Investigations into the Structure and Function of Type I Polyketide Synthases

by

Aaron A. Koch

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Cancer Biology)
in the University of Michigan
2017

Doctoral Committee:

        Professor David H. Sherman, Chair
        Professor Mark Day
        Professor Tom Kerppola
        Professor Benjamin Margolis
        Professor Janet Smith

Aaron Andrew Koch

aakoch@umich.edu

ORCID iD: 0000-0001-7488-9074

Acknowledgements

My scientific journey has been as linear in trajectory as a sinusoidal wave function. I am forever indebted to a multitude of mentors who invested their time and energy into keeping my path towards a productive endpoint. During my undergraduate studies at Grand Valley State University I was blessed to work for mentors fully committed to the education and advancement of their students. I thank Professor David Leonard for instigating my fascination with biochemistry, especially enzymes. I am also grateful to Professor Richard Rediske for providing me the opportunity to enter the world of natural products research, a passion with which I am still afflicted today despite my initial assignment of collecting and analyzing raw sewage. During these years I was also fortunate to receive the opportunity to perform research at my future graduate school as a summer student in Dr. Colin Duckett's lab. I thank professor Duckett for graciously providing me with this opportunity, scientific mentorship, and a paycheck; all of which were desperately needed. The inspiration and guidance of these mentors during my formative years ultimately led me to the laboratory of Dr. David Sherman at the University of Michigan where I have been fortunate to pursue my scientific interests in an exceptional research environment. I thank Dr. Sherman for affording me the opportunity to carry out the research outlined in this thesis under his guidance during my graduate studies. I also thank my committee members Dr. Mark Day, Dr. Tom Kerppola, Dr. Benjamin Margolis, and Dr. Janet Smith for providing insightful and encouraging feedback during the course of my time at the University of Michigan. Finally, I thank both former and present fellow scientists within the Sherman lab, especially Dr. Joe Chemler, Dr. Doug Hansen, Dr. Jeff Kittendorf, Dr. Andrew Lowell, Dr. Sean Newmister, and Vik Shende for their contributions and advice. I would be remiss if I did not make special mention of Dr. Hansen who served as my mentor during my graduate career on a volunteer basis (I hope that decision does not haunt you) and contributed significantly to many of the projects contained within this thesis.

Table of Contents

List of Figures

vi

List of Schemes

List of Tables

Abstract

The polyketide (PK) class of natural products constitutes an abundant array of secondary metabolites produced in microorganisms, many of which possess potential medicinal value, especially in the area of oncology. Polyketides are assembled biosynthetically via the megaenzymes polyketide synthases (PKSs) through an assembly line process of stepwise condensations of simple malonic acid building blocks derived from primary metabolism. Despite the usefulness of natural products in medicine, the development of polyketide natural products into new drugs is often hindered by their suboptimal pharmacological properties, highlighting the need for their modification by medicinal chemistry. However, low natural abundance and high structural complexity often necessitates lengthy and expensive synthetic routes to natural product analogs, thus impeding their clinical development. A promising method for expanding the chemical diversity within polyketide natural products is PKS bioengineering, whereby natural product analogs are generated by engineering new functionality into the enzymes responsible for their production instead of through synthetic derivatization. While notable successes in PKS engineering have been achieved, many attempts result in decreased product yields or fail to produce the predicted molecules entirely.

The studies in this thesis focus on investigating the structural and mechanistic parameters that govern PKS catalysis in order to increase the potential of harnessing these enzymes as biocatalysts for the production of new polyketide analogs. First, a series of engineered PKS modules was generated by combining modules from the pikromycin, erythromycin, and juvenimycin biosynthetic pathways with non-native TE domains and analyzed for substrate flexibility in vitro. The results from this study implicated the TE domain as the dominant catalytic bottleneck in the full-module processing of unnatural substrates. We next focused our investigations on probing the TE directly as an excised domain, subsequently confirming the previously observed

catalytic bottleneck. Mutational analysis of the Pik TE domain resulted in an engineered variant (S148C) with improved substrate flexibility and catalytic efficiency, which eliminated the aforementioned bottleneck and allowed for the production of diastereomeric macrolactone analogs. Finally, we performed molecular dynamics (MD) simulations coupled with quantum mechanical (QM) calculations of the native and engineered TE domains to provide a mechanistic rational for our experimental observations. Taken together, the results herein provide further insight into the catalytic and mechanistic parameters that govern the productive functioning of engineered PKSs. Our identification of the thioesterase domain as a key catalytic bottleneck in the processing of unnatural substrates builds the groundwork for future engineering of PKS TE domains in order to generate more flexible catalysts for the production of novel natural product analogs.

Chapter I

Introduction

Therapeutic natural products

Natural products (secondary metabolites) are organic compounds which are not directly required for the normal growth, development, or reproduction of the producing organism, yet provide a selective long-term advantage in their survival.[1] Human use of natural products from both marine and terrestrial sources as therapeutic agents to combat a diverse array of diseases predates recorded human history. Some of the earliest recorded examples of the medicinal use of natural products come from cuneiform inscriptions on clay tablets from Mesopotamia (2600 B.C.) which documented the therapeutic application of many plant derived substances including *Cupressus sempervirens* (Cypress) and *Commiphora* species (myrrh) oils in the treatment of respiratory ailments.[2]

Historically, the use of natural products as medicines has been in the form of crude traditional medicines, remedies, and potions with the bioactive component(s) unidentified. The arrival of modern chemistry starting in the 18th century marked a new era in the study and use of natural products. Technological advances in analytical and structural chemistry made possible the purification and characterization of the individual components of natural product extracts. This shift in paradigm is exemplified by the German pharmacist Friedrich Wilhelm Sertürner's isolation of morphine from opium, constituting the first pure natural product medicine.[3] Since that time, Western pharmaceutical companies have pursued natural product drug discovery and development extensively, resulting in numerous clinically successful natural product therapeutics. Examples[4] of clinically employed natural products with diverse bioactivities

1

include the antibiotic (erythromycin A), anticancer (epothilone), cholesterol-lowering agent (lovastatin), antiparasitic (avermectin), insecticide (spinosyn A), anagelsic (morphine), and immunosuppressant (rapamycin) (Figure 1.1). Highlighting the ubiquitous use of natural product derived molecules in the clinic, Newman et al. noted in a recent review[1] that of the 246 small molecule therapeutic agents approved worldwide from the 1930s to 2014 in the area of oncology alone, 160 (77%) of these were either natural products or synthetic compounds inspired by natural products.

Erythromycin A
antibiotic

Epothilone B
anticancer

Lovastatin
cholesterol lowering

Avermectin
antiparasitic

Spinosyn A
insecticidal

Morphine
anagelsic

Rapamycin
immunosuppressant
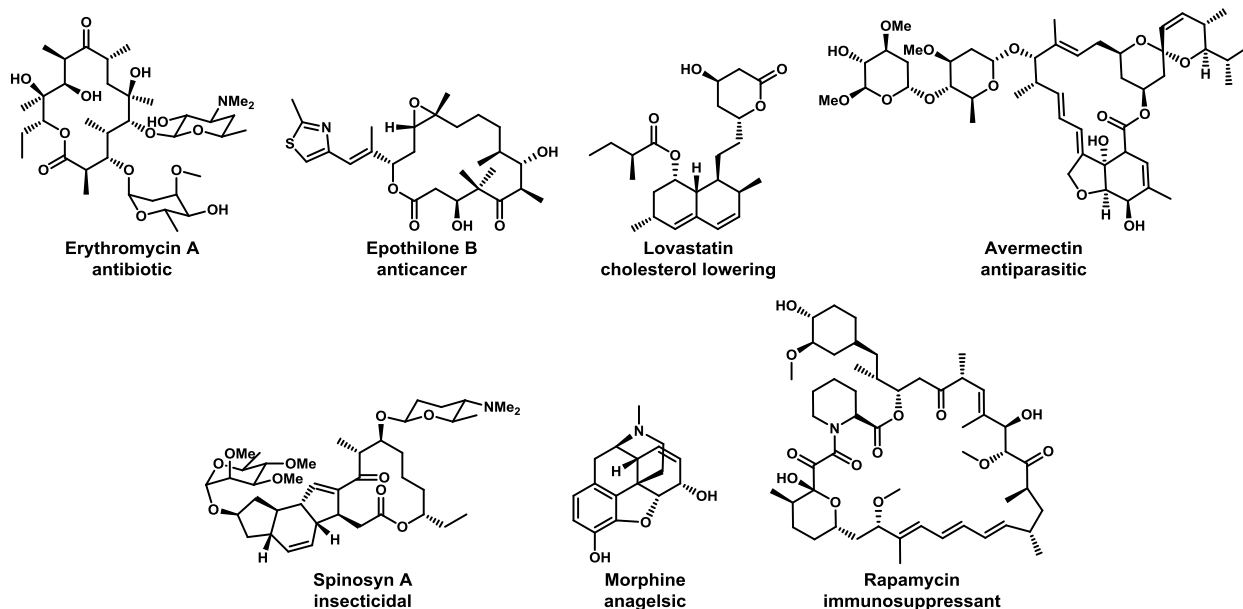
Figure 1.1 Examples of clinically valuable bioactive natural products.

Despite this clinical success, pharmaceutical research into natural products has largely declined over the past two decades.[2] A number of factors contribute to this contraction, with two prominent developments being the introduction of high-throughput screening (HTS) methods against specific molecular targets and the development of combinatorial chemistry. HTS methods have allowed for the focused targeting of a growing number of precise molecular targets, however, these assays are typically incompatible with natural product extracts and have traditionally relied on synthetic libraries. Concomitant with the development of HTS methods was the introduction of combinatorial chemistry which promised to supply HTS assays with chemical libraries of simpler, more drug-like molecules. As a result of these technological changes, the current climate in pharmaceutical dug discovery relies on rapid screening of simple

2

chemical libraries, initial hit identification, followed by significant medicinal chemistry on the lead molecule to establish structure-activity relationships (SAR). Thus, the comparably resource intensive natural product drug discovery model involving extract-library preparation and screening, bioassay-guided isolation, structure elucidation, and analog generation has been largely abandoned.

However, this shift away from natural products drug discovery has resulted in significant negative consequences. It is now well recognized that synthetic chemical libraries contain molecules that are economically efficient to generate, and thus significantly less complex than their natural products counterparts.[3] Compared to synthetic chemical libraries, natural product-structures possess increased levels of chemical diversity arising from their high molecular mass, increased steric complexity from a large number of chiral centers, increased molecular flexibility from a large number of rotatable bonds, and unique distribution of heteroatoms.[4] These molecular properties combined with their biochemical specificity lend natural products as favorable lead structures for drug discovery. Natural products serve as excellent lead compounds in drug discovery as they have high chemical diversity, biochemical specificity and other favorable molecular properties.[2] Additionally, natural products often contain "privileged scaffolds" referring to the fact that natural products have been evolved under selective pressure for interaction with a wide variety of biological targets for a specific purpose.[2,5] It is this evolutionarily optimized scaffold that enables many natural products to interact with challenging molecular targets and imparts them with their exquisite biological activities, thus providing an advantage over synthetic libraries often employed in drug discovery efforts.

Polyketide natural products

The polyketide class of natural products constitutes an abundant array of secondary metabolites, many of which possess therapeutic or industrial value[1], constituting sales in the billions of dollars.[6] Polyketides represent complex chemical scaffolds arising from their highly diverse composition of functional groups. The manifold skeletal and stereochemical complexity decorating polyketides, along with their

potent bioactivities, has made them a keen area of interest for a broad range of scientific disciplines.[7]

Of particular allure, polyketides often contain a macrocyclic core, defined as a ring architecture containing 12 or more atoms.[8] Macrocyclic polyketides (e.g. all except lovastatin and morphine, Figure 1.1) are of great interest to medicinal chemists as the conformational preorganization of these large rings allows for the precise display of key functional groups to interact with challenging drug targets, such as protein-protein interactions.[8] Despite the proven therapeutic potential of macrocyclic natural products (there are currently >100 marketed macrocyclic drugs[8]), macrocyclic compounds remain an under-explored section of drug discovery. One significant contributing factor to this lack of exploration is the formidable challenge in generating macrocyclic compounds from their linear precursors. Regio- and stereoselective ring-closures are notoriously difficult and often require dilute reactions conditions with large volumes and low reactant concentrations.[8,9] Furthermore, as evidenced from the pioneering total synthesis of Erythromycin A[10], macrolactonization of acyclic analogs is often unachievable even in optimized conditions as the productive conformations necessary for the linear substrate to achieve for ring closure are likely perturbed from the natural "competent conformation".[11] These synthetic hurdles to the production of macrocyclic polyketide analogs thus limit their development into clinically useful therapeutics, despite their potential for interacting with challenging macromolecular targets.

Type I polyketide biosynthesis

Polyketides are assembled biosynthetically via the megaenzymes polyketide synthases (PKS) through an assembly line process of stepwise condensations of simple monomer building blocks in a manner analogous to fatty acid biosynthesis.[7] Modular type I polyketide synthases are large bacterial proteins that form complexes likened to molecular assembly lines that are responsible for the production of numerous therapeutically valuable polyketide natural products.[12] Advances in microbial genetics allowed for the pioneering sequence elucidation of the erythromycin biosynthetic pathway that produces the polyketide core of the erythromycin A antibiotic.[12-14] This

initial sequence analysis, as well as the many subsequent reports following it, allowed for the characterization of the mechanism through which type I PKS biosynthesis occurs: Each PKS module is responsible for catalyzing the two-carbon elongation of a growing polyketide chain and requires a minimum of three enzymatic domains (Figure 1.2): An acyltransferase (AT) domain which selects an appropriate acyl-coenzyme A extender unit and transfers it to the acyl carrier protein (ACP) which presents the loaded extender unit to the ketosynthase (KS) domain that accepts the upstream acyl chain
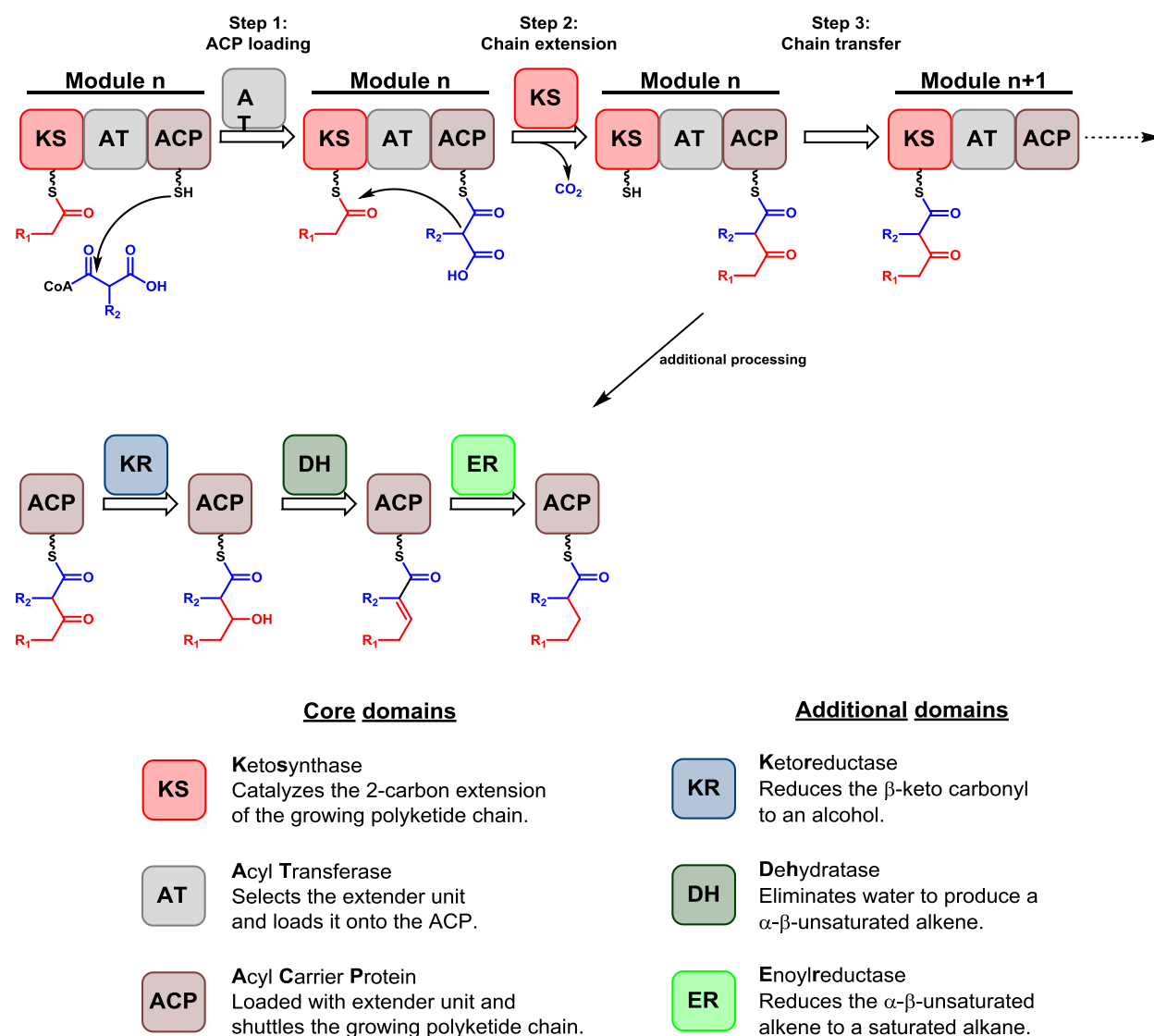
Figure 1.2 Type I polyketide synthase catalytic cycle. A malonyl CoA extender unit (blue) is incorporated into the growing nascent polyketide chain (red) through the successive action of three core domains. The extended intermediate is then available for additional reductive processing at the β-keto position by a combination of three additional reductive domains.

5

and catalyzes a decarboxylative Claisen condensation to incorporate the extender unit into the growing acyl chain.[12] Additionally, PKS modules may also contain a combination of tailoring domains including a ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) which reduce the β-keto position to a hydroxyl, alkene, or alkane, respectively. Furthermore, the terminal module often contains a thioesterase (TE) domain which is responsible for the regio- and stereoselective offloading of the acyl intermediate either as a linear acid or macrolactone.[15]

The pikromycin biosynthetic pathway: a model system

The pikromycin biosynthetic pathway (Pik) is a representative type I PKS system that has served as a workhorse for the Sherman Lab for nearly 20 years.[16] The Pik pathway is unique in its ability to produce two macrocyclic polyketides, the 12-membered ring 10-deoxymethnolide (10-dml) and the 14-membered ring narbonolide (Figure 1.3). Both macrolactone products are then further diversified by the processing of post-assembly line tailoring enzymes to yield a suite of five macrolide natural products.[17] The inherent catalytic flexibility of the Pik pathway renders it an attractive target for engineering studies. Sequencing analysis of Streptomyces venezuelae (S. Venezuelae) revealed the genetic architecture of the Pik pathway.[16] The Pik biosynthetic pathway is composed of 18 clustered genes: two ribosomal methyl transferases, four polyketide synthases, a type II thioesterase, nine genes involved in desosamine biosynthesis and attachment, a p450, and a pathway regulator.

In a landmark study, four laboratories at the University of Michigan collaborated on solving the first three-dimensional structure of an entire intact PKS module through the use of cryo-electron microscopy (cryo-EM).[18,19] In this pioneering work, Dutta et al. subjected PikAIII (Figure 1.3) to a painstakingly in-depth purification strategy followed by cryo-EM analysis and the selection of hundreds of thousands of particles to yield a final set of structures with resolutions of 7 to 10 Å. In contrast to long-standing expectations, fitting the spheroid particles to a model system derived from the crystal structures of several homologous excised domains from the DEBS biosynthetic pathway yielded a three-dimensional reconstruction of PikAIII as a homodimer aligned head-to-
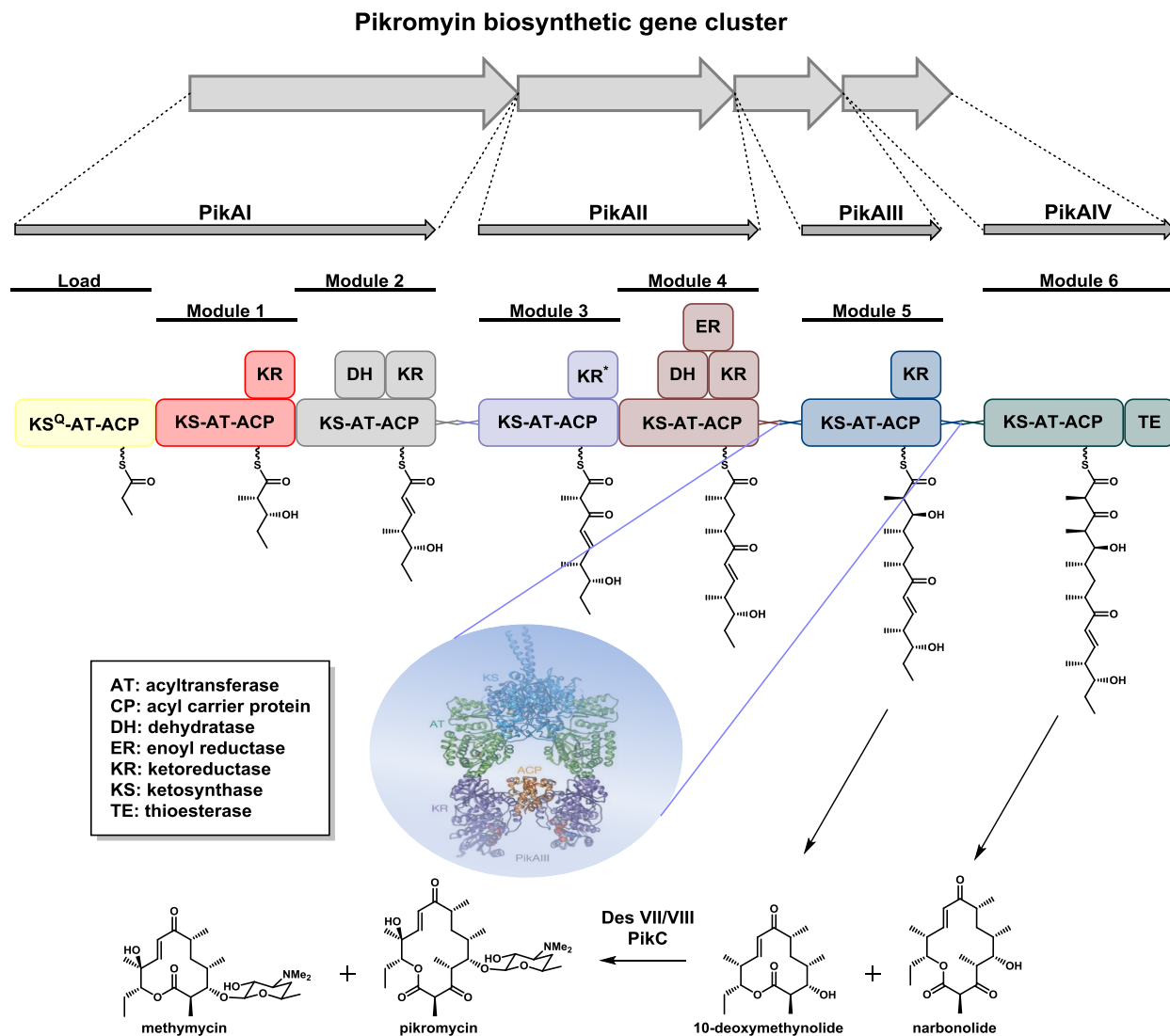
Figure 1.3 The pikromycin (Pik) biosynthetic pathway. The macrolactones 10-dml and narbonolide generated by the Pik PKS undergo further tailoring to the macrolides methymycin and pikromycin, respectively. Open reading frames are represented as arrows while enzymatic domains are represented as boxes. KS[Q], KS-like domain; KR[*], inactive KR. The three-dimensional cryo-EM structure of PikAIII is highlighted.

head and tail-to-tail in a horseshoe shape which encloses a single inner reaction chamber composed of each domain active site.[18] Furthermore, by making use of the natural substrate (Pik pentaketide) and extender unit (methylmalonyl-coenzyme A) Whicher et al. successfully determined the cryo-EM structure of PikAIII in three key biochemical states of its catalytic cycle representing a single round of polyketide chain extension and reduction[19] (Figure 1.3). Comparing the overall architecture of PikAIII in its various catalytic states provided significant insight into the dramatic conformational rearrangements that take place to allow for productive catalysis. These flagship

structural studies of PikAIII provide the first ever structural and dynamic blueprint for future PKS bioengineering efforts.

Engineering new polyketides

Despite the usefulness of polyketide natural products in medicine, the development of polyketides into new drugs is often hindered by their need for synthetic derivatization. Although many polyketide natural products possess favorable bioactivities, their development into clinically viable therapeutics often represents a significant challenge due to inherently poor pharmacological properties. Thus it is often necessary to optimize polyketide natural products for therapeutic development through the generation of analogs. However, the high structural complexity of polyketide natural products limits their chemical space accessible by synthetic chemistry and necessitates lengthy and expensive synthetic routes.[20,21] Traditionally, natural product analogs have been generated through semi-synthesis, whereby a naturally occurring molecular entity is isolated from the producing organism and further derivitized through synthetic chemistry (1, Figure 1.4). While providing notable successes, semi-synthetic strategies often require lengthy and expensive routes that are constrained by accessible functionality and reactivity. Alternatively, the collinear relationship between PKS genetic sequence and polyketide product provides an additional route to polyketide analogs.[7,12] Biosynthetic pathway engineering represents an attractive method for increasing the chemical diversity of polyketide natural products whereby novel natural product analogs are generated by manipulation of the enzymes responsible for their biosynthesis.[22] Mutasynthesis uses synthetic starter units that chemically complement early pathway knockouts in native or heterologous producers as a means to generate natural product analogs[23,24] (2, Figure 1.4). Combinatorial biosynthesis generates polyketide chemical diversity by engineering of the PKS biosynthetic enzymes responsible for their production.[6,25,26] The strategies utilized in combinatorial biosynthesis include alterations at both the individual domain level to modify the functional groups present on the polyketide backbone and at the module or subunit level to change the overall chain
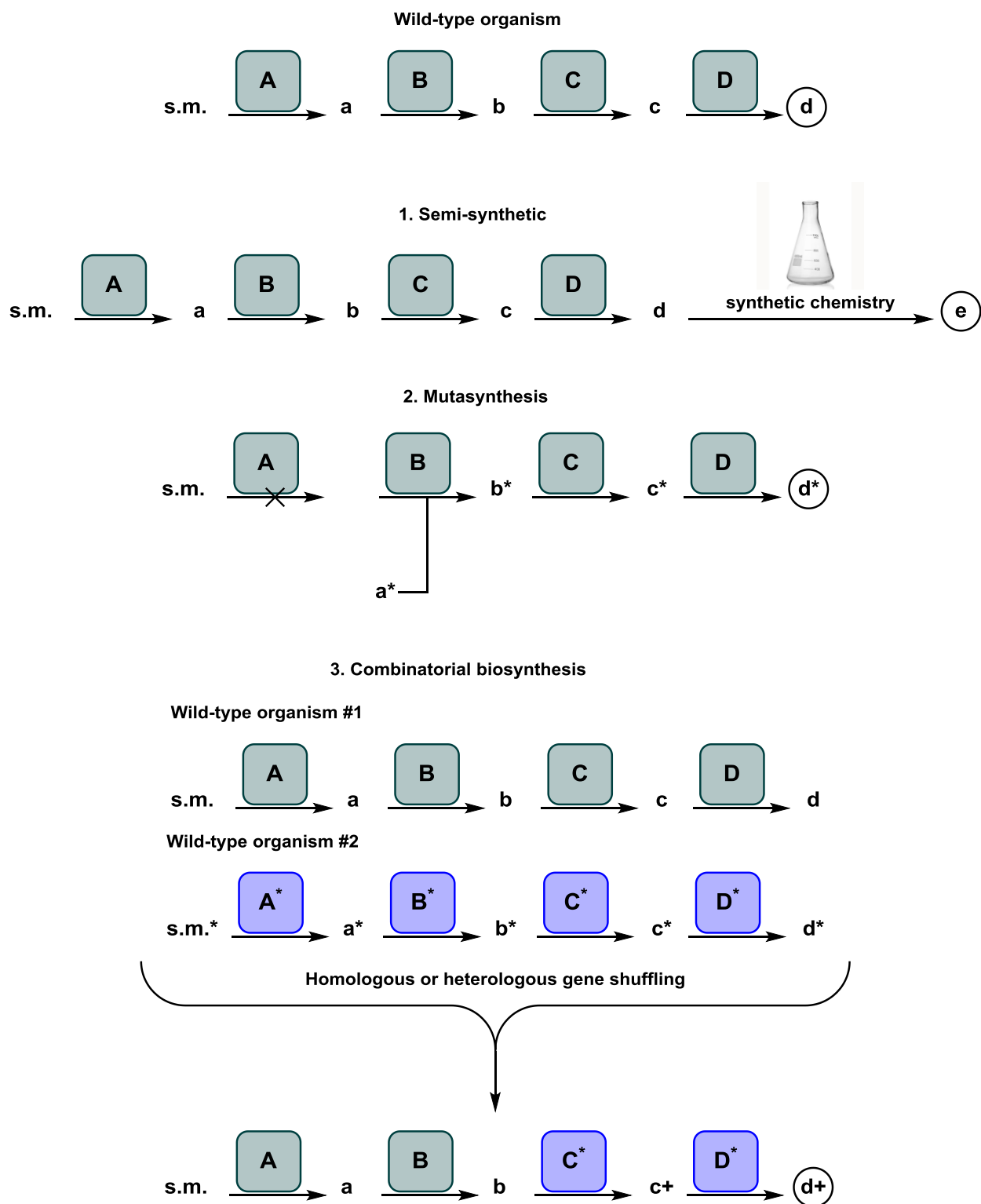
Figure 1.4 Laboratory methods for generating polyketide analogs. Boxes represent enzymatic domains while letters represent chemical intermediates. Figure adapted from reference 23.

length (3, Figure 1.4). New chemical diversity is achieved through a number of different forms of PKS bioengineering that alter the length of the polyketide chain (determined by the number of modules that comprise the polyketide synthase), the choice of primer and extender units (controlled by the AT domains), the degree of reduction of the polyketide backbone (determined by the set of enzyme domains present in each module), and the stereochemistry at centers carrying alkyl and hydroxyl substituents (controlled by enzyme domains that are responsible for generating the stereocenter in question).

Although combinatorial biosynthesis has been successfully pursued by many research groups for the production of polyketide analog libraries, these bioengineering attempts are often accompanied by a concomitant decrease in product yields or fail to produce new molecules entirely.[27-29] The mechanistic rational for the diminished activity observed from PKS engineering attempts remains unclear. To date, the majority of PKS engineering studies have targeted genetic modifications in early pathway PKS modules.[27,28,30-33] As a result it is unknown whether the attenuated product formation results from the engineered module or the downstream modules that must accept and process the unnatural intermediate. A significant hindrance in the mechanistic troubleshooting of these PKS engineering studies is the in vivo context in which they were performed which prevents the isolation and characterization of shunt products arising from catalytic points of failure. Thus for a more general framework for rational engineering of efficient PKSs, it is critical moving forward for more targeted investigations to probe the catalytic details of individual domains.

Thesis outline

Given the interest in generating natural product analogs to support drug discovery efforts, and the significant synthetic challenges encountered in this pursuit, we sought to further our knowledge of PKS structure and function in efforts to harness them as biocatalysts for the biosynthetic production of polyketide analogs. Due to the innate catalytic flexibility of the pikromycin biosynthetic machinery, we chose to utilize Pik as a model system for performing these investigations. The studies presented in this dissertation aim to unveil the catalytic parameters that govern PKS function by directly

probing PKS enzymatic domains and modules in vitro. To begin, the studies in Chapter II focus on identifying catalytic bottlenecks in the processing of unnatural substrates through the use of in vitro methods that simulate combinatorial biosynthesis. Taking advantage of previously optimized in vitro PKS biochemistry conditions, we employed a series of Pik pentaketide analogs that mimic unnatural intermediates arising from upstream PKS engineering modifications, and use them to probe the substrate flexibility of PikAIII-TE. Analysis of the resulting shunt products indicated the TE domain as a key catalytic bottleneck in the processing of unnatural substrates. To test this hypothesis, we next generated a series of hybrid type I PKS modules containing TE domains from three homologous pathways and used these hybrid PKS modules to demonstrate the critical role of the TE domain in the context of PKS engineering. Furthermore, we next applied these catalytic insights towards the chemoenzymatic production of 14-membered macrolactone analogs. The advances made in Chapter II have been partially reported in references 35.

Chapter III builds upon the TE gatekeeper hypothesis generated in Chapter II by focusing studies on the TE directly as an excised domain. Prior to enzymatic analysis, it was necessary to stabilize the Pik TE native substrate (Pik hexaketide). This was accomplished through the use of two distinct protection strategies, including a photocleavable 2-nitrobenzyloxymethyl ether protecting group. These advances in substrate stabilization allowed for the evaluation of a panel of substrate esters in order to optimize diffusive loading onto Pik TE. This evaluation revealed a dramatic variation in catalytic processing with each protected hexaketide. Cloning, expression, and purification of a series of TE domain mutants allowed for probing the catalytic parameters required for TE catalyzed macrolactonization reactions and provided the serendipitous discovery of a Pik TE mutant (S148C) that is a more efficient macrolactonization catalyst. Finally, incorporation of this engineered TE into PikAIII-TE relieved the catalytic bottleneck observed in Chapter I, allowing for the processing of two unnatural macrolactones. The findings from this chapter have been reported in references 34 and 36.

Chapter IV concludes this dissertation with a thorough investigation into the increased catalytic efficiency observed with Pik TE S148C. Working towards a structural

analysis of Pik TE, a double mutant strategy was successfully employed to stably capture the protein as an acyl-enzyme intermediate. However, all attempts to crystalize this labeled mutant to date have failed to yield suitable material for X-ray crystallography. Thus we pursued an alternative strategy using computational techniques and the available three-dimensional structures of Pik TE as an apoprotein to model the catalytic steps post acyl-enzyme formation. Using a combination of molecular dynamics (MD) simulations along with quantum mechanical (QM) calculations, we provide further insight into the catalytic parameters governing macrocyclization in Pik TE and describe a mechanistic rational for the previously observed strict substrate stereospecificity. We next perform analogous serine to cysteine mutations in homologous type I TE domains to preliminarily determine the generality of this mutation, and discuss the biosynthetic implications of naturally occurring TE domains with a native cysteine nucleophile. The findings from this chapter have been reported in reference 36.

References

(1)   Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2016**, *79*, 629.

(2)   Koehn, F. E.; Carter, G. T. *Nat Rev Drug Discov* **2005**, *4*, 206.

(3)   Henkel, T.; Brunne, R. M.; Müller, H.; Reichel, F. *Angew. Chem. Int. Ed.* **1999**, *38*, 643.

(4)   Feher, M.; Schmidt, J. M. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 218.

(5)   Evans, B. E.; Rittle, K. E.; Bock, M. G.; DiPardo, R. M.; Freidinger, R. M.; Whitter, W. L.; Lundell, G. F.; Veber, D. F.; Anderson, P. S.; Chang, R. S. L.; Lotti, V. J.; Cerino, D. J.; Chen, T. B.; Kling, P. J.; Kunkel, K. A.; Springer, J. P.; Hirshfield, J. *J. Med. Chem.* **1988**, *31*, 2235.

(6)   Weissman, K. J. *Nat. Prod. Rep.* **2016**, *33*, 203.

(7)   Hertweck, C. *Angew. Chem. Int. Ed.* **2009**, *48*, 4688.

(8)   Driggers, E. M.; Hale, S. P.; Lee, J.; Terrett, N. K. *Nat. Rev. Drug. Discov.* **2008**, *7*, 608.

(9)   Parenty, A.; Moreau, X.; Niel, G.; Campagne, J. M. *Chem. Rev.* **2013**, *113*, PR1.

(10)   Woodward, R. B.; Au-Yeung, B. W.; Balaram, P.; Browne, L. J.; Ward, D. E.; Au-Yeung, B. W.; Balaram, P.; Browne, L. J.; Card, P. J.; Chen, C. H. *J. Am. Chem. Soc.* **1981**, *103*, 3213.

(11)   Hoffmann, R. W. *Angew. Chem. Int. Ed.* **2000**, *39*, 2054.

(12)   Staunton, J.; Weissman, K. J. *Nat. Prod. Rep.* **2001**, *18*, 380.

(13)   Cortes, J.; Haydock, S. F.; Roberts, G. A.; Bevitt, D. J.; Leadlay, P. F. *Nature* **1990**, *348*, 176.

(14) Donadio, S.; Staver, M.; McAlpine, J.; Swanson, S.; Katz, L. *Science* **1991**, *252*, 675.

(15) Du, L.; Lou, L. *Nat. Prod. Rep.* **2010**, *27*, 255.

(16) Xue, Y.; Zhao, L.; Liu, H.-w.; Sherman, D. H. *Proceedings of the National Academy of Sciences* **1998**, *95*, 12111.

(17) Kittendorf, J. D.; Sherman, D. H. *Bioorg. Med. Chem.* **2009**, *17*, 2137.

(18) Dutta, S.; Whicher, J. R.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Congdon, G. R.; Narayan, A. R.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. *Nature* **2014**, *510*, 512.

(19) Whicher, J. R.; Dutta, S.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Dosey, A. M.; Narayan, A. R.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. *Nature* **2014**, *510*, 560.

(20) Lam, K. S. *Trends Microbiol.* **2007**, *15*, 279.

(21) Maier, M. E. *Org. Biomol. Chem.* **2015**, *13*, 5302.

(22) Weissman, K. J.; Leadlay, P. F. *Nat. Rev. Microbiol.* **2005**, *3*, 925.

(23) Kirschning, A.; Taft, F.; Knobloch, T. *Org. Biomol. Chem.* **2007**, *5*, 3245.

(24) Weist, S.; Sussmuth, R. D. *Appl. Microbiol. Biotechnol.* **2005**, *68*, 141.

(25) Kittendorf, J. D.; Sherman, D. H. *Curr. Opin. Biotechnol.* **2006**, *17*, 597.

(26) Wong, F. T.; Khosla, C. *Curr. Opin. Chem. Biol.* **2012**, *16*, 117.

(27) McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Betlach, M.; Ashley, G. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 1846.

(28)   Xue, Q.; Ashley, G.; Hutchinson, C. R.; Santi, D. V. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 11740.

(29)   Menzella, H. G.; Reid, R.; Carney, J. R.; Chandran, S. S.; Reisinger, S. J.; Patel, K. G.; Hopwood, D. A.; Santi, D. V. *Nat. Biotechnol.* **2005**, *23*, 1171.

(30)   Jacobsen, J. R.; Keatinge-Clay, A. T.; Cane, D. E.; Khosla, C. *Bioorg. Med. Chem.* **1998**, *6*, 1171.

(31)   Leaf, T.; Cadapan, L.; Carreras, C.; Regentin, R.; Ou, S.; Woo, E.; Ashley, G.; Licari, P. *Biotechnol. Prog.* **2000**, *16*, 553.

(32)   Kinoshita, K.; G Williard, P.; Khosla, C.; Cane, D. E. *J. Am. Chem. Soc.* **2001**, *123*, 2495.

(33)   Harvey, C. J.; Puglisi, J. D.; Pande, V. S.; Cane, D. E.; Khosla, C. *J. Am. Chem. Soc.* **2012**, *134*, 12259.

(34)   Hansen, D. A.; Koch, A. A.; Sherman, D. H. *J. Am. Chem. Soc.* **2015**, *137*, 3735.

(35)   Hansen, D. A.; Koch, A. A.; Sherman, D. H. *J. Am. Chem. Soc.* **2017**, *139*, 13450.

(36)   Koch, A. A.; Hansen, D. A.; Shende, V. V.; Furan, L. R.; Houk, K. N.; Jiménez-Osés, G.; Sherman, D. H. *J. Am. Chem. Soc.* **2017**, *139*, 13456.

Chapter II

Investigations into late stage PKS engineering

DH contributed to Schemes 2.1-2.1 and Figure 2.2

AK contributed to all others.

Introduction

The biocatalytic production of polyketide analogs through genetic engineering of PKS biosynthetic pathways is a long-standing goal.[1-4] Several successful examples have been reported by Kosan Biosciences, highlighting the ability to generate libraries of polyketide analogs through modular PKS engineering.[5-7] These studies utilized canonical combinatorial biosynthesis strategies by incorporating AT and KR domain substitutions into modules 2, 5, and 6 of the 6-deoxyerythronolide B (DEBS) biosynthetic pathway followed by overexpression in heterologous *Streptomyces* or *E. coli* hosts to generate a suite of 6-deoxyerythronolide (6-dEB) analogs. While these PKS engineering studies introduced an impressive array of functional and skeletal diversity into the 6-dEB macrolactone core, they also highlighted the significant shortcomings with this approach as the yields of the 14-membered macrolactone analogs dropped precipitously with the highest being only 3.5% relative to 6-dEB production.[5]

Furthermore, a perhaps more significant limitation of these studies is the in vivo experimental design which precludes the identification of catalytic bottlenecks due to the difficulties encountered with isolating shunt products from the complex fermentation medium. Thus the biochemical basis for the decreased product titers or entirely failed engineered PKS pathways remains elusive. These engineering strategies often times target early pathway domains or modules[5,6,8-11] and the resulting unnatural intermediate must then be effectively processed by all downstream enzymatic domains for the successful production of the anticipated natural product analog. Therefore, in the absence of precise mechanistic information, it is unclear whether the attenuated product formation stems from the engineered module or the downstream modules (or both) that must accept and process the resulting unnatural intermediates. This uncertainty highlights the need for more targeted investigations into PKS catalysis to provide a more general framework for the rational engineering of efficient PKS pathways.

Towards this goal, investigations in the Sherman lab have evolved over the past two decades from in vivo studies relying on genetic alterations to targeted in vitro experiments directly probing the mechanistic underpinning of PKS catalytic domains. A particularly productive strategy has been the application of synthetic chemistry to

provide full length chain elongation intermediates for probing PKS machinery with its native substrate. Prior to this advancement, the more easily synthetically achievable diketide and triketide *N*-acetylcysteamine thioester substrates were employed for studying PKS catalysis in vitro. While the use of these simpler substrates has provided many fundamental insights into the nature of PKS catalysis, it was quickly noted that their relevance in determining the substrate selectivity of late stage PKS machinery was inadequate as these simplified substrates do not adequately represent the natural chain elongation intermediates processed by the enzymatic domains in question.[12] Although it requires a significant synthetic investment, it is essential to study these enzymes with fully elaborated polyketide chain elongation intermediates to adequately understand the complex kinetic and mechanistic parameters governing PKS catalysis. Recent work in the Sherman lab has focused especially on late stage biosynthetic enzymes as they are responsible for the final processing of the most mature biosynthetic intermediates and thus may possess the strictest level of substrate specificity.[12] Efforts by Mortison et al[13]. investigating the substrate flexibility of the homologous penultimate modules from the Pik and DEBS pathways highlighted the high level of substrate specificity inherent in these final PKS modules. Consistent with previous reports, PikAIII-TE paired with its native substrate Pik pentaketide was able to efficiently generate 10-dml, while DEBS Mod5-TE paired with its native substrate DEBS pentaketide produced the anticipated 12-membered macrolactone. However, when tested for the ability to process the non-native yet similar pentaketides from the noncognate pathways, no macrolactone products were identified, indicating a low level of substrate flexibility within these late stage PKS modules.

Probing PikAIII-TE with unnatural pentaketides

Developing this paradigm further, Hansen et al. pursued a substrate-centric strategy to "simulate late stage PKS engineering" which provided increased mechanistic insight into PKS catalysis with unnatural substrates.[14] In this approach, two classes of Pik pentaketides were generated; aliphatic truncations and diastereomers that mimic the upstream engineering of early pathway AT or KR domains, respectively to probe the

substrate flexibility of the late stage module PikAIII-TE.[12,13,15] PikAIII-TE displayed moderate flexibility for truncated substrates (**2-5,** Scheme 2.1) and was able to process C-10 desmethyl **2** with near wild type efficiency (56% yield), followed by C-8 desmethyl **3** with a 39% yield. However, C-8,10 didesmethyl **4** and C-9 desethyl **5** afforded poor conversion to the expected products (17% and 9% yield, respectively). PikAIII-TE failed to yield detectable macrocyclic products entirely from diastereomers **11-13** designed to interrogate the PKS flexibility toward substrates bearing unnatural C-8 and C-9 stereocenters (Scheme 2.2). HPLC purification of the isolable products from these reactions followed by ^{1}H NMR analysis revealed small quantities of extended material off-loaded from the PKS assembly line as linear shunt products (**14**-**19**, Scheme 2.2).
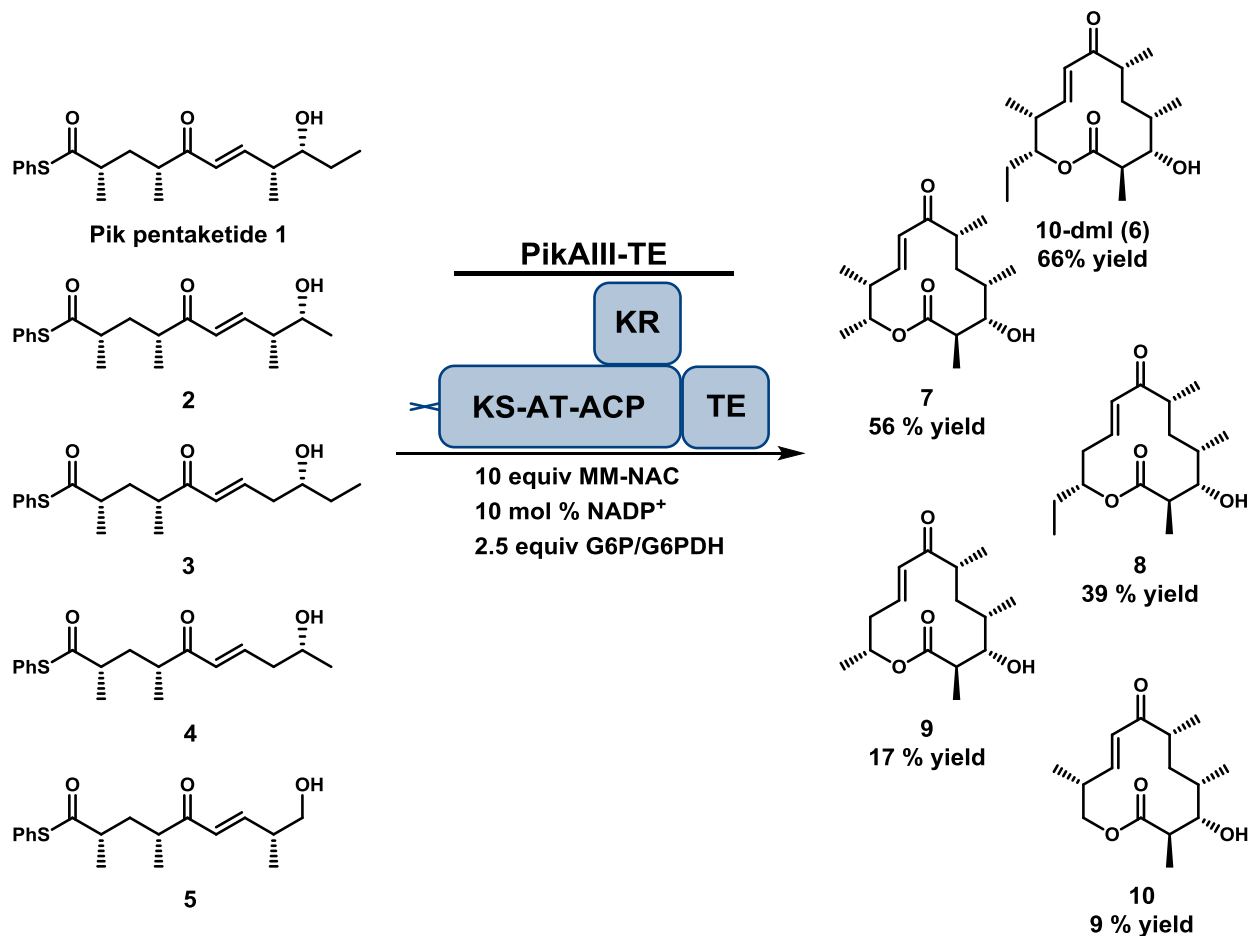
Scheme 2.1 Evaluation of PikAIII-TE with a panel of truncated pentaketides. These reactions were performed to interrogate the substrate flexibility of PikAIII-TE for pentaketides that simulate the upstream engineering of early pathway AT domains Enzymatic reaction conditions: 4 mM Pik pentaketide, 40 mM (10equiv) MM-NAC, 20 mM (5 equiv) 2-vinylpyridine, 0.4 mM (10 mol%) NADP+, 10 mM (2.5 equiv) glucose-6-phosphate, glucose-6-phosphate dehydrogenase (2 units/mL), 4 μM (0.1 mol%) cell-free PikAIII-TE, 8 hours, stationary, RT.
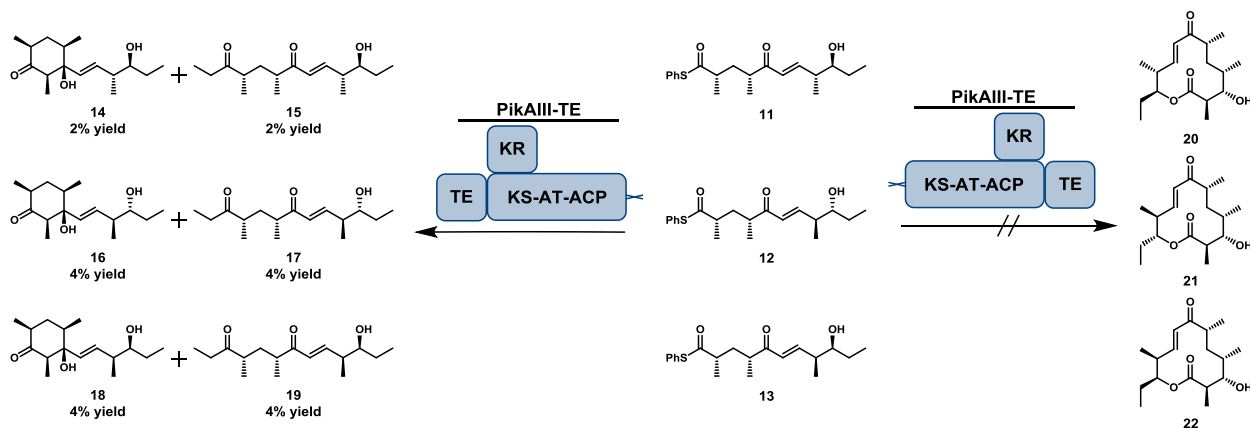
Scheme 2.2 Reaction of PikAIII-TE with a panel of stereoisomer pentaketides. These reactions were performed to interrogate the substrate flexibility of PikAIII-TE for pentaketides that simulate the upstream engineering of early pathway KR domains Enzymatic reaction conditions: 4 mM Pik pentaketide, 40 mM (10 equiv) MM-NAC, 20 mM (5 equiv) 2-vinylpyridine, 0.4 mM (10 mol %) NADP+, 10 mM (2.5 equiv) glucose-6-phosphate, glucose-6-phosphate dehydrogenase (2 units/mL), 4 μM (0.1 mol %) cell-free PikAIII-TE, 8 hours, stationary, RT.

The severe drop in product yields from the native pentaketide **1** to the truncated substrates **4** and **5** was surprising as these modifications are distal from the location of the Claisen condensation and β-keto reduction catalyzed by the KS and KR domains, respectively. We interpreted this to suggest that the TE domain responsible for the release of the macrolactone from the PKS assembly line could be responsible for the compromised catalytic efficiency with the truncated substrates **2-5**. The total failure in processing of diastereomer **11,** which differs from the native pentaketide **1** only by epimerization of the C-9 hydroxyl group which serves as the nucleophile during macrolactonization, further implicated the TE domain as the catalytic point of failure as previous studies from the DEBS TE have revealed a strict stereospecificity in the macrolactonization of unnatural substrates.[16] NMR and MS analysis of the reaction products from a large scale reaction with **11** produced the shunt products **14** and **15** which were successfully extended by the KS domain and thus we hypothesized that these products resulted from stalling and premature off-loading from the PKS via hydrolysis and subsequent decarboxylation (Figure 2.1). Isolation of these extended yet linear off loaded intermediates indicates the TE domain as the dominant catalytic bottleneck to the formation of macrocyclic products. This hypothesis was then confirmed by synthesis of the corresponding C-11-epimerized hexaketide and subsequent probing

20
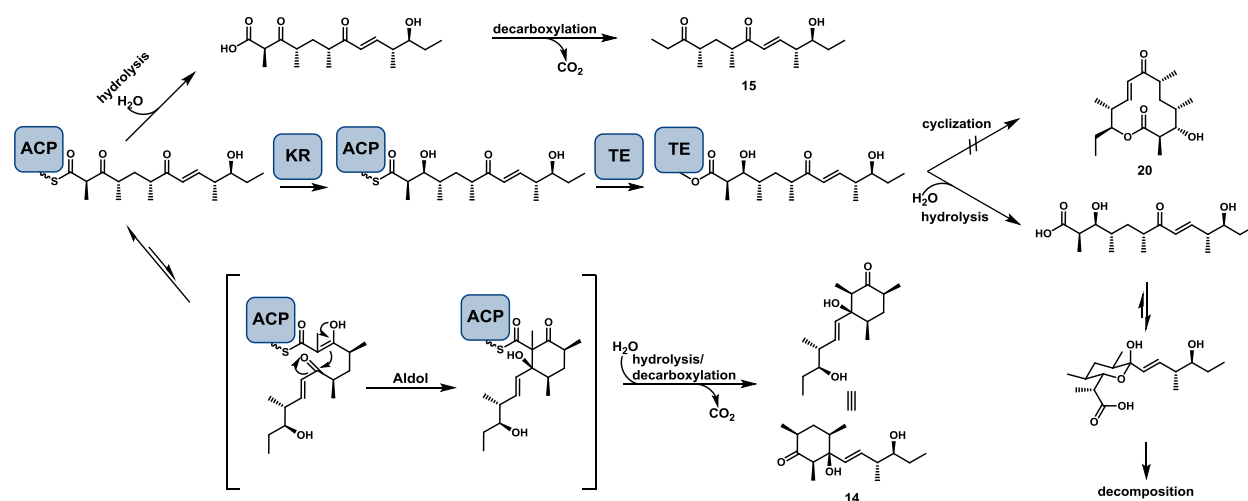
Figure 2.1 Proposed shunt pathways when macrocyclization is compromised. Release mechanisms of unnatural Pik chain elongation intermediates from PikAIII-TE and the subsequent degradative pathways when TE catalyzed cyclization is impaired.

of the Pik TE domain to provide direct evidence for the strict stereospecificity of Pik TE (see Chapter III).

Generation of hybrid TE PKS modules

To investigate the role of the TE domain in engineered PKS modules, we generated a series of hybrid type I PKS modules by combining modules from the pikromycin (Pik), erythromycin (DEBS), and juvenimycin (Juv) biosynthetic pathways with each corresponding TE domain. The TE domains from these 3 related pathways generate macrolactones that vary in size from 12 to 16-membered rings and contain unique chemical functionality (Figure 2.2). By generating this suite of hybrid TE PKS modules, we were able to test both the ability of each TE domain to function in an engineered PKS module and also probe PKS modules with both natural and unnatural substrates to ascertain the intrinsic substrate flexibility of each TE domain for macrolactonization. The TE domains from the Pik[17] and DEBS[18] biosynthetic pathways have been cloned and expressed as excised domains previously. In both cases, the C-terminus of the TE contained the natural stop codon from the corresponding module while the N-terminus contained a portion of the ACP-TE interdomain linker region. Regarding Pik TE, the length of the linker region included in the TE was originally determined[17] empirically by optimizing the protein for in vitro activity. Furthermore,

21

10-dml (6)   narbonolide (23)   6-dEB (24)   tylactone (25)
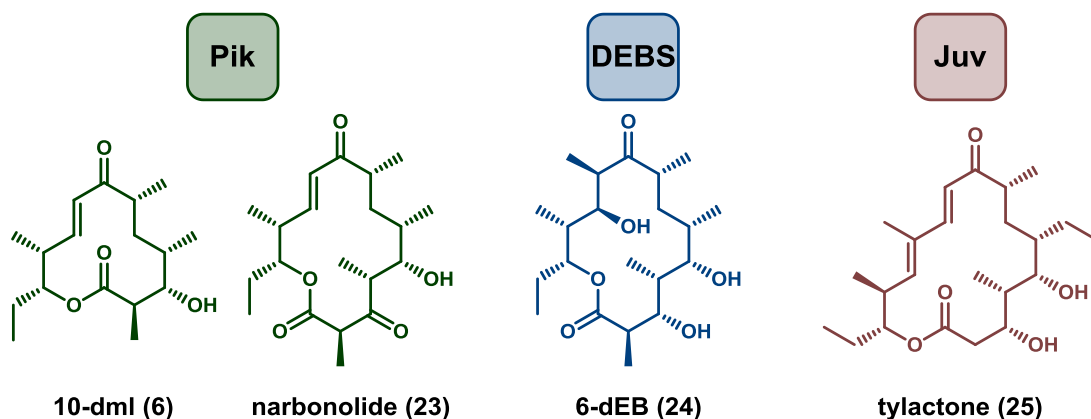
Figure 2.2 The thioesterase domains utilized in this study and the macrolactones they produce.

fusing the optimized Pik TE downstream of the Pik Mod5 ACP domain to generate PikAIII-TE[15] resulted in a hybrid module capable of producing 10-dml (**6**)[12,13] from the Pik pentaketide **1**.

For generating the hybrid TE modules in this study, we followed a similar strategy for each TE domain by including the natural stop codon from the C-terminus of each respective module, and determined the optimal N-terminal boundary through sequence alignment with the previously optimized Pik TE (Figure 2.3). Pik Mod5 and DEBS Mod5 have been previously constructed[13,15,19] as hybrid fusion proteins with the Pik and DEBS TE domains, respectively, and were modified slightly in this study to include the full selection of the aforementioned TE domains. Pik Mod6, DEBS Mod6, Juv Mod6, and Juv Mod7 hybrids were generated by inserting a restriction site at the 3' end of the
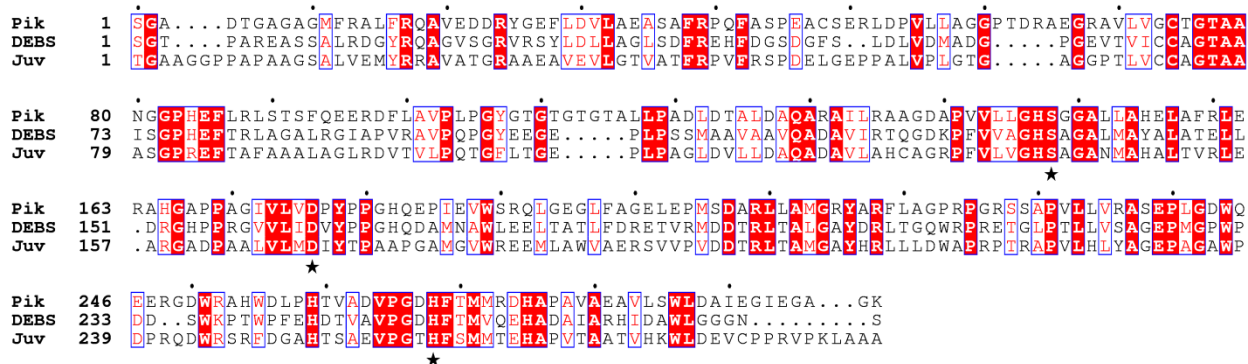
Figure 2.3 Sequence alignment of the Pik, DEBS, and Juv thioesterase domains. The Pik and DEBS TEs have been cloned and described previously. The N-terminal portion of the Juv TE domain was determined in this study by sequence alignment with Pik and DEBS. Sequences were aligned using T-coffee[20] and rendered with ESPript[21] The stars denote the positions of residues in the catalytic triad.

22

alignment consensus for each antiSMASH[22] annotated ACP domain to allow for the insertion of the respective non-native TE domains (Figure 2.4). Following the cloning of
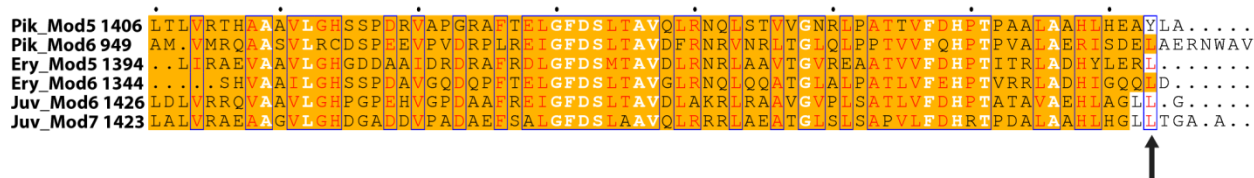
Figure 2.4 Determination of the post-ACP restriction site for incorporating non-native TE domains. The location (black arrow) was chosen from alignment of the antiSMASH[22] annotated ACP domain regions (orange highlight) for each PKS module. Sequences were aligned using T-coffee[20] and rendered with ESPript.[21]

each hybrid, we next performed expression of each His$_6$-tagged protein in the E. *coli* strain BAP1[23], resulting in the production of 13 new hybrid TE modules. A single pass purification scheme was chosen to minimize the length of time each protein was processed since previous work has shown the most reproducible in vitro activity is achieved with short purification times.[24] Purification by Ni-NTA affinity chromatography provided each protein in yields (Table 2.1) and purities (Figure 2.5) sufficient for enzymatic analysis. Although the Juv Mod6 hybrids contained significant contaminating species within the purified protein fractions (lanes 13-16, Figure 2.5), previous work[25] had shown Juv Mod6 retains high catalytic efficiency as a cell-free lysate, thus alleviating our concern over the purity of these proteins.

Role of the TE domain in engineered PKS modules

With a library of hybrid TE PKS modules in hand, we set out to assess the substrate flexibility of each TE domain and examine the role of the TE in each engineered system. Using previously optimized in vitro conditions[24], we initially tested Pik Mod5 (the native module for processing Pik pentaketide **1**, DEBS Mod5, DEBS Mod6, and Juv Mod6 hybrid modules for the ability to generate 10-dml (**6**) using the Pik pentaketide **1**[24] as a substrate. This initial scheme was chosen due to our access to the authentic standards for product formation, as each of these hybrid modules share the same domain composition, and thus should produce either 10-dml (**6**) or 3-keto-10-dml **26**. HPLC quantification of the reaction products answered two key questions: First, TE
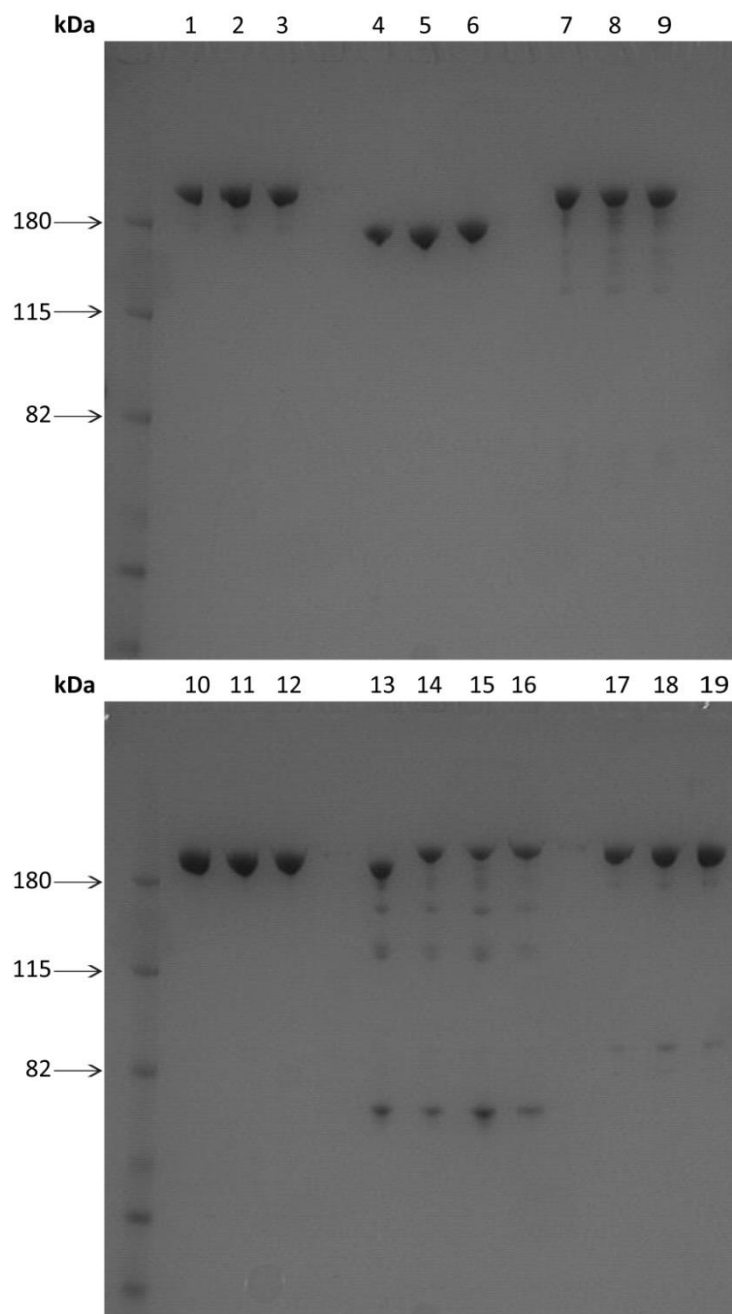
Figure 2.5 PAGE gel of hybrid TE PKS modules. NuPAGE® Bis-Tris Mini, MOPS running buffer, BenchMark™ Pre-stained Protein Ladder. Lanes: **1**. Pik Mod5-DEBS TE (186 kDa), **2**. Pik Mod5-Juv TE (186 kDa), **3**. Pik Mod5-Pik TE (187 kDa), **4**. Pik Mod6-DEBS TE (138 kDa), **5**. Pik Mod6-Juv TE (139 kDa), **6**. Pik Mod6-Pik TE (143 kDa), **7**. DEBS Mod5-DEBS TE (186 kDa), **8**. DEBS Mod5-Juv TE (187 kDa), **9**. DEBS Mod5-Pik TE (188 kDa), **10**. DEBS Mod6-DEBS TE (178 kDa), **11**. DEBS Mod6-Juv TE (178 kDa), **12**. DEBS Mod6-Pik TE (179 kDa), **13**. Juv Mod6 (165 kDa), **14**. Juv Mod6-DEBS TE (187 kDa), **15**. Juv Mod6-Juv TE (188 kDa), **16**. Juv Mod6-Pik TE (189 kDa), **17**. Juv Mod7-DEBS TE (186 kDa), **18**. Juv Mod7-Juv TE(187 kDa), **19**. Juv Mod7-PikTE (188 kDa).

Module	TE	Lane number	MW (kDa)	Purification yield (mg)
Pik Mod5	DEBS	1	186	41
	Juv	2	186	40
	Pik	3	187	37
Pik Mod6	DEBS	4	138	39
	Juv	5	139	34
	Pik	6	143	26
DEBS Mod5	DEBS	7	186	41
	Juv	8	187	27
	Pik	9	188	32
DEBS Mod6	DEBS	10	178	24
	Juv	11	178	22
	Pik	12	179	22
Juv Mod6	---	13	165	35
	DEBS	14	187	26
	Juv	14	188	31
	Pik	16	189	30
Juv Mod7	DEBS	17	186	26
	Juv	18	187	20
	Pik	19	188	21

Table 2.1 Yields from the Ni-NTA column purification of the hybrid TE PKS modules.

domains from homologous type I PKS pathways can function in a combinatorial manner, and secondly, proper pairing of the TE domain to the incoming linear intermediate is critical for achieving productive catalysis. Indeed, even in reactions pairing all non-native core PKS domains, we achieved significant conversions to 10-dml (**6**) if the Pik TE was in place (Table 2.2). These results demonstrate that the KS-AT-ACP and KR domains from non-cognate PKS modules in our system maintained a suitable level of catalytic function when the downstream TE domain is correctly matched to the incoming unnatural intermediate. This is evidenced by the 21, 17, 41, and 75% conversions from the reactions containing, DEBS Mod5-Pik TE, DEBS Mod6-Pik TE, Juv Mod6-Pik TE, Pik Mod5-Pik TE, respectively (Table 2.2). These results also indicate that the DEBS and Juv TE domains are poorly suited for generating 12-membered macrolactones as the percent conversions dropped rapidly for hybrid modules paired with these TE domains. Thus, we interpret these stark variations achievable through replacement of a single domain to indicate that the TE domain is acting as a key gatekeeper in engineered PKS reactions.

Table 2.2 Screening of hybrid TE modules with Pik pentaketide. Enzymatic reaction conditions: sodium phosphate buffer (400 mM, pH = 7.2), 0.5 mM Pik pentaketide, 10 mM (20 equiv) MM-NAC, 4 mM (8 equiv) 2-vinylpyridine, 0.25 mM (50 mol%) NADP$^+$, 1.25 mM (2.5 equiv) glucose-6-phosphate, glucose-6-phosphate dehydrogenase (2 units/mL), 2 μM (0.4 mol %) PKS module, 4 hours, stationary, RT. Conversion to **6** (R$_1$=OH, R$_2$=H) or **26** (R$_1$=R$_2$=O) was monitored by HPLC with data represented as the mean ± standard deviation where n = 3. Trace=detected by LC-HRMS but below the detection limit of HPLC. ND= not detected.

Module	TE	Conversion to 6 (%)	Conversion to 26 (%)
DEBS Mod5	DEBS	1.2 ± 0.2	trace
	Juv	ND	ND
	Pik	21 ± 0.7	trace
DEBS Mod6	DEBS	6.6 ± 0.5	ND
	Juv	ND	ND
	Pik	17 ± 0.5	6.6 ± 0.1
Juv Mod6	DEBS	2.3 ± 0.4	trace
	Juv	ND	ND
	Pik	41 ± 2.8	trace
Pik Mod5	DEBS	2.9 ± 0.2	trace
	Juv	ND	ND
	Pik	75 ± 5.6	trace

In effort to further our TE gatekeeper findings, we next probed the hybrid TE modules with the Pik hexaketide **27** (see Chapter III for details) to determine the extent to which the modules could accept and process the hexaketide to a 14-membered macrolactone. LC-HRMS screening of reactions containing 2-nitrobenzyloxymethyl (NBOM) ether protected thiophenol activated hexaketide[26] **27** with our panel of PKS modules revealed that DEBS Mod6, Juv Mod6, and Pik Mod6 were able to generate 14-membered macrolactones. A large scale reaction containing DEBS Mod6-DEBS TE provided sufficient quantities of **28** for structural characterization. HPLC quantification of the reaction products indicated that our panel of PKS modules followed a similar reactivity trend as with the pentaketide (Table 2.3). Consistent with our previous report[26] Pik Mod6 is able to efficiently process the Pik hexaketide to narbonolide (**23**), however

The figure shows substrate **27** (a phenyl thioester with OH and O groups) converting via PKS module and Pik hexaketide to product **23 or 28** (a macrolactone with R₁ and R₂ substituents).

Substrate labeled **27**. Product labeled **23 or 28**. Arrow labeled "PKS module" / "Pik hexaketide".
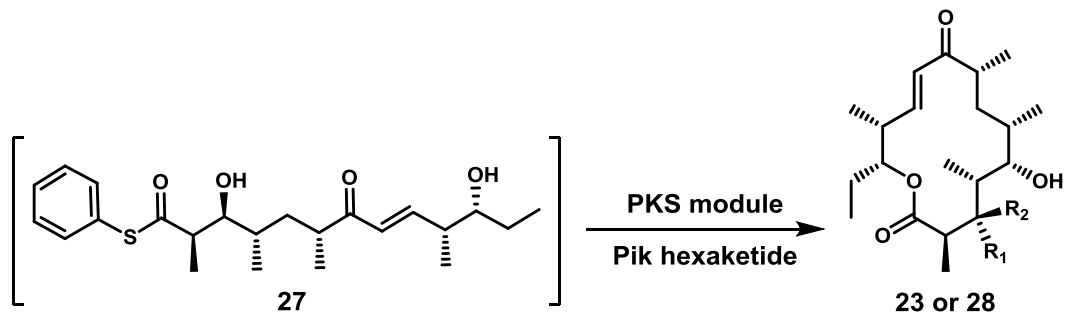
Module	TE	Conversion to 23 (%)	Conversion to 28 (%)
DEBS Mod6**	DEBS	34 ± 4.0	trace
	Juv	9.6 ± 1.0	ND
	Pik	3.5 ± 0.1	22 ± 0.7
Juv Mod6**	DEBS	trace	ND
	Juv	ND	ND
	Pik	ND	trace
Pik Mod6*	DEBS	ND	trace
	Juv	ND	ND
	Pik	ND	24 ± 0.7

Table 2.3 Evaluation of TE hybrids with Pik hexaketide. **27** is generated in situ by photolysis of the 2-nitrobenzyloxymethyl ether (NBOM) protected native hexaketide[22]. Enzymatic reaction conditions: (*) sodium phosphate buffer (400 mM, pH = 7.2), 1 mM Pik hexaketide, 20 mM (20 equiv) MM-NAC, 8 mM (8 equiv) 2-vinylpyridine, 1 mM sodium metabisulfite and 25 mM ascorbic acid (**) and 0.5 mM (50 mol%) NADP$^+$, 2.5 mM (2.5 equiv) glucose-6-phosphate, glucose-6-phosphate dehydrogenase (2 units/mL), 2.5 μM (0.25 mol %) PKS module, 4 hours, stationary, RT. Conversion to **23** (R$_1$=OH, R$_2$=H) or **28** (R$_1$=R$_2$=O) was monitored by HPLC with data represented as the mean ± standard deviation where n = 3. Trace=detected by LC-HRMS but below the detection limit of HPLC. ND= not detected.

substitution of the TE domain to either the DEBS or Juv TE causes a precipitous drop in conversions, which we interpret to indicate the DEBS and Juv TE domains require the C-3 reduced intermediate for cyclization (vide infra). DEBS Mod6 when paired with either the DEBS or Juv TE was able to process the hexaketide to 3-hydro-narbonolide **28** at 34 and 9.6% conversions, respectively. However, DEBS Mod6-Pik TE produces predominantly the unreduced narbonolide (**23**), indicating that fusion with the Pik TE interferes with the native catalytic cycle. It is unclear at this time what the mechanism for this switch in product formation stems from. We note that the native Pik substrate is unreduced at the C-3 position and the Pik TE may have a higher binding affinity for the ACP-tethered intermediate prior to β-keto reduction thus leading to premature cyclization. Reactions containing Juv Mod6 with the DEBS and Pik TE yielded trace quantities of **23** and **28**, respectively, indicating that Juv Mod6 does not efficiently

accept or process the Pik hexaketide. In contrast with the results from our reactions with **1** (Table 2.2), DEBS TE retained a significant level of reactivity when generating the 14-membered macrolactones. This is perhaps unsurprising as the native product from the DEBS PKS is the 14-membered 6-dEB (**24**, Figure 2.2) and the macrolactones **23** and **28** are similar in both size and functionality. Of note, both the DEBS and Juv TEs appear to be sensitive to the identity of the functional group at the C-3 position and only are capable of effectively forming the 3-hydro-narbonolide **28** species which contains the same functionality as their native products at this position.

Recently, the Sherman lab reported the cloning and in vitro activity of the penultimate (Juv Mod6) and terminal (Juv Mod7) modules from the Juvenimycin biosynthetic pathway which are homologous with the corresponding modules from the tylosin pathway but are more amenable to in vitro analysis.[25] We were curious to test the ability of the hybrid TE strategy to mimic late stage PKS engineering by performing reactions requiring substrate processing by two PKS modules to understand if the hybrid TE modules retained viability for productive interaction with the upstream module. We thus expanded our investigations into TE substrate flexibility by analyzing Juv Mod7 TE hybrids paired with Juv Mod6 in conjunction with the Tyl hexaketide **29** (Table 2.4). LC-HRMS analysis of the Juv Mod7 reactions revealed all three TE domains possessed the ability to generate the 16-membered macrolactone when fused to Juv Mod7. HPLC quantification of tylactone (**25**) production demonstrated Juv Mod7-Juv TE to be the most efficient at 34% conversion, followed closely by Juv Mod7-DEBS TE at 30% conversion. Juv Mod7-Pik TE was significantly less adept at producing tylactone with levels below the detection limit of the HPLC assay. Notably this represents the first example of the DEBS and Pik TE domains producing a 16-membered macrolactone in vitro[27] and confirms that systems containing engineered TE domains are capable of functioning productively with the upstream module.

PKS module
Tyl hexaketide

29

tylactone (25)
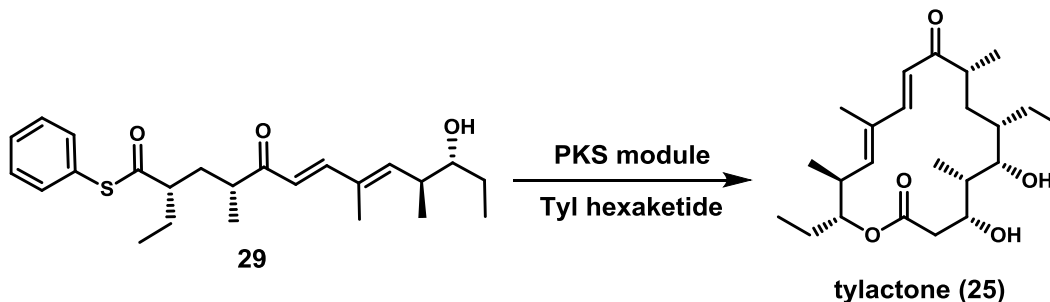
Module	TE	Conversion to 25 (%)
Juv Mod6 + Juv Mod7	DEBS	30 ± 0.8
	Juv	34 ± 3.3
	Pik	trace

Table 2.4 Reaction of Juv Mod7 hybrids with Tyl hexaketide **29**. Enzymatic reaction conditions: sodium phosphate buffer (400 mM, pH = 7.2), 0.5 mM Tyl hexaketide, 10 mM (20 equiv) MM-NAC, 10 mM (20 equiv) M-NAC, 4 mM (8 equiv) 2-vinylpyridine, 0.25 mM (50 mol%) NADP$^+$, 1.25 mM (2.5 equiv) glucose-6-phosphate, glucose-6-phosphate dehydrogenase (2 units/mL), 4 µM (0.8 mol %) PKS module, 4 hours, stationary, RT. Conversion to **25** was monitored by HPLC with data represented as the mean ± standard deviation where n = 3. Trace=detected by LC-HRMS but below the detection limit of HPLC.

Biocatalytic production of macrolactone analogs

After identifying the TE as a crucial gatekeeper domain in the processing of unnatural substrates, we next sought to harness this information for the production of novel macrolactones. To accomplish this, we screened our library of hybrid PKS modules with the Tyl hexaketide **29** to probe each of them for the ability to generate a 14-membered diene macrolactone. Analytical experiments assessed by LC-HRMS indicated reactions containing **29** with Juv Mod6-Pik TE successfully produced a new macrolactone product. To validate our analytical findings, we performed a 0.1 mmol scale reaction and purified the reaction products by preparatory reverse phase HPLC. Initial characterization of the purified reaction products by LC-HRMS (Figure 2.6) indicated that the scale-up reaction did successfully produce the 14-membered macrolactone **30** in a 12% yield (Scheme 2.3). We are currently performing a full structural characterization of the suspected macrolactone product using two-dimensional nuclear magnetic resonance spectroscopy (2D-NMR) to confirm the mass

spectrometry results. The production of this macrolactone was of significant interest as we predicted the conjugated diene would produce a highly strained macrolactone ring structure. Furthermore this product was unobtainable through reactions containing native PKS modules, highlighting the power of TE domain engineering for producing new polyketide products.
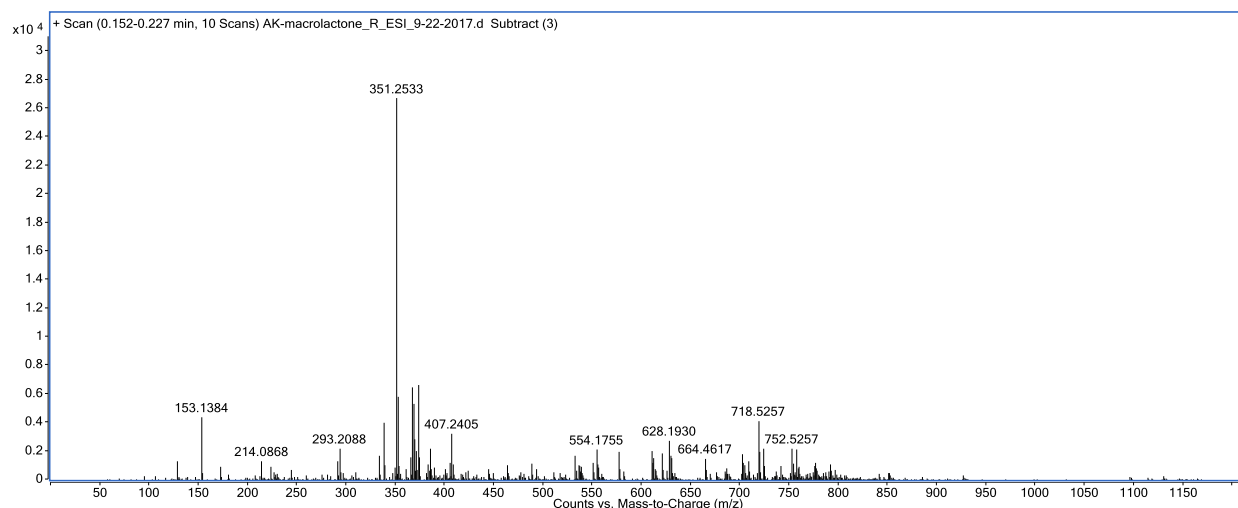
Figure 2.6 LC-HRMS analysis of the purified product from reactions containing Juv Mod6-Pik TE with Tyl hexaketide **29**. The observed molecular ion at m/z 351.2533 matches closely with the calculated value of m/z 351.2529. Full characterization of this molecule is currently in progress.
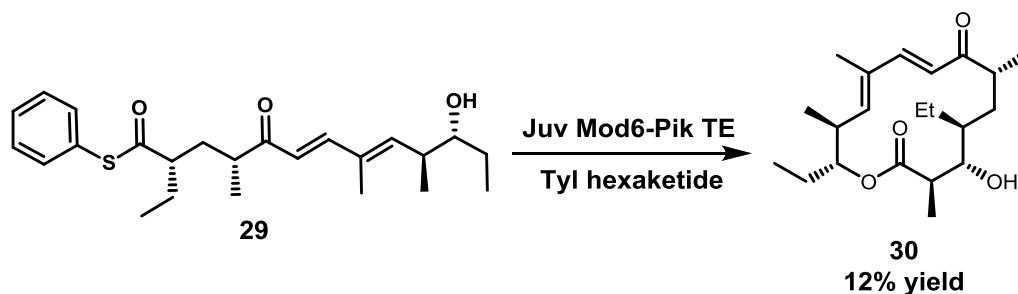
Scheme 2.3 Reaction of Juv Mod6-Pik TE with Tyl hexaketide **29**. Enzymatic reaction conditions: sodium phosphate buffer (400 mM, pH = 7.2), 0.5 mM Tyl hexaketide, 10 mM (20 equiv) MM-NAC, 0.25 mM (50 mol%) NADP$^+$, 1.25 mM (2.5 equiv) glucose-6-phosphate, glucose-6-phosphate dehydrogenase (2 units/mL), 4 µM (0.8 mol %) Juv Mod6-Pik TE, 20 hours, stationary, RT.

Experimental procedures

Chemistry

Reactions were performed in evacuated (<0.05 torr) flame dried glassware backfilled with dry N_2 and run under a positive pressure of dry N_2 provided by a mineral oil bubbler unless stated otherwise (open flask). Reactions at elevated temperatures were controlled by IKA RET Control Visc (model RS 232 C), room temperature (RT) reactions were conducted at ~23 °C, reactions run cooler than room temperature were performed in ice (0 °C) or dry ice/acetone (-78 °C) baths. Analytical thin-layer chromatography (TLC) was performed with EMD 60 F_{254} pre-coated glass plates (0.25 mm) and visualized using a combination of UV, p-anisaldehyde, $KMnO_4$, and bromocresol green stains. Flash column chromatography was performed using EMD 60 Gerduran® (particle size 0.04-0.063) silica gel. Commercial purification system MBraun-MB-SPS # 08-113 provided all dry solvents unless stated otherwise (technical grade). NMR spectra were recorded on a Varian 600 MHz spectrometer. ^{1}H NMR spectra were recorded relative to residual solvent peak (CDCl$_3$ δ_H 7.26 ppm, D$_6$-DMSO δ_H 2.50 ppm, D$_6$-acetone δ_c 2.05 ppm) and reported as follows: chemical shift (ppm), multiplicity, coupling constant (Hz), and integration. Multiplicity abbreviations are as follows: s = singlet, d = doublet, t = triplet, q = quartet, quint = quintet, h = hextet, ovlp = overlap, br = broad signal. ^{13}C NMR spectra were recorded relative to residual solvent peaks (CDCl$_3$ δ_C 77.0 ppm, D$_6$-DMSO δ_c 39.5 ppm, D$_6$-acetone δ_c 29.8 ppm). High resolution mass spectrometry was performed on an Agilent quadrapole time-of-flight spectrometer (Q-TOF 6500 series) by electrospray ionization (ESI).

26: A 4-dram vial was charged with **6**[26] (50 mg, 0.168 mmol, 1 equiv), technical grade DMSO (0.5 mL), IBX (0.09 g, 0.34 mmol, 2 equiv) was added and the reaction was placed in a water bath, and heated to 50 °C with stirring. The reaction was monitored by TLC, and after consumption of starting material (~3 h) the reaction mixture was concentrated under rotary high vacuum and purified directly. Flash chromatography: EtOAc/Hexanes (30:70) gave **26** as a white solid (47 mg, 0.159 mmol, 95% yield).

¹H NMR (700 MHz; CD₃OD): δ 6.88 (dd, *J* = 15.8, 5.3 Hz, 1H), 6.70 (dd, *J* = 15.8, 0.9 Hz, 1H), 5.11 (ddd, *J* = 8.7, 5.1, 2.2 Hz, 1H), 3.87 − 3.82 (m, 1H), 3.35 − 3.32 (m, 1H), 2.84 − 2.79 (m, 1H), 2.62 − 2.56 (m, 1H), 2.41 (dqd, *J* = 12.8, 6.4, 1.9 Hz, 1H), 2.20 − 2.14 (m, 1H), 1.82 − 1.75 (m, 1H), 1.72 − 1.63 (m, 2H), 1.29 (d, *J* = 7.0 Hz, 3H), 1.26 (d, *J* = 7.0 Hz, 3H), 1.22 (d, *J* = 6.9 Hz, 3H), 0.99 (d, *J* = 6.4 Hz, 3H), 0.96 (t, *J* = 7.4 Hz, 3H).

¹³C NMR (700 MHz; CD₃OD): δ 208.5, 206.5, 174.5, 150.0, 127.0, 77.1, 50.7, 46.9, 43.0, 39.8, 38.8, 26.2, 17.31, 14.0, 13.7, 10.7, 9.7.

HRMS: Calculated [M+Na]⁺ 317.1723, found 317.1727.

Cloning of hybrid PKS modules

The cloning, expression, and purification of Pik Mod5[28], Pik Mod5-PikTE[15], Pik Mod6[28], Pik TE[17], DEBS Mod5[13], DEBS Mod5-DEBS TE[13], DEBS Mod6[13], DEBS TE[18], Juv Mod6[25], and Juv Mod7[25] have been previously reported.

The hybrid TE proteins utilized in this study were generated by the following procedures:

The Juv TE domain region was determined by sequence alignment to Pik and DEBS and PCR amplified from pET21-Juv Mod7 using primers 5'-ccaaccgaattcaccggcgcggcgggcgggccacc-3' and 5'-ccaaccctcgagtcatgcggccgcaagcttcggaacgcg-3' and subsequently inserted into pET28b using *Eco*RI and *Xho*I. This construct was then used as the template for two rounds of mutagenesis following the QuikChange site-directed mutagenesis (Stratagene) protocol to remove an internal *Bam*HI site using primers 5'- ggggcctggccggaccccccggcaggactg-3' and 5'-cagtcctgccggggggtccggccaggcccc-3', and *Hind*III site using primers 5'-gtgcggccgcaagtttcggaacgcgag-3' and 5'-cgcctggcggtaccggatcgcctggc-3'. This modified construct was then used for generating the Juv TE hybrid PKS modules as described below.

Pik Mod5-TE hybrids:

pET24b Pik Mod5-Pik TE[15] was digested with *Hind*III and *Xho*I to allow for fusion with the DEBS TE using primers 5'-ccaaccaagcttagcgggactcccgcccgggaagcg-3' and 5'-ccaaccctcgagtgaattccctccgcccagccaggc-3', and Juv TE using primers 5'-ccaaccaagcttaccggcgcggcgggcgggccacc-3' and 5'-ccaaccctcgagtgcggccgcaagcttcggaacgcg-3'.

Pik Mod6–TE hybrids:

The Pik Mod6-TE hybrids were constructed from pET24b Pik Mod6[28] in two steps. First, Pik Mod6 was truncated at the 3' end to its corresponding ACP domain by digestion with *Eco*RI and *Hind*III followed by ligation with a similarly digested PCR amplification product generated using primers 5'-gacagctcacccgaattc-3' and 5'-ccaaccaagcttcagctcgtcgctgatgcgctcggc-3'. Next this intermediate was digested with *Hind*III and *Xho*I to allow for fusion with the DEBS TE using primers 5'-ccaaccaagcttagcgggactcccgcccgggaagcg-3' and 5'-ccaaccctcgagtgaattccctccgcccagccaggc-3', and the Juv TE using primers 5'-ccaaccaagcttaccggcgcggcgggcgggccacc-3' and 5'-ccaaccctcgagtgcggccgcaagcttcggaacgcg-3'.

DEBS Mod5-TE hybrids:

pET28b DEBS Mod5-DEBS TE[13] was digested with *Bam*HI and *Xho*I to allow for fusion with Pik TE using primers 5'-ccaaccggatcctccggggccgacaccggc-3' and 5'-ccttccctcgagtcagcccgccccctcgatgcc-3', DEBS TE using primers 5'-ccaaccggatccagcgggactcccgcccgggaagcg-3' and 5'-ccaaccctcgagtcatgaattccctccgcccagccaggc-3', and Juv TE using primers 5'-ccaaccggatccaccggcgcggcgggcgggccacc-3' and 5'-ccaaccctcgagtcatgcggccgcaagcttcggaacgcg-3'.

DEBS Mod6-TE hybrids:

The DEBS Mod6-TE hybrids were constructed from pET24b DEBS Mod6[13] in two steps. First, DEBS Mod6 was truncated at the 3' end to its corresponding ACP domain by digestion with *Xba*I and *Eco*RI followed by ligation with a similarly digested PCR amplification product generated using primers 5'-cccctctagaaataattttgtttaactttaagaagg-3' and 5'-ccaaccgaattcgagctgctgtcctatgtggtcg-3'. Next this intermediate was digested with *Eco*RI and   *Hind*III to allow for fusion with the Pik TE using primers 5'-ccaaccgaattctccggggccgacaccggc-3' and 5'-ccaaccaagcttgcccgcccccctcgatgcc-3', and the Juv TE using primers 5'-ccaaccgaattcaccggcgcggcgggcgggccacc-3' and 5'-ccaaccaagctttgcggccgcaagcttcggaacgcg-3'.

Juv Mod6-TE hybrids:

The Juv Mod6-TE hybrids were constructed from pET28b Juv Mod6[25] in two steps. First, Juv Mod6 was truncated at the 3' end to its corresponding ACP domain by digestion with *Nde*I and *Hind*III followed by ligation with a similarly digested PCR amplification product generated using primers 5'- ccaacccatatgtcgaacgagcagaagctccgc-3' and 5'- ccaaccaagcttgagcagcccggccaggtgctcggc-3'. Next this intermediate was digested with *Hind*III and *Xho*I to allow for fusion with the Pik TE using primers 5'-ccaaccaagctttccggggccgacaccggc-3' and 5'- ccaaccctcgagtcacttgcccgcccccctcga-3', the DEBS TE using primers 5'- ccaaccaagcttagcgggactcccgcccgggaagcg-3' and 5'-ccaaccctcgagtcatgaattccctccgcccagccaggc-3', and the Juv TE using primers 5'-ccaaccaagcttaccggcgcggcgggcgggccaacc-3' and 5'-ccaaccctcgagtcatgcggccgcaagcttcggaacgcg-3'.

Juv Mod7-TE hybrids:

The Juv Mod7-TE hybrids were constructed from pET21b Juv Mod7[25] in two steps. First, Juv Mod7 was truncated at the 3' end to its corresponding ACP domain by digestion with *Kpn*I and *Hind*III followed by ligation with a similarly digested PCR amplification product generated using primers 5'- cgcctggcggtaccggatcgcctggc-3' and 5'- ccaaccaagcttgagcaggccgtgcaggtgcgcggc-3'. Next this intermediate was digested

with *Hind*III and *Xho*I to allow for fusion with the Pik TE using primers 5'-ccaaccaagctttccgggggccgacaccggc-3' and 5'- ccaaccctcgagcttgcccgcccccctcga-3', and the DEBS TE using primers 5'- ccaaccaagcttagcgggactcccgcccgggaagcg-3' and 5'-ccaaccctcgagtgaattccctccgcccagccaggc-3'.

PKS Biochemistry

Cell culture, protein purification, and enzymatic reactions were performed using water obtained from a Millipore Milli-Q system with Millipore Q-Gard 2/Quantum Ex Ultrapure organex cartridges. *E. coli* culture growth was performed in 15 mL sterile tubes for the seed cultures and 2.8 L Corning Fernbach flasks with deep baffles (3x) for protein expression cultures. Reagents were obtained from the following sources: LB broth (Miller) and glycerol were obtained from EMD. Isopropyl-β-D-thiogalactopyranoside (IPTG), Kanamycin sulfate (Kan), and Ampicillin (Amp) were obtained from Gold Biotechnology. NaCl, $CaCl_2$ and imidazole were obtained from Fisher Scientific. Lysozyme was purchased from RPI, PD-10 columns were purchased from GE scientific, and Ni-NTA agarose resin was purchased from Qiagen. The pH of all solutions was monitored via a Symphony SB70P pH meter calibrated according to the manufacturer's specifications. Optical density (OD_{600}) was determined using an Eppendorf Biophotometer and cell lysis was accomplished using a 550 Sonic Dismembrator purchased from Fisher Scientific. All solutions were autoclaved or sterile filtered (0.2 μm) prior to use.

Buffers:
 lysis: HEPES (50 mM), NaCl (300 mM), imidazole (10 mM), glycerol (10% v/v), pH 8.0.
wash: HEPES (50 mM), NaCl (300 mM), imidazole (30 mM), glycerol (10% v/v), pH 8.0.
elution: HEPES (50 mM), NaCl (300 mM), imidazole (300 mM), glycerol (10% v/v), pH 8.0.
storage: HEPES (50 mM), NaCl (150 mM), EDTA (1 mM), glycerol (20% v/v), pH 7.2.

Protein expression

A starter culture was generated by inoculating 5 mL of LB broth containing Kan or Amp (50 mg/L) with fresh transformants of E. coli (BAP1)[23] cells containing the corresponding plasmids for expression of the respective PKS proteins and grown overnight at 37 °C. Following the overnight growth, the entire starter culture was subsequently used to inoculate an expression culture of 1 L of TB containing Kan or Amp (50 mg/L) and grown at 37 °C to an OD_{600} of 0.3-0.4. The expression cultures were then cooled to 18 °C and growth was maintained until an OD_{600} of 0.7-0.8 was reached, at which point protein expression was induced via addition of IPTG (350 µM) and the cultures were incubated at 200 RPM at 18 °C for 20 hours.

Protein purification

Protein expression cultures were cooled to 4 °C and harvested by centrifugation (6,500 x g, 10 min, 4 °C). The pelleted cells were then suspended in 5 mL of lysis buffer per gram of cells via vortex. Cell lysis was accomplished by the addition of 0.4 mg/mL lysozyme and the solution was then sonicated on ice (100 x 3s with 10s rest periods). The resulting cellular lysate was then pelleted by centrifugation (60,000 x g, 30 min, 4 °C) and the supernatant was applied to 6 mL of pre-equilibrated Ni-NTA resin. After binding, the column was washed with 15 column volumes of wash buffer and the target protein was subsequently eluted with 4 column volumes of elution buffer. Elution fractions were determined by their absorption at 280 nm, pooled, and buffer exchanged into storage buffer using a pre-equilibrated PD-10 column. After buffer exchange, the elution fractions were once again monitored via their absorption at 280 nm, pooled, flash frozen in liquid N_2, and stored at -80 °C.

Analytical enzymatic reactions

All analytical scale reactions were performed in triplicate at a volume of 50 µL and quenched with 3 volumes of MeOH (150 µL), clarified by centrifugation (17,000 x g, 30 min, 4 °C) and analyzed for product formation by HPLC. 2-vinylpyridine (Sigma) was employed as a thiol scavenger. Enzymatic reactions utilizing NBOM protected substrates were performed over two steps. First, a solution of ascorbic acid (25 mM final

concentration), sodium metabisulfite (1 mM final concentration), NBOM protected substrate (1 mM final concentration), and H_2O (requisite dead volume) was irradiated under a consumer facial tanning lamp at a height of 14 cm (Verseo #AH129c) for 20 min to furnish the deprotected Pik hexaketide. NOTE: Irradiation through the side of the microtubes employed (Axygen #MCT-175-C) did not interfere with photolysis, and this process was reproducible over the course of this study. After photolysis, the solution was diluted with reaction buffer, MM-NAC, and catalysis was initiated via the addition of enzyme.

HPLC analysis:

Macrolactone production was monitored via analytical high performance liquid chromatography (HPLC) using a Shimadzu LC-20AD.

Product formation in Table 2.2 was quantified using Phenomenex Luna 5µ C18 250 x 4.6 mm column (serial 466013-1) monitoring at a wavelength of 236 nm. Separation was accomplished by the following method: 1.5 mL/min, solvent A: H_2O 0.1% formic acid, solvent B: MeCN 0.1% formic acid, 5% B 0-1 min, 5-100% B linear gradient 1-12 min, 100% B 12-15 min, 5% B 15-17.5 min.

Product formation in Table 2.3 and 2.4 was quantified using a Zorbax SB-Phenyl 3.5 µM 4.6 x 150 mm column (part number 863953-912) monitoring at a wavelength of 236 nm. Separation was accomplished by the following method: 3.0 mL/min, solvent A: H_2O 0.1% formic acid, solvent B: MeCN 0.1% formic acid, 5% B 0-1 min, 5-70% B linear gradient 1-13 min, 100% B 13-15 min, 5% B 15-17 min.

LC-HRMS analysis:

Analytical liquid chromatography-mass spectrometry (LC-MS) was performed on an Agilent LC system (1290 series) coupled to an Agilent QTOF mass spectrometer (6500 series) using a Phenomenex Synergi 4µ Hydro RP 100 x 2 mm column (serial 48836-5) heated to 50 °C. Method: 0.4 mL/min, solvent A: H_2O 0.1% formic acid, solvent B: MeCN 0.1% formic acid, 0% B 0-2 min, 0-100% B linear gradient 2-10 min, 100% 10-11 min, 0% B 11-12 min, 0-1 min were diverted to waste.

Large scale reactions for product characterization

Incubation of **27** with DEBS Mod6-DEBS TE:

Reaction conditions: sodium phosphate buffer (400 mM, 20% v/v glycerol, 92 mL total, pH = 7.2), hexaketide **27** (51 mg, 0.09 mmol, 1 mM), MM-SNAC (20 equiv, 20 mM), NADP$^+$ (0.5 equiv, 0.5 mM), glucose-6-phosphate (2.5 equiv, 2.5 mM), glucose-6-phosphate dehydrogenase (2 units/mL), 2-vinylpyridine (8 mM), ascorbic acid (25 mM), sodium metabisulfite (1 mM), DