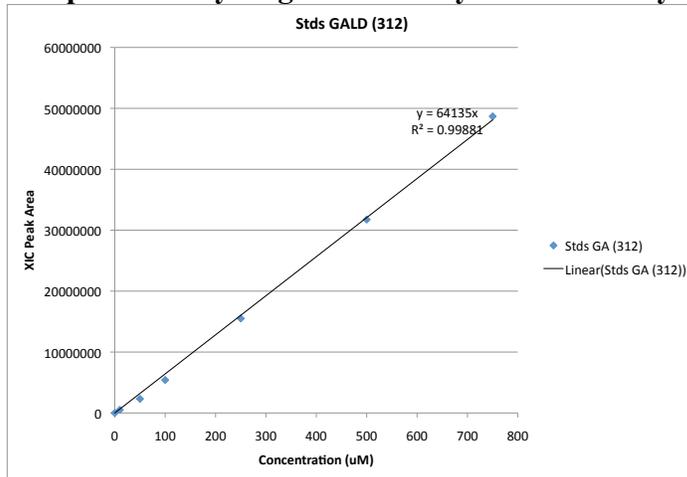
Figure 8.1.1 – Glycolaldehyde standard curve, peaks were integrated using extraction ion chromatograms of the 312 ion

DHA (414)

y = 997.75x
R² = 0.99126
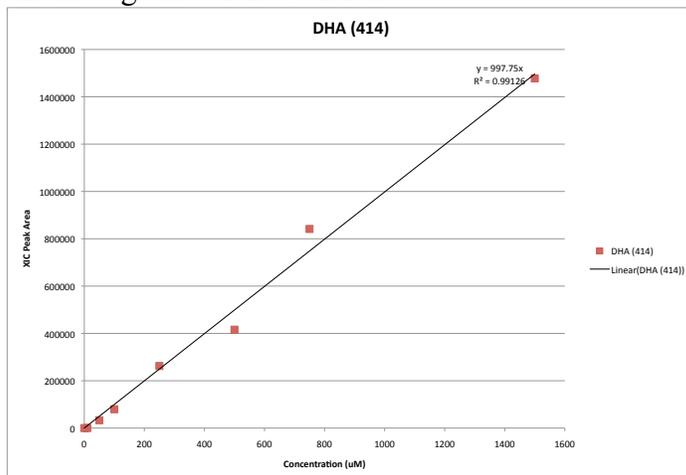
Figure 8.1.2 – Dihydroxyacetone standard curve, peaks were integrated extraction ion chromatograms of the 414 ion.
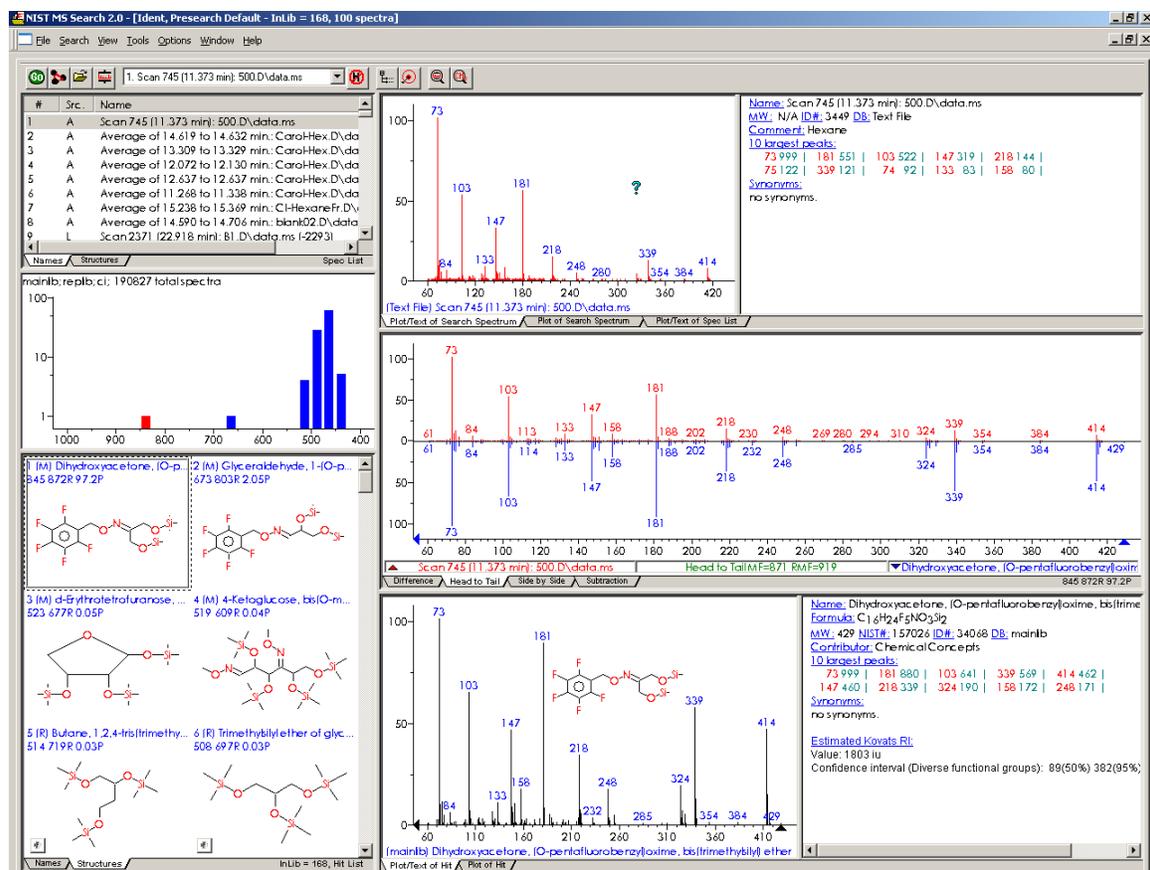
Figure 8.1.3 - Fragmentation pattern of dihydroxyacetone standard and NIST Library matches.

Figure 8.1.4 - Fragmentation pattern of glycolaldehyde standard and NIST Library matches.

Exact Mass: 312.0

[M-CH$_3$]$^+$ GALD-PFBOA-BSTFA Derivative

Exact Mass: 414.1

[M-CH$_3$]$^+$ DHA-PFBOA-BSTFA Derivative

Figure 8.1.5 – Structures and masses of of [M-CH$_3$]$^+$ ions from dihydroxyacetone and glycolaldehyde PFBOA-BSTFA derivatives

8.2. Supplementary information for Chapter 3 – Understanding the Role of Histidine in the GHSxG Acyltransferase Active Site Motif: Evidence for Histidine Stabilization of the Malonyl-Enzyme Intermediate

Detailed Materials and Methods:

Reagents and Chemicals

Malonyl CoA, β-NAD+, NADH, α-ketoglutarate dehydrogenase (porcine heart) (αKGDH), α-ketoglutaric acid, thiamine pyrophosphate (TPP), TCEP and trypsin were purchased from Sigma (St. Louis, MO). HisPur Ni-NTA affinity resin was from Thermo. The Hi-Trap Q anion exchange column was from GE Healthcare. All primers were from IDT.

Cloning: The plasmid pSP8, expressing the wildtype yersiniabactin PKS was cloned by PCR amplification of the DNA sequence encoding amino acids 1-1895 (the PKS fragment) of HMWP1(Miller et al., 2002) using the forward primer 5' – caaca**ccatgg**ataacttgcgcttc - 3' (NcoI site in bold) and the reverse primer 5' - caacac**ctcgag**cgcctctggcttagaacaga - 3' (XhoI site in bold) using pET - 28 HMWP1 as a template.(Suo et al., 2000) The PCR amplicon was then cloned into the NcoI-XhoI sites of pET28b, encoding a C-terminally his-tagged yersiniabactin PKS. The sequence of the resulting pSP8 plasmid is publicly available on the JBEI public registry (public-registry.jbei.org, Part ID: JPUB_003691). Using pSP8 as a template, the S641A mutation was constructed by performing nested PCR encoding the mutation using the primer pairs
CAACAC**CATATG**ATCGTCGCCTCGCTGCCC/CAAATTCACCGAC*GGC*ATGCCC AATGGCGAAGTCTGGC (NdeI site in bold and S641A mutation in bold italics) and CATTGGGCAT*GCC*GTCGGTGAATTTGCCGCTGCCG/CAACAC**GAATTC**CTGAT ACAGCACTTCCACGCCG (XhoI site in bold and S641A mutation in bold italics) to construct a PCR fragment with flanking NdeI-EcoRI sites, encoding a S641A mutation in the PKS sequence between the unique NdeI-EcoRI sites in pSP8. Primers were designed using j5 (Hillson et al., 2011). This PCR fragment was then cloned into the unique NdeI-EcoRI sites in pSP8. To construct the mutations for the H640A and H640A+S641A mutants, Kunkel mutagenesis was performed as described (Linshiz et al., 2014) using the following antisense primers H640A: 5' -
CAAATTCACCGACGGA*AGC*CCCAATGGCGAAGTCTG - 3'. H6540A+S641A 5' - GGCAAATTCACCGAC*GGCAGC*CCCAATGGCGAAGTCTG - 3' (sequence encoding mutations in bold italics). All constructs were sequenced verified by Quintara Biosciences (Richmond, CA).

Purification: An E. coli BLR strain harboring acyl transferase mutants were grown in 1 L 2xYT medium supplemented with appropriate antibiotics at 37°C until the OD600 reached 0.7-0.9. The cultures were then cooled to 18°C and induced with 75 µM isopropyl-β-D-galactopyranoside for 18 h. The cells were harvested by centrifugation

(10000 x g, 5 min) and resuspended in lysis/wash buffer (25 mM Tris pH 8, 500 mM NaCl, 20 mM imidizole). The cells were lysed by sonication and cellular debris was removed by centrifugation (10000 x g, 10 min, 4°C). The supernatant was poured through a fritted column with 5 mL Nickel-NTA agarose resin twice, washed with 10 resin volumes of lysis/wash buffer (4°C), and eluted with 10 mL of elution buffer (25 mM Tris, pH 8, 150 mM imidazole, 4°C). The eluted protein was then applied to a HiTRAP Q anion exchange column (GE Healthcare), washed with 10 resin volumes of start buffer (25 mM Tris pH 8), and eluted at approximately 375 mM NaCl. The eluted protein was concentrated using an Amicon Ultra-15 Centrifugal Filter, 100K device (Millipore). Glycerol was added to the concentrated protein to 8% and was flash frozen in liquid nitrogen and stored at -80°C. The presence of holo-ACP was verified by proteomics. Protein concentrations were determined using Quick Start 1x Bradford Dye Reagent (Bio-Rad) using BSA as a standard.

Hydrolysis assay: Hydrolysis assays were based on a fluorometric transacylase assay described previously.(Molnos et al., 2003) Hydrolysis assays were performed as described previously, using a Synergy H4 Hybrid Multi-Mode Microplate Reader (BioTek), except that EDTA and BSA were omitted from the assay mixture.(Dunn et al., 2013) Final assay concentrations were: 50 mM sodium phosphate, pH 7.4, 10% glycerol, 1 mM TCEP, 0.4 mU/μL αKGDH, 0.4 mM NAD+, 0.4 mM TPP, and 2 mM α-ketoglutaric acid. The concentration of yersiniabactin PKS variants was 1 μM and the concentration of malonyl-CoA was 35 μM.(Bennett et al., 2009) NADH was used to generate a standard curve. All reactions were run in triplicate.

Proteomics-based acylation/transacylation assay: Protein variants at a concentration of 1 μM were incubated with 35 μM malonyl-CoA in 100 mM ammonium bicarbonate, pH 7.8, and 1 mM TCEP for 20 seconds. The addition of malonyl-CoA was used to start the reaction. Reactions were quenched with 50% acetonitrile. The concentration of acetonitrile was diluted to 25% with water and proteins were digested with trypsin (0.5 μg/μl) at 37 °C for 4.5 hours at a ratio of 1:12.5 trypsin:protein. Samples were analyzed on an Agilent 1290 UHPLC system (Agilent Technologies, Santa Clara, CA) coupled to an Agilent 6550 Q-TOF mass spectrometer operating in MS1 mode. 0.25 microgram (μg) of protein was injected and separated on a Sigma Ascentis Peptide Express C-18 column (2.1 mm × 50 mm, 2.7 um particle size; Sigma-Aldrich, St. Louis, MO) at a flow rate of 400 μl/min with gradient conditions as follows: starting with 95% Buffer A (98% water, 2 % acetonitrile, 0.1 % formic acid) and 5% Buffer B (98% acetonitrile, 2% water, 0.1 % formic acid) for one minute, followed by an increase to 35% Buffer B over 5.5 min, followed by a rapid increase to 80% B and a flow rate of 600 μl/min in 1 min, where it was held for 4 min. The solvent composition was quickly ramped to 5% B and the flow rate reduced to 400 μl/min, where it was subsequently held for 2 min to allow the column to equilibrate for the next run. The peptides eluting from the column were ionized by using an Agilent Jet Stream source (sheath gas flow: 11 l/min, sheath gas temperature: 250 c, nozzle voltage: 1,000 v, nebulizing pressure: 35 psi, chamber voltage: 5000 V) operating in positive-ion mode. The data were acquired with MassHunter B.05.00 operating in MS1 mode within 300 m/z to 1400 m/z mass range.

Preparation of extracted ion chromatograms for Figure 2: A) Malonyl-ACP intensity peak intensity is plotted from extracted ion chromatograms of m/z = 1141.0562 +/- 0.02, which is the monoisotopic m/z of the tryptic peptide containing the phosphopantetheine arm of the ACP with a malonyl moiety attached (tryptic ACP peptide sequence: LSDPASLHPNQDLLQLGMDSLLFLELSSDIQHYLGVR). B) Malonyl-AT intensity peak intensity is plotted from extracted ion chromatograms of m/z =1000.9903 +/- 0.02 or m/z = 948.4849 +/- 0.02, which is the monoisotopic mass of the tryptic peptide containing the active site serine of the AT plus a malonyl moiety for wildtype and the H640A mutant, respectively (Wildtype AT tryptic peptide sequence: AEGLKPDFAIGHSVGEFAAAVVCGHYTIEQVMPLVCR).

8.3. Supplementary Information for Chapter 4 – Divergent Mechanistic Routes for the Formation of gem-Dimethyl Groups in the Biosynthesis of Complex Polyketides

Detailed Materials and Methods

Reagents and Chemicals:

Malonyl CoA, methylmalonyl-CoA, S-adenosyl methionine (SAM), cysteine, tris (2-carboxyethyl) phosphine (TCEP) and trypsin were purchased from Sigma (St. Louis, MO). HisPur Ni-NTA affinity resin was from ThermoFisher Scientific (Waltham, MA). Sodium Salicylate was purchased from Fisher (Waltham, MA). The Hi-Trap Q anion exchange column was from GE Healthcare (Little Chalfont, Buckinghamshire, UK). All primers were from IDT (Coralville, IA). Sfp was purified as described.(Quadri et al., 1998)

Cloning:

The plasmid pSP8, expressing the wildtype yersiniabactin PKS was cloned by PCR amplification of the DNA sequence encoding amino acids 1-1895 (the PKS fragment) of HMWP1(Miller et al., 2002) using the forward primer 5' – caacaccatggataacttgcgcttc - 3' (NcoI site in bold) and the reverse primer 5' - caacacctcgagcgcctctggcttagaacaga - 3' (XhoI site in bold) using pET - 28 HMWP1 as a template.(Suo et al., 2000) The PCR amplicon was then cloned into the NcoI-XhoI sites of pET28b, encoding a C-terminally his-tagged yersiniabactin PKS. The sequence of the resulting pSP8 plasmid is publicly available on the JBEI public registry (public-registry.jbei.org, part ID: JPUB_003691). The epothilone M8 PKS DNA was synthesized by DNA 2.0 (Menlo Park, CA) with flanking NdeI and BamHI sites in a pUC vector. This vector was digested with NdeI/BamHI and ligated into a similarly digested version of pET28b. The sequence of the resulting pSP47 plasmid is publicly available on the JBEI public registry (public-registry.jbei.org, Part ID: JPUB_004475). Plasmids pHMWP2.CH8 (Keating et al., 2000) for expression of the first NRPS in the yersinaibactin cluster and pH6YbtE,(Gehring et al., 1998) which expresses a standalone adenylation domain in the cluster were obtained from the Christopher T. Walsh laboratory at Harvard University and are available on the JBEI public registry (public-registry.jbei.org, Part ID: JPUB_004473 and JPUB_004474, respectively).

Protein purification:

HMWP2

HMWP2 was purified as described, (Keating et al., 2000; McLoughlin and Kelleher, 2005) with minor modifications. pHMWP2.CH8 was transformed into BL21(DE3) *E. coli* and plated on LB agar with 50 µg/mL kanamycin. Cultures of the HMWP2 were grown at 30°C at 200 rpm in 2xYT media supplemented with 2 mM MgCl (1 L) and 50 µg/mL kanamycin. When the OD_{600} reached 0.4, the temperature of the culture was

lowered to 18°C. When the OD_{600} reached 0.8, the HMWP2 was induced with 50 μM IPTG. Growth continued 15 h and the cells were harvested by centrifugation. The cell pellets were resuspended in 25 mM Tris-HCl pH 8.0 / 500 mM NaCl / 20 mM imidazole before lysis by several passages through an Emulsiflex C3 homogenizer (Avestin, Mannheim, Germany). The lysates were clarified by centrifugation (30 min at 10,000 x g) followed by two passages through 2 mL Ni-NTA resin. The resin was then washed with buffer containing 20 mM imidizole prior to the elution of the proteins in elution buffer, 25 mM Tris-HCl pH 8 / 500 mM NaCl / 150 mM imidazole. The eluted protein was concentrated using an Amicon Ultra-15 Centrifugal Filter, 100K device (Millipore, Billerica, MA) and dialyzed against 25 mM Tris-HCl pH 8 / 1 mM TCEP / 10% and was flash frozen in liquid nitrogen and stored at -80°C. Protein concentrations were determined using Quick Start 1x Bradford Dye Reagent (Bio-Rad, Hercules, CA) using BSA as a standard.

YbtE

YbtE was purified as described,(Gehring et al., 1998; McLoughlin and Kelleher, 2005) with minor modifications. pH6YbtE was transformed into BLR *E. coli* and plated on LB agar with 100 μg/mL amp. Cultures of the YbtE were grown at 37°C at 200 rpm in 2xYT media (1 L) with 100 μg/mL amp. When the OD_{600} reached 0.8 the culture was induced with 1 mM IPTG. Growth continued for 4 h and the cells were harvested by centrifugation. The cell pellet was resuspended in 50 mM Phosphate pH 7.6 / 500 mM NaCl / 10 mM imidazole before lysis with a sonicator. The lysates were clarified by centrifugation (10 min at 10,000 x g) followed by incubation for 1 hour 4°C with 5 mL Ni-NTA resin. The resin was then washed with wash buffer, 50 mM Phosphate pH 7.6 / 500 mM NaCl / 20 mM prior to the elution of the proteins in elution buffer, 150 mM Phosphate pH 7.6 / 150 mM imidazole. The eluted protein was concentrated using an Amicon Ultra-15 Centrifugal Filter, 100K device and buffer exchanged using the concentrator into (8% Glycerol, 50 mM Phosphate pH 7.6 , 2 mM DTT, 10 mM $MgCL_2$) and was flash frozen in liquid nitrogen and stored at -80°C. Protein concentrations were determined using Quick Start 1x Bradford Dye Reagent using BSA as a standard.

Yersiniabactin PKS

An *E. coli* BLR strain with pSP8 was grown in 1 L 2xYT medium supplemented with 50 μg/mL kanamycin 37°C until the OD_{600} reached 0.7-0.9. The cultures were then cooled to 18°C and induced with 75 μM isopropyl-β-D-galactopyranoside for 18 h. The cells were harvested by centrifugation (10000 x g, 5 min) and resuspended in lysis/wash buffer (25 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidizole). The cells were lysed by sonication and cellular debris was removed by centrifugation (10000 x g, 10 min, 4°C). The supernatant was poured through a fritted column with 5 mL Nickel-NTA agarose resin twice, washed with 10 resin volumes of lysis/wash buffer (4°C), and eluted with 10 mL of elution buffer (25 mM Tris, pH 8.0, 150 mM imidazole, 4°C). The eluted protein was then applied to a HiTRAP Q anion exchange column, washed with 10 resin volumes of start buffer (25 mM Tris pH 8.0), and eluted at approximately 375 mM NaCl. The eluted

protein was concentrated using an Amicon Ultra-15 Centrifugal Filter, 100K device. Glycerol was added to the concentrated protein to 8% and was flash frozen in liquid nitrogen and stored at -80°C. Protein concentrations were determined using Quick Start 1x Bradford Dye Reagent using BSA as a standard.

Apo and holo Epothilone M8 PKS

An *E. coli* BLR (for apo protein) or BAP1(Pfeifer et al., 2001) (for holo protein) strain harboring pSP47 mutants were grown in 4 L 2xYT medium supplemented with 50 mg/L kanamycin at 37°C until the OD_{600} reached 0.7-0.9. The cultures were then cooled to 18°C and induced with 75 µM isopropyl-β-D-galactopyranoside for 18 h. The cells were harvested by centrifugation (10000 x g, 5 min) and resuspended in lysis/wash buffer (25 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidizole). The cells were lysed by sonication and cellular debris was removed by centrifugation (10000 x g, 10 min, 4°C). The supernatant was poured through a fritted column with 1.5 mL Nickel-NTA agarose resin twice, washed with 10 resin volumes of lysis/wash buffer (4°C), and eluted with 10 mL of elution buffer (25 mM Tris, pH 8.0, 150 mM imidizole, 4°C). The eluted protein was concentrated using an Amicon Ultra-15 Centrifugal Filter, 100K device. Glycerol was added to the concentrated protein to 8% and was flash frozen in liquid nitrogen and stored at -80°C. Protein concentrations were determined using Quick Start 1x Bradford Dye Reagent using BSA as a standard.

Synthesis of Substrates:

Diketide-SNAC

(2*S*,3*R*)-3-hydroxy-2-methylpentanoyl-*S*-*N*-Acetyl-cysteamine. Prepared as described.(Sharma and Boddy, 2007) Evans' acyl oxazolidinone chemistry was used to construct the two stereocenters in enantiopure form from the aldol condensation reaction. ^{1}H NMR (400 MHz, CDCl$_3$): δ 6.10 (s, br, 1H), 3.90-3.84 (m, 1H), 3.51-3.40 (m, 2H), 3.11-3.00 (m, 2H), 2.80-2.73 (m, 1H), 2.69 (s, br, 1H), 1.99 (s, 3H), 1.60-1.43 (m, 2H), 1.22 (d, *J* = 6.8 Hz, 3H), 0.99 (t, *J* = 7.2 Hz, 3H).

NAME   DEN-C-86
EXPNO   1
PROCNO   1
Date_   20131118
Time   12.12
INSTRUM   AVB-400
PROBHD 5 mm PABBO BB-
PULPROG   zg30
TD   65536
SOLVENT   CDCl3
NS   15
DS   0
SWH   8278.146 Hz
FIDRES   0.126314 Hz
AQ   3.9584243 sec
RG   90.5
DW   60.400 usec
DE   6.00 usec
TE   294.2 K
D1   4.00000000 sec
TD0   1

======== CHANNEL f1 ========
NUC1   1H
P1   8.20 usec
PL1   -3.00 dB
PL1W   23.05461311 W
SFO1 400.1324710 MHz
SI   131072
SF   400.1300000 MHz
WDW   EM
SSB   0
LB   0.30 Hz
GB   0
PC   4.00
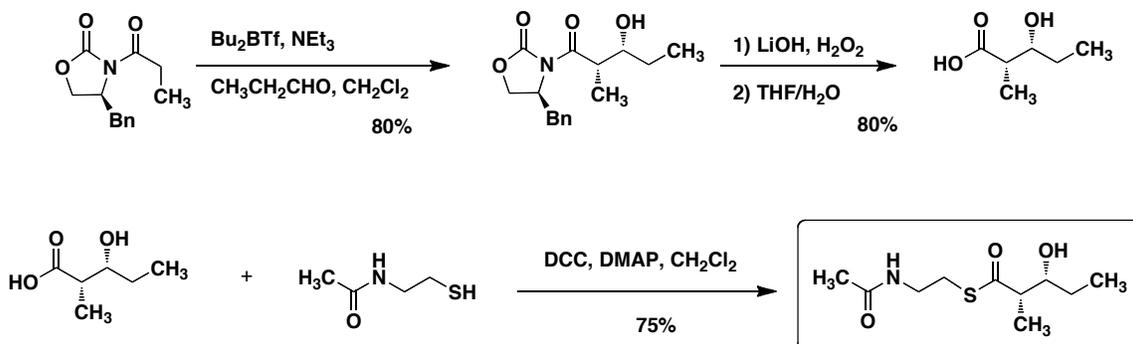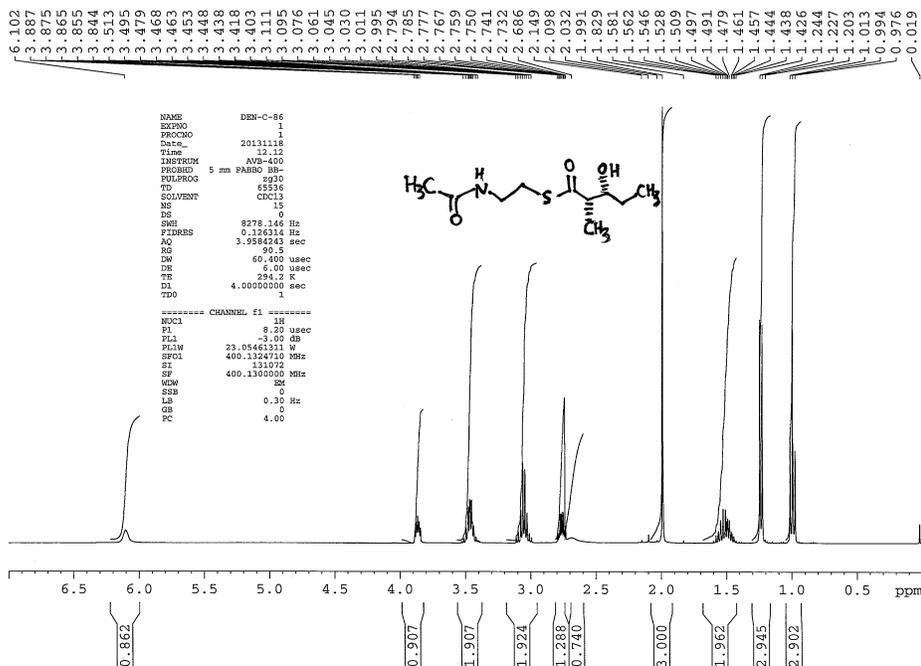
^{1}H NMR spectrum (CDCl3, 400 MHz) of (2S,3R)-3-hydroxy-2-methylpentanoyl-S-N-Acetyl-cysteamine.

Dimethylmalnonyl-thiophenol Thioester

A dry 25 mL round bottom flask was charged with 150 mg (1.13 mmol, 1.6 eq.) dimethylmalonic acid, 10 mL CH$_3$CN, 236 mg (1.13, 1.6 eq.) thiophenol and subsequently cooled to 0 °C. To this 79 mg (0.67 mmol, 1.0 eq) diisopropylcarbodiimide was dissolved in 3 mL CH$_3$CN and added dropwise over 10 min. The reaction was let warm to room temperature and continue over 12 h at which point it was quenched with 90 mL saturated aqueous sodium bicarbonate solution and filtered. The aqueous layer was extracted with Et$_2$O (2 x 25 mL) and then brought to pH 3 with 1 N HCl. The aqueous layer was then extracted with EtOAc (5x 20 mL, or until the monothiophenol thioester is fully extracted). EtOAc layers were combined and washed with brine (1 x 30 mL) and dried over Na$_2$SO$_4$. The EtOAc was removed in vacuo and the residue was, due to compound lability, quickly recrystallized from Et$_2$O /Hexanes to afford a fine white crystal (33.2 mg/24% yield). Diisopropyl urea was unable to be purified from title compound and residual byproduct from the coupling agent remained with the final compound. TLC 1:1:0.02/Hexanes:EtOAc:AcOH R$_f$ = 0.4. ^{1}H NMR (CDCl$_3$, 400 MHz) δ 1.64, (s, 1H), 7.46 (s, 5H).
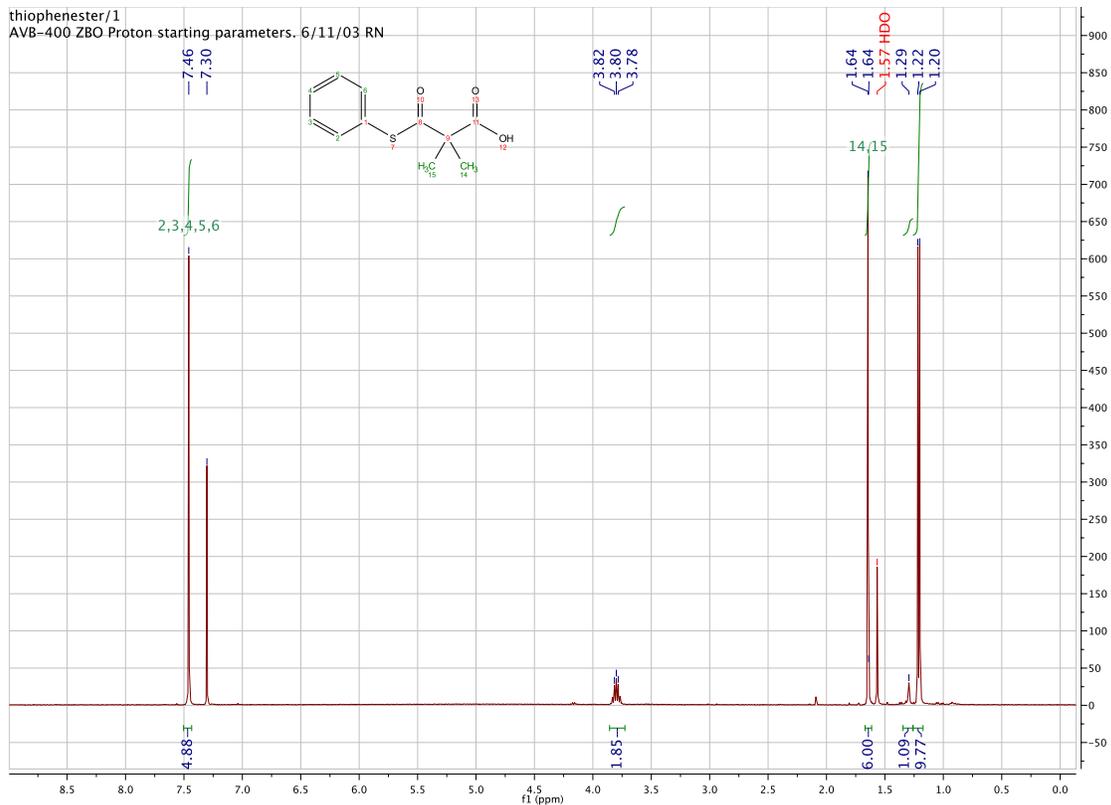
^{1}H NMR spectrum (CDCl$_3$, 400 MHz) of dimethylmalnonyl-thiophenol thioester.

Dimethylmalnonyl-CoA

9.8 mg (12.7 μMol, 1.0 eq) coenzyme A was dissolved in 2 mL 1:1 H_2O and THF with 5.36 mg (63.2 μMol, 1.0 eq) sodium bicarbonate and cooled to 5 °C. To the cooled solution was added 11.2 mg (50.4 μMol, 4 eq.) dimethylmalonyl thiophenol thioester. Reaction proceeded at 5 °C and was allowed to warm to room temperature over 4 h at which time the pH was lowered to 4.0 and extracted 3x with 500 μL EtOAc. Combined extracts were dried over Na_2SO_4 and concentrated to 500 μL for purification by HPLC. Title compound was purified on an Agilent 1260 fitted with a Zorbax SB-C18 Semi-preparative column (9.4 x 50 mm, 5 micron). A gradient consisting of buffer A (50 mM NaH_2PO_4) and B (MeOH) was used: 5% B to 10 min, ramp to 90% B at 15 min, hold for 2 min and back to 5% B at 20 min. Compound matched previously reported dimethylmalonyl CoA.(Batchelar et al., 2008; Bretschneider et al., 2012) (Bretschneider et al., 2012) ^{1}H NMR (400 MHz, D_2O) δ = 8.49 (s,1H), 8.23 (s, 1H), 6.11 (d, *J* = 4.0 Hz, 1H), 4.51 (s, 1H), 4.15 (s, 2H), 3.92 (s, 1H), 3.74 (m, 1H), 3.46 (m, 1H), 3.36 (t, *J* = 8.0 Hz, 2H), 3.23 (t, *J* = 8.0 Hz, 2H), 2.91 (t, *J* = 8.0 Hz, 2H), 2.34 (t, *J* = 8.0 Hz, 2H), 1.28 (s, 6H), 0.80 (s, 3H), 0.66 (s, 3H). H 41 and 43 are obscured by the H_2O signal.

^{1}H NMR spectrum (D_2O, 400 MHz) of dimethylmalonyl-CoA.

In vitro loading and extension assays:

In vitro experiments were performed with both epothilone and yersinaibactin PKS proteins at a concentration of 2.5 µM in 100 mM phosphate buffer, pH 7.2 (for yersiniabactin experiments) and pH 7.6 (for epothilone M8 PKS experiments), 1 mM TCEP at a temperature of 30°C.

For epothilone M8 PKS full reconstitution experiments (Figure 8.3.4), reagent concentrations were: 5 mM diketide-SNAC, 0.75 mM SAM, and 0.5 mM methylmalonyl-CoA or malonyl-CoA. Holo protein was utilized in the full reconstitution experiments. The addition of diketide-SNAC was used to start the reaction and SAM was omitted for the minus SAM experiment in Figure 4 and Figure 8.3.7; all samples were quenched after 16 min. For epothilone Sfp dimethylmalonyl-ACP loading experiments (Figure 2), 10 mM MgCl$_2$, 10 µM Sfp and 0.1 mM dimethylmalonyl-CoA were incubated at 30°C with the apo-PKS protein for 45 minutes, to charge dimethylmalonyl-ACP prior to the addition of diketide-SNAC to start the reaction. For the determination of the AT substrate preference (Figure 8.3.1): malonyl-CoA; methylmalonyl-CoA; and both malonyl-CoA and methylmalonyl-CoA were incubated with the EpoM8 PKS at a concentration of 0.5 mM and quenched after 16 min.

For yersiniabactin PKS full reconstitution experiments (Figure 8.3.4) experiments both the HMWP2 and yersiniabactin PKS were converted to their holo forms using 10 uM sfp in the presence of 10 mM MgCl$_2$, 0.5 mM CoA, 0.5 mM cysteine, and 0.5 mM salicylic acid for 25 min at 30°C. 10 mM ATP and 10 uM YbtE were added to form acyl KS and then reactions were started by the simultaneous addition of 0.5 mM malonyl-CoA and 0.75 mM SAM. Acetyl-ACP was also monitored during full reconstitution experiments in the presence of SAM, as shown in Figure 8.3.2. To form unmethylated extended product as shown in Figure 3, SAM was omitted for 10 min, and then added to start the reaction. For yersiniabactin Sfp dimethylmalonyl-ACP loading experiments (Figure 2), 50 µM PKS was incubated with 10 mM MgCl$_2$, 10 µM Sfp and 0.1 mM dimethylmalonyl-CoA at 30°C for 60 minutes, to form dimethylmalonyl-ACP. Simultaneously, HMWP2 was converted to its holo form using 10 uM sfp in the presence of 10 mM MgCl$_2$, 0.5 mM CoA, 0.5 mM cysteine, and 0.5 mM salicylic acid for 25 min at 30°C. 10 mM ATP and 10 uM YbtE were added and allowed to react for 60 min to form acyl HMWP2. Subsequently, dimethylmalonyl-ACP loaded PKS was mixed with acyl-HMWP2 to start the reaction.

Reactions were quenched with 50% acetonitrile after 20 seconds for yersiniabactin PKS dimethylmalonyl-ACP extension experiments (Figure 2 B) and SAM omission experiments (Figure 3). Reactions were quenched with 50% acetonitrile after 64 minutes for epothilone M8 PKS dimethylmalonyl-ACP extension experiments (Figure 2), as the epothilone M8 enzyme as this enzyme took much longer than the yersiniabactin PKS to turn over. After quenching the concentration of acetonitrile was diluted to 25% with water and proteins were digested with trypsin (0.5 µg/µl) at 37 °C for 2 hours at a ratio of 1:10 trypsin:protein prior to LC/MS/MS analysis.

LC/MS/MS Proteomics methods:

Samples were analyzed on an AB Sciex (Foster City, CA) 4000 Q-Trap mass spectrometer operating in MRM (SRM) mode coupled to an Agilent 1100 system. 1-2 μg of total peptide was injected onto a Sigma (St. Louis, MO) Ascentis Peptide Express C-18 column (2.1 mm x 50 mm) via an autosampler. For the epothilone module 8 samples a 32.5-minute method was used and consisted of a 400 μL/minute flow rate, starting with 95 % Buffer A (2% Acetonitrile, 0.1% formic acid) and 5% Buffer B (98 % Acetonitrile, 0.1% formic acid) for 1.2 minutes, followed by a rapid increase to 25% Buffer B in 1 minute and then a gradual increase to 37.5% B in 19.8 minutes. This was followed by a sharp increase to 90% B, where it was held for 2 minutes, followed by a quick ramp back down to 5% B, where it was subsequently held for 6 minutes to allow for column equilibration for the next run. For yersiniabactin samples, a similar 40.5 minute method was used with the same flow-rate and buffer compositions as above with a longer gradual increase, as the yersiniabactin tryptic ACP peptide was more hydrophobic. As above, the method began with 5% buffer B for 1.2 minutes followed by a rapid rise to 25% over 1 minute and then a very slow rise to 51% B over 25.8 minutes. After the slow gradient step, buffer B was rapidly increased to 90%, held, and dropped back down to re-equilibrate the column as above. The peptides eluting from the column were ionized by using Turbo V Ion source (curtain gas flow: 20 L/min, temperature: 400 C, ion spray voltage: 4,800 V, ion source gas flow: 50 L/min, entrance potential: 10 V) operating in positive-ion mode. Targeted methods were designed using Skyline (MacLean et al., 2010) and data collected in Analyst 1.5.1 and data was quantified in MultiQuant 2.1 (AB Sciex).

Peptides and MS/MS Transitions Monitored

Epothilone M8 ACP tryptic peptide:

HVPFSNLGMDSLIGLELR

Table 8.3.1 - Transitions monitored for epothilone M8 PKS

Name	Structure	*m/z* parent	*Z* (parent)	*m/z* daughter
Malonyl-ACP	(structure)	808.7	3	347.12
Methylmalonyl-ACP	(structure)	813.4	3	361.14
Isobutyryl-ACP	(structure)	803.4	3	331.16
Unmethylayed extended-ACP	(structure)	832.1	3	417.20
Singly methylayed extended-ACP	(structure)	836.7	3	431.22
Methylated extended-ACP	(structure)	841.1	3	445.23

Yersiniabactin PKS ACP tryptic peptide:

LSDPASLHPNQDLLQLGMDSLLFLELSSDIQHYLGVR

Table 8.3.2 - Transitions monitored for yersiniabactin PKS

Name	Structure	m/z parent	Z	m/z daughter
Acetyl-ACP		904.2	5	303.13
		1130.1	4	
Isobutyryl-ACP		909.9	5	331.16
		1137.0	4	
dimethylmalonyl - ACP		916.6	5	375.15
		1148.0	4	
Unmethylated extended-ACP		962.2	5	593.16
		1202.6	4	
Singly methylated extended-ACP		965.1	5	607.17
		1206.1	4	
Methylated extended-ACP		967.9	5	621.18
		1209.6	4	
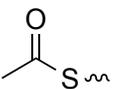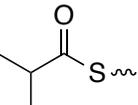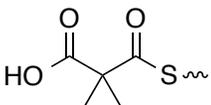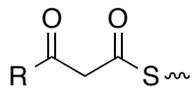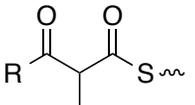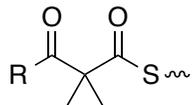
R represents the acyl chain transferred from the previous NRPS module: a hydroxyphenylthiazolinylthiazolinyl (HPTT) moiety in this case.

Supplementary Figures
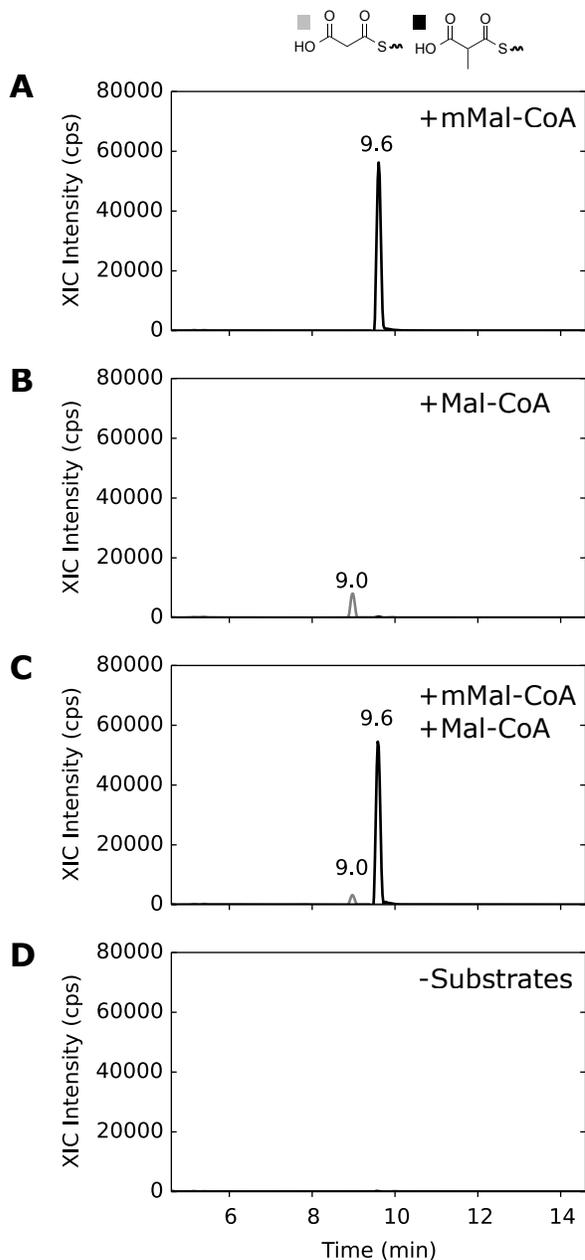
Figure 8.3.1 – Determination of Epo M8 AT loading substrate specificity by incubation with various substrates. A) Methylmalonyl-CoA loading. B) Malonyl-CoA Loading. C) Competition between methylmalonyl-CoA and malonyl-CoA. D) No substrate control. Methylmalonyl-CoA appears to be preferred by the EpoM8 PKS, however some activity toward malonyl-CoA was observed.
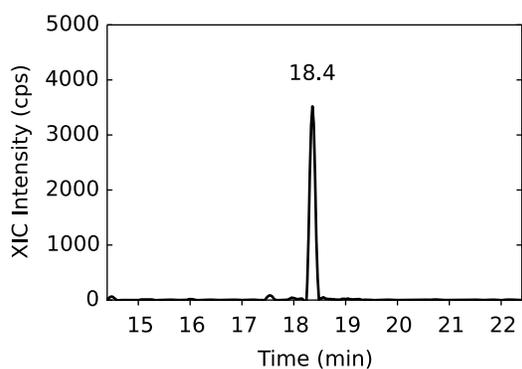
Figure 8.3.2 – Acetyl-ACP was still present on the yersiniabactin enzyme after 10 min incubation in the presence of SAM, suggesting that acetyl-ACP is a dead end intermediate and cannot be methylated to form isobutryl-ACP (parent ion: z=4).
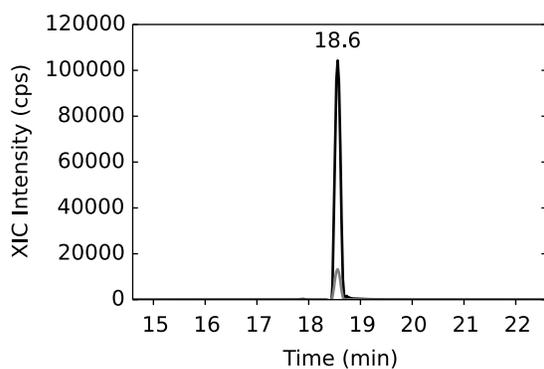
Figure 8.3.3 –Dimethylmalonyl-ACP chromatogram for yersiniabactin PKS after loading using Sfp. Parent ion: z=4 (black) and z=5 (grey).
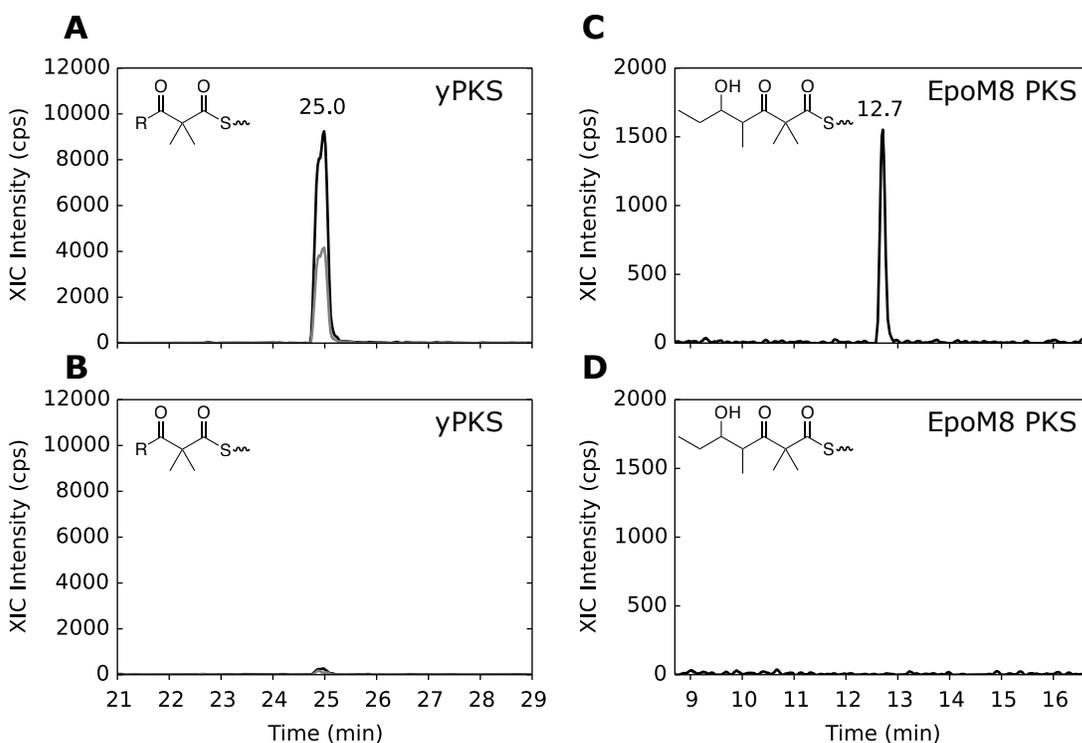
Figure 8.3.4 – Yersiniabactin PKS and Epothilone M8 PKS reactions with malonyl-CoA or methylmalonyl-CoA substrate in the presence of SAM (full reconstitution). A) Yersiniabactin PKS extended methylated product after 20 sec incubation. (parent ion: $z=5$ (black) and $z=4$ (grey)) B) Yersiniabactin PKS before the addition of substrates to form acyl-KS. C) Epothilone module 8 PKS extended, methylated product chromatogram after 16 minute incubation with the diketide-SNAC. D) Epothilone module 8 PKS before the addition of diketide-SNAC. For the yersiniabactin case, R represents the acyl chain transferred from the previous NRPS module: a HPTT moiety.
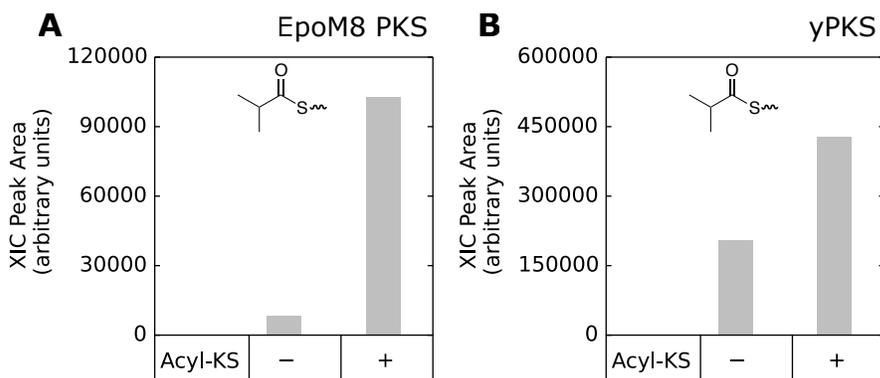
Figure 8.3.5 – Isobutyryl-ACP was produced from dimethylmalonyl-ACP after the formation of acyl-KS by both the epothilone PKS A) and yersinaibactin PKS B). Minus acyl-KS samples were taken after the sfp loading reactions, immediately before the formation of acyl-KS. These data are from the same experiments performed to generate the data in Figure 2 of the main text (parent ion $z=4$).
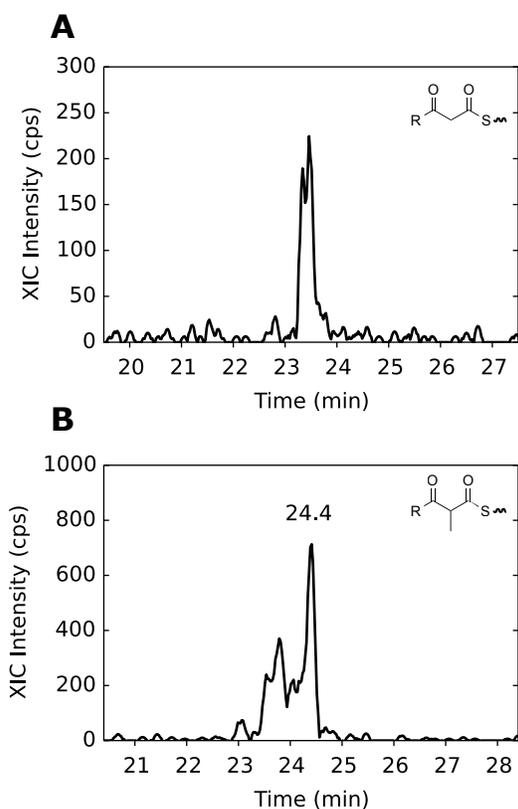
Figure 8.3.6 – Unmethylated A) and singly methylated B) products produced by yersiniabactin PKS during full reconstitution reactions after 20 second incubation. R represents the acyl chain transferred from the previous NRPS module: a HPTT moiety (parent ion z=5).
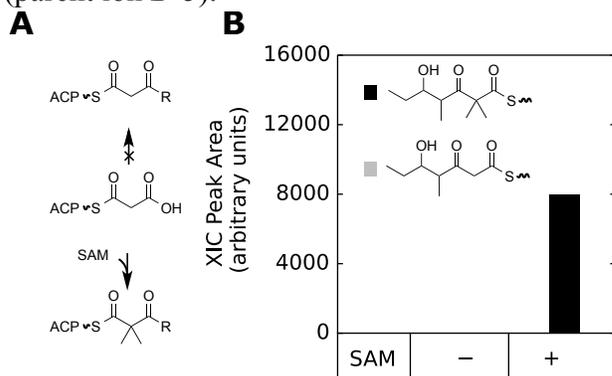
Figure 8.3.7 – Characterization of Route 1 (KS catalyzed condensation precedes methylation) for the epothilone PKS. A) Experiment overview: using malonyl-CoA as an extender unit SAM was omitted from the overall reaction (no product formed), or SAM was included, allowing the complete reaction to take place. B) Unmethylated extended-ACP extracted ion chromatogram (XIC) peak areas (gray) and methylated extended-ACP

peak areas (black). All XIC m/z values monitored in this study are summarized in Tables 8.3.1 and 8.3.2.

8.4. Supplementary information for Chapter 5 – Incorporation of Gem-Dimethyl Groups into Engineered Polyketide Products

K_{cat}/K_m determination protocol:

k_{cat}/K_M Determination

I)   Prepare enzyme master mix:
   a.   Measure protein aliquot concentration: 2ul of 1:10 dilution into 1mL Bradford dye. Incubate for 5 minutes, measure OD_{595} and calculate [enzyme] using spreadsheet.
   b.   Combine enzyme and phosphate buffer
      i.   Only use phosphate buffer, pH 8. Do not use phosphate lysis buffer.
      ii.   Final enzyme MM volume = 86ul x (#rxn wells + 1)
      iii.   Based on this volume, add enzyme to 11.63uM in MM. This will give [Enzyme]$_{final}$ = 10uM.
   c.   Aliquot 86ul of enzyme MM into a separate 0.6ml tube for each [SNAC] condition (usually 5-6).
II)   Prepare assay master mixes:
   a.   In separate tubes labeled 0.3, 1, 3, 10, 25, 50mM SNAC, add:
      i.   Vol. 10mM DTNB = 4ul
      ii.   Vol 20% DMSO = 10ul - (Vol. SNAC)
      iii.   Vol. SNAC:
         1.   0.3mM: 3ul of 10mM stock
         2.   1mM: 1ul of 100mM stock
         3.   3mM: 3ul of 100mM stock
         4.   10mM: 10ul of 100mM stock
         5.   25mM: 5ul of 500mM stock
         6.   50mM: 10ul of 500mM stock
   b.   If testing multiple enzymes with the same concentration gradient, prepare assay master mix for the number of proteins tested + 1.
III)   Perform plate reader assay:
   a.   Configure Spectramax:
      i.   Login to plate reader PC and open Spectramax software.
      ii.   From "Protocols" select "Enzymology > Michaelis-Menten"
      iii.   Under "Plate 01 – *Settings*" select a kinetic screen:
         1.   Take absorbance readings at 412nm every 15s for 2min.
         2.   Only read the wells containing the reaction.
      iv.   Under "Plate 01 – *Options*:"
         1.   Assign each well a concentration and a name (default – Enz 01)
IV)   Run assay:
   a.   Sequentially add 14ul of assay MM to to 86ul of enzyme MM for each condition.

b. Mix thoroughly using a P100 before pipetting 91ul of the reaction into its assigned well on the 96 well reader plate.

c. When all conditions have been mixed and added to the plate, click the green "READ" button in the Spectramax software.

V) Process data:

a. When assay is complete, save the .pda file and export the data to a .txt document.

b. Extract data points using bash, copy and paste into a template Excel spreadsheet.

c. Obtain parameter estimates:

i. Estimate the V_{max}, K_M, and k_{cat} using the charts in the template. Feed these values into Sean's Least Squared template spreadsheet to calculate the kcat/Km.

ii. Alternatively, use a MATLAB script that reads the first 6 columns from the Spectramax .txt output file

Robotic Ellman's Based Screening:

384 Well Protein Expressions and Prep

- Day 1: OVERNIGHTS
    - Pick a single colony from plate and inoculate 0.18ml of room temp LB/Kan+10%Glycerol (NOT TB) in STERILE shallow well plate (Costar3795)
    - Pick +/- controls: if using QPix robot, leave wells H10-H12 as negative controls, inoculate by hand.
        - H10-H12: Blank – Negative – Positive control
    - Cover with breathable seal
    - Shake at 37deg 16-24hrs in warm room plate shaker (260 rpm; check humidity) to ensure they are all fully grown. Stack in groups of 4.
- Day 2: EXPRESSION
    - Inoculate a fresh STERILE plate of 180 μl of TB/Kan in a shallow well plate with 3.6μL of overnight cultures (FX protocol: Sean LB to TB transfer).
    - Cover growth plate with breathable seal.
    - Grow at 37deg as above for 4 hours (target OD approx 1). Stack in groups of 4 with an empty plate on top as the 5th plate to prevent evaporation.
    - Add IPTG to 0.25mM (5 uL per well of 10mM IPTG) using the FX (FX protocol: Induce_new tips_10uL, change volume)
    - Transfer plate to 37 deg shaker (260 rpm; check humidity) for 3 hrs
- Day 3: STORE
    - Transfer 4x 96 wells to 384 well plate. (Find FX Protocol "Sean 384 to 96 pooling_short)
    - Spin down 384 well plates at 4000rpm for 5 min
    - Pour off supernatant
    - Store cells in -80 until ready for purification
- Assay (~60min/plate)

- o Use Abs 412nm Assay
- o Lysis buffer/60 uL per well: (50 mM HEPES pH 7.4, 1% Tween, 2 mM MgCl$_2$, 150 mM NaCl, 0.1 mg/mL Lysozyme)
- o 384 Well Assay mix: 15 uL Assay Master Mix, 20 uL Lysate from each well, Final Concentration SNAC = 3 mM (7 mM in master mix), Final Concentration Ellman's Reagent 0.4 = mM (0.93 mM in master mix)

Kunkel Mutagenesis:

Kunkel mutagenesis was performed as described (Linshiz et al., 2014). The 96 amino acids nearest the active site of the pikromycin TE were selected using pymol pSP37 numbering (16, 19, 20, 24, 25, 26, 27, 28, 29, 29, 30, 31, 32, 33, 74, 75, 76, 77, 78, 79, 80, 81, 82, 85, 104, 107, 122, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 205, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 236, 237, 241, 251, 252, 253, 254, 255, 267, 268, 269, 270, 271, 272, 273). NNK kunkel oligos were designed using the oligo_desiger MATLAB script from the Ostermeier lab (Firnberg and Ostermeier, 2012). Oligos were pooled into 8 reactions each with 8 primers for annealing and extension. All 8 reactions were pooled after extension and transformed into electrocompetent BLR.

In Vitro and In Lysate Assays:

For screening of DEBSM5 and yersiniabactin PKS engineered linker constructs (**Figure 5.10**) all polyketide constructs were overexpressed in BAP1 (Pfeifer et al., 2001) overnight at 18°C in LB media with 75 μM IPTG. The ACP4 construct was overexpresed in *E. coli* strain BLR with 1 mM IPTG for 4 hr at 30°C in LB media. Aliquots of cognate lysates (0.75 mL each) were combined, centrifuged and frozen at -80°C for later use. These pellets were lysed using sonication in 500 μL 10% glycerol, 1 mM TCEP, 100 mM phosphate buffer pH 7.2 and centrifuged. Frozen 0.5 mL aliquots of ACP4 were lysed in 500 μL of the same buffer and centrifuged. To 200 μL of clarified ACP4 lysate, 10 mM MgCl$_2$, 5 μM sfp and 250 μM isobutryl-CoA were added to form isobutyryl-ACP4 from apo-ACP4 in lysate. The loading reaction was allowed to run for 45 min at 30°C. 25 μL of the cognate PKS clarified supernatants were combined with 25 μL of the loaded ACP4 lysate; as well as 0.5 mM methyl malonyl-CoA and 0.5 mM NADPH (substrates for DEBSM5); and 0.5 malonyl-CoA, and 0.75 mM SAM (substrates for the yersiniabactin PKS). After a 30 min reaction at 30°C, lysates were trypsinized with 1:10 trypsin:protein, cleaned up using a C18 tip (see lysate tip clean up protocol below) and injected onto a Agilent 6460 LC-MS.

For PCP loading and extension experiments (**Figure 5.11**), acyl-PCP was formed in a reaction with 10 mM MgCl$_2$, 5 μM sfp, 100 mM phosphate buffer pH 7.2, 1 mM TCEP, 250 μM isobutryl-CoA/3-hydroxy-butryryl-CoA and 10 μM SP5 protein for 45 min at 30°C. After PCP loading, yersiniabactin PKS protein (construct pSP8) was added to 5 μM along with 0.5 mM NADPH, 0.5 malonyl-CoA, and 0.75 mM SAM. The extension

reaction as incubated for 30 min reaction at 30°C, and then were trypsinized with 1:10 trypsin:protein, and then injected onto a Agilent 6460 LC-MS.

Lysate tip clean up protocol:

Desalting for mass spectrometry (modified from http://qb3.berkeley.edu/qb3/pmsl/docs/Sample%20desalting%20for%20mass%20spectrometry.pdf)

Equipment:

C18 Spec tips, Agilent, cat #A57203

p200 and p1000 pipetmen

You can use a p1000 pipetman to wash and elute spec tips. For each wash, add the required volume into the top of the spec tip, then press a p1000 tip firmly into the spec tip and depress the plunger to drive the liquid through the device.

1. Make 15 uL lysate samples up to 100 ul with 85 uL of 5% formic acid

2. Wet the spec tip by pushing through 200 μl of HPLC grade methanol.

3. Wash the spec tip with 3 x 200μl Water /5% formic acid.

4. Push the sample through the spec tip. You may pass the sample through twice, ((20 uL of digested PKS, diluted with 85 uL H2O pushed through 3x) washing with water/5% formic acid between each pass, if you would like to maximize binding.

5. Wash with 2 x 200 μl water/ 5% formic acid.

6. Elute with 2 x 100 μl 80% acetonitrile/5% formic acid.

7. Inject 5 to 20 μg protein

Reagents:

HPLC grade methanol

HPLC grade water/5% formic acid

50% HPLC grade acetonitrile/5% formic acid

Make solutions with MilliQ or better water. NOTE: Do not pipet concentrated formic acid with plastic pipet tips; use glass.

8.5. Detailed Methods for Using Skyline for Phosphopantetheine Proteomics

Step 1: Get Fasta File for Sequence of interest (yPKS.fasta for this example, the yersiniabactin PKS)

Step 2: Paste fasta file into skyline (skyline calculates peptide fragmentation per http://www.matrixscience.com/help/fragmentation_help.html, we're going to teach it ppant ejection ion fragmentation)

Step 3: Add holo-ACP transition to structural modification list

Add modification: Settings -> Peptide Settings -> Modifications Tab -> Edit List -> Add -> Show All -> Phosphopantetheine -> Click on Variable

Turn on DSL containing peptide with Phosphopantetheine arm: Protein -> Right Click -> Pick Children -> Find DSL containing peptide (at end, example sequence LSDPASLHPNQDLLQLGMDSLLFLELSSDIQHYLGVR).

To calculate fragmentation using skyline move phosphopantetheine mod to C terminus, add neutral loss: Peptide Settings -> Modifications Tab -> Edit List -> Edit -> Move to C terminus -> Loss -> Add loss of Arginine (or lysine if C terminus is a lysine) and PO_3 -> Turn on updated peptide per B) -> select y_1^+ ion with neutral loss

Step 4: Add Acyl-ACP (malonyl-ACP as example, see chem draw file)

Peptide Settings -> Modifications Tab -> Edit List -> Copy -> Edit copy -> Add acid weight, subtract water to chemical formula -> select y_1^+ ion with neutral loss

Step 5: Optimize Collision Energies (35 is optimal for ppant ejection for Aglient 6460)

Settings -> Transition Settings -> Collision Energy -> Add -> Charge 4 (change depending on charge state of your peptide of interest), Slope 0, Intercept 35

Step 6: Export Method

File -> Export -> Method -> Instrument type Aglient 6400 series -> choose template method

Open method in Aglient Software, Add in Sample Bypass, Run Method

Step 7: Import Results

File -> Import -> Results

Step 8: Repeat with your favorite protein!

yPKS.fasta:
>yPKS
MDNLRFSSAPTADSIDASIAQHYPDCEPVAVIGYACHFPESPDGETFWQNLLEGR
ECSRRFTREELLAVGLDAAIIDDPHYVNIGTVLDNADCFDATLFGYSRQEAESMD
PQQRLFLQAVVWHALEHAGYAPGAVPHKTGVFASSRMSTYPGREALNVTEVAQV
KGLQSLMGNDKDYIATRAAYKLNLHGPALSVQTACSSSLVAVHLACESLRAGES
DMAVAGGVALSFPQQAGYRYQPGMIFSPDGHCRPFDASAEGTWAGNGLGCVV
LRRLRDALLSGDPIISVILSSAVNNDGNRKVGYTAPSVAGQQAVIEEALMLAAID
DRQVGYIETHGTGTPLGDAIEIEALRNVYAPRPQDQRCALGSVKSNMGHLDTAA
GIAGLLKTVLAVSRGQIPPLLNFHTPNPALKLEESPFTIPVSAQAWQDEMRYAGV
SSFGIGGTNCHMIVASLPDALNARLPNTDSGRKSTALLLSAASDSALRRLATDYA
GALRENADASSLAFTALHARRLDLPFRLAAPLNRETAEALSAWAGEKSGALVYS
GHGASGKQVWLFTGQGSHWRTMGQTMYQHSTAFADTLDRCFSACSEMLTPSL
REAMFNPDSAQLDNMAWAQPAIVAFEIAMAAHWRAEGLKPDFAIGHSVGEFAA
AVVCGHYTIEQVMPLVCRRGALMQQCASGAMVAVFADEDTLMPLARQFELDL
AANNGTQHTVFSGPEARLAVFCATLSQHDINYRRLSVTGAAHSALLEPILDRFQD
ACAGLHAEPGQIPIISTLTADVIDESTLNQADYWRRHMRQPVRFIQSIQVAHQLG
ARVFLEMGPDAQLVACGQREYRDNAYWIASARRNKEASDVLNQALLQLYAAG
VALPWADLLAGDGQRIAAPCYPFDTERYWKERVSPACEPADAALSAGLEVASR
AATALDLPRLEALKQCATRLHAIYVDQLVQRCTGDAIENGVDAMTIMRRGRLLP
RYQQLLQRLLNNCVVDGDYRCTDGRYVRARPIEHQQRESLLTELAGYCEGFQAI
PDTIARAGDRLYEMMSGAEEPVAIIFPQSASDGVEVLYQEFSFGRYFNQIAAGVL
RGIVQTRQPRQPLRILEVGGGTGGTTAWLLPELNGVPALEYHFTDISALFTRRAQ
QKFADYDFVKYSELDLEKEAQSQGFQAQSYDLIVAANVIHATRHIGRTLDNLRPL
LKPGGRLLMREITQPMRLFDFVFGPLVLPLQDLDAREGELFLTTAQWQQQCRHA
GFSKVAWLPQDGSPTAGMSEHIILATLPGQAVSAVTFTAPSEPVLGQALTDNGD
YLADWSDCAGQPERFNARWQEAWRLLSQRHGDALPVEPPPVAAPEWLGKVRL
SWQNEAFSRGQMRVEARHPTGEWLPLSPAAPLPAPQTHYQWRWTPLNVASIDH
PLTFSFSAGTLARSDELAQYGIIHDPHASSRLMIVEESEDTLALAEKVIAALTASA
AGLIVVTRRAWRVEENEALSASHHALWALLRVAANEQPERLLAAIDLAENTPW
ETLHQGLSAVSLSQRWLAARGDTLWLPSLAPNTGCAAELPANVFTGDSRWHLV
TGAFGGLGRLAVNWLREKGARRIALLAPRVDESWLRDVEGGQTRVCRCDVGD
AGQLATVLDDLAANGGIAGAIHAAGVLADAPLQELDDHQLAAVFAVKAQAASQ
LLQTLRNHDGRYLILYSSAAATLGAPGQSAHALACGYLDGLAQQFSTLDAPKTL
SVAWGAWGESGRAATPEMLATLASRGMGALSDAEGCWHLEQAVMRGAPWRL
AMRVFTDKMPPLQQALFNISATEKAATPVIPPADDNAFNGSLSDETAVMAWLKK
RIAVQLRLSDPASLHPNQDLLQLGMDSLLFLELSSDIQHYLGVRINAERAWQDLS
PHGLTQLICSKPEALEHHHHHH*

Structure of Holo Daughter Ion:
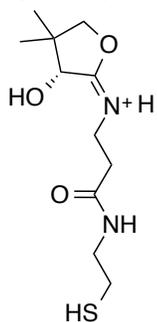
Chemical Formula: $C_{11}H_{21}N_2O_3S^+$
Exact Mass: 261.13

Structure of a malonyl daughter ion:
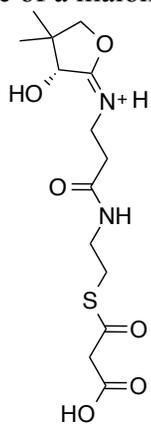
Chemical Formula: $C_{14}H_{23}N_2O_6S^+$
Exact Mass: 347.13

m/z of 4+ holo parent ion, made using biopolymer tool in chem. draw

H—L·S·D·P·A·S·L·H·P·N—Q D·L·L·Q·L·G M D—N... L·L·F·L·E·L·S·S·D·I—Q H Y·L·G V·R—OH
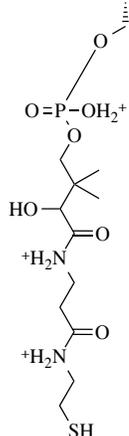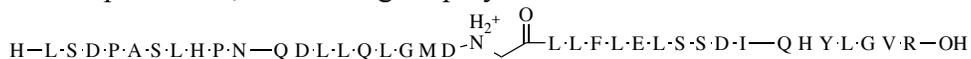
m/z: 1120.06 (100.0%), 1119.81 (95.3%), 1120.31 (45.9%), 1119.56 (45.2%), 1120.56 (35.6%), 1120.31 (23.6%), 1120.31 (18.5%), 1120.06 (17.6%), 1120.56 (13.2%), 1120.31 (12.5%), 1120.56 (9.0%), 1120.31 (8.6%), 1120.56 (8.5%), 1119.81 (8.3%), 1120.81 (7.9%), 1120.81 (6.3%), 1120.81 (6.0%), 1120.81 (5.8%), 1120.56 (4.4%), 1120.06 (4.3%), 1120.81 (4.2%), 1120.06 (4.1%), 1121.06 (3.2%), 1121.06 (2.6%), 1121.06 (2.2%), 1120.81 (2.1%), 1120.06 (1.7%), 1120.81 (1.7%), 1120.31 (1.6%), 1120.56 (1.6%), 1120.06 (1.5%), 1121.06 (1.2%), 1120.81 (1.1%), 1120.81 (1.1%)

8.6. Plasmids used

Name	Description	JBEI Part ID	Chapter described
pET28-L90T	Most evolved Formolase in a pET 29 vector	JBx_024128	2
pET28-BAL	Benzaldehyde Lyase in a pET 29 vector	JBx_045355	2
pSP8	Expresses the yersiniabactin PKS in pET28	JBx_014927	3,4,5
pSP40	Yersiniabactin PKS – S641A	JBx_025255	3
pIY1	Yersiniabactin PKS – S641A+H640A	JBx_026080	3
pSP43	Yersiniabactin PKS –H640A	JBx_040191	3
pSP47	Expresses the epothilone module 8 PKS in pET28	JBx_027026	4
pH6YbtE	Expresses the standalone adenylation domain in the yersiniabactin cluster	JBx_022169	4
pHMWP2.CH8	Expresses of the first NRPS in the yersinaibactin cluster	JBx_020657	4
pSP2	Yersiniabactin PKS with erythromycin TE appended	JBx_014739	5
pSP5	PCP1-S1439A-Cy2-PCP2 (acyl donor to yersiniabactin PKS)	JBx_014743	5
pSY48	Expresses the DEBS thioesterase in a pET vector	JBx_021101	5
pSP37	Expresses the Pikromycin thioesterase in a pET vector	JBx_022926	5
pSP37-A218T	Pikromycin TE – A218T	JBx_024733	5
pSP37-G150A	Pikromycin TE – G150A	JBx_045354	5
pBS31	DEBSM5+ Yersiniabactin NRPS C-term linker	JBx_031175	5
pBS32	DEBSM5+CurG ddKS	JBx_031181	5
pBS33	DEBSM5+ CurK ddKS	JBx_031182	5
pBS34	DEBSM5+ ddPLM1	JBx_031183	5
pBS35	DEBSM5+ EryM4 ddKS	JBx_031185	5
pBS36	DEBSM5+ EryM2 ddKS	JBx_031184	5
pBS41	CurL ACP dd+ YPKS	JBx_031177	5
pBS42	CurH ACP dd + YPKS	JBx_031176	5
pBS43	Plm2 C-term linker+ YPKS	JBx_031178	5
pBS44	EryM3 ACP dd + YPKS	JBx_031180	5
pBS45	EryM5 ACP dd+ YPKS	JBx_031179	5
pBS48	Erythromyin ACP4 in pET28 Vector	JBx_039691	5

8.7. Curriculum Vitae

SEAN POUST

Email: poust@berkeley.edu
Address: Joint BioEnergy Institute, 5885 Hollis St., 4th Floor, Emeryville, CA 94608
Phone: (847) 644-2711

EDUCATION

University of California-Berkeley — Berkeley, CA
PhD, Chemical Engineering — Expected May 2015
Dissertation: Engineering polyketide synthases using mass spectrometry

University of Illinois-Urbana Champaign — Urbana, IL
MS, Environmental Engineering — August 2010
Dissertation: Engineered vesicles for perchlorate degradation

University of Illinois-Urbana Champaign — Urbana, IL
BS with highest honors, Civil and Environmental Engineering — May 2007
Thesis: The effect of short-term shear on membrane fouling and soluble EPS release
Minors in International Studies and Environmental Studies

RESEARCH EXPERIENCE

University of California-Berkeley — Berkeley, CA
Graduate Student Researcher (Ph.D. Candidate); Advisor: Jay D. Keasling — 2010-Present
Understanding and engineering polyketide synthases using mass spectrometry
— Established lab-wide utilization of tandem and high-resolution mass spectrometry to examine acyl intermediates on polyketide synthases for mechanistic studies and enzyme engineering.
— Created robotic methods and informatics pipeline for high-throughput screening, hit picking and sequencing for evolution of two different enzyme classes.
— Developed LC-MS/MS and GC-MS methods for polyketide product formation, aldolase activity and terpene production.

University of Illinois-Urbana Champaign — Urbana, IL
Graduate Student Researcher (Master's level); Advisor: Julie L. Zilles — 2009-2010
Nanoreactors imbued with functional biomolecular machinery
— Expressed, purified and demonstrated active encapsulation of perchlorate reducing enzymes in polymer vesicles for perchlorate contaminated water treatment.

PROFESSIONAL EXPERIENCE

Carollo Engineers — Sarasota, FL
Engineer — 2007-2009
— Lead engineer for three design areas of the Lake Bradford Road Water Reuse Facility in Tallahassee, FL.
— Coordinated pilot studies for UV reactor validation, biological fixed-bed perchlorate removal and performance evaluation of new water filtration systems.

AWARDS

Best Poster, Society for Industrial Microbiology and Biotechnology Conference   2014
NSF Graduate Research Fellowship   2009
Hertz Fellowship Finalist   2009
Research Assistant of the Year, Army Corps of Engineers - CERL   2009
City of the Future Competition Winner, IBM Headquarters, Chicago, IL   2007
Ira O. Baker First Prize (Given to top student in Civil Engineering), Univ. of Illinois   2007
Bronze Tablet (GPA in top 3% of graduating class), Univ. of Illinois   2007
Knight of St. Patrick (Engineering leadership award), Univ. of Illinois   2007
Langelier, George Farnsworth Jr. Scholarship, other undergraduate scholarships   2003-2007

PUBLICATIONS

Poust, S., Phelan, R., Katz, L., Petzold, C. J., and Keasling, J. D. Divergent Mechanistic Routes for gem-Dimethyl Group Formation in Complex Polyketide Biosynthesis, *Angewandte Chemie 127*, 2400-2403.

Siegel, J.B., Smith A.L., **Poust, S.**, Wargacki, A.J., Bar-Even, A., Louw, C., Shen, B.W., Eiben, C.B., Tran, H.M., Noor, E., Gallaher, J., Bale, J., Yoshikuni. Y, Gelb, M.H., Keasling, J.D., Stoddard, B.L., Lidstrom, M.E., and Baker, D. Design and Construction of a Novel One-Carbon Assimilation Pathway, *Proceedings of the National Academies of Sciences*. (accepted)

Poust, S., Hagen, A., Katz, L., and Keasling, J. D. (2014) Narrowing the Gap Between the Promise and Reality of Polyketide Synthases as a Synthetic Biology Platform, *Current Opinion in Biotechnology 30*, 32-39.

Poust, S., Yoon, I., Adams, P. D., Katz, L., Petzold, C. J., and Keasling, J. D. (2014) Understanding the Role of Histidine in the GHSxG Acyltransferase Active Site Motif: Evidence for Histidine Stabilization of the Malonyl-Enzyme Intermediate, *PloS One 9*, e109421.

Hagen, A., **Poust, S.**, de Rond, T., Yuzawa, S., Katz, L., Adams, P. D., Petzold, C. J., and Keasling, J. D. (2014) In Vitro Analysis of Carboxyacyl Substrate Tolerance in the Loading and First Extension Modules of Borrelidin PKS, *Biochemistry 53*, 5975-5977.

Linshiz, G., Stawski, N., Goyal, G., Bi, C., **Poust, S.**, Sharma, M., Mutalik, V., Keasling, J. D., and Hillson, N. J. (2014) PR-PR: Cross-Platform Laboratory Automation System, *ACS Synthetic Biology 3*, 515-524.

Hutchison, J. M., **Poust, S.**, Kumar, M., Cropek, D. M., MacAllister, I. E., Arnett, C. M., and Zilles, J. L. (2013) Perchlorate Reduction Using Free and Encapsulated *Azospira oryzae* Enzymes, *Environmental Science & Technology 47*, 9934-9941.

Linshiz, G., Stawski, N., **Poust, S.**, Bi, C., Keasling, J. D., and Hillson, N. J. (2012) PaR-PaR Laboratory Automation Platform, *ACS Synthetic Biology 2*, 216-222.

Kumar, M., Payne, M. M., **Poust, S.**, and Zilles, J. L. (2012) Polymer-Based Biomimetic Membranes for Desalination, In *Biomimetic Membranes for Sensor and Separation Applications*, pp 43-62, Springer Netherlands.

Brown, J., Lauderdale, C., Estavo, G., Ettori, A., Shih, W., **Poust, S.**, Walker, S., Raskin, L., Upadhyaya, G., and Li, X. (2008) Direct Fixed-Bed Biological Perchlorate Destruction

Demonstration, Department of Defense Environmental Security Certification Program, Report Number ER-0544, Arlington, Virginia.

PUBLICATIONS IN PREPARATION

Poust, S., Piety, J., Bar-Even, A., Louw, C., Baker, D., Keasling, J. D., and Siegel, J.B. Mechanistic Analysis of a Engineered Enzyme that Catalyzes the Formose Reaction.

Hagen, A., **Poust, S.**, Li, R., Katz, L., Petzold, C. J., and Keasling, J. D. Characterization of the iterative module of the Borrelidin PKS.

Hagen, A., **Poust, S.**, de Rond, T., Fortman, J.L., Katz, L., Petzold, C. J., and Keasling, J. D. Biological production of the nylon precursor adipic acid using a highly-engineered PKS.

SELECTED CONFERENCE PRESENTATIONS

Poust, S., Hagen, A., Phelan, R., Katz, L., Petzold, C. J., and Keasling, J. D. "Debugging Polyketide Synthases Using Mass Spectrometry." Poster presentation at the Synthetic Biology Engineering Research Conference. Boston, MA, September 2014.

Poust, S., Phelan, R., Katz, L., Petzold, C. J., and Keasling, J. D. "Understanding the Order of Methylation and Condensation in Geminal-Dimethyl Producing Polyketide Synthase Domains." Poster presentation at the Society for Industrial Microbiology and Biotechnology Conference. St. Louis, MO, July 2014. (Best poster award)

Poust, S., Phelan, R., Katz, L., Petzold, C. J., and Keasling, J. D. "Understanding the Order of Methylation and Condensation in Geminal-Dimethyl Producing Polyketide Synthase Domains." Poster presentation at the Bioorganic Chemistry Gordon Conference. Andover, NH, June 2014.

Poust, S., Kumar, M., Cropek, D., Zilles, J., MacAllister, I. "Synthetic Cells with Embedded Biomolecular Machinery." Oral presentation at the US Army Corps of Engineers Research and Development Conference. Memphis, TN, November 2009.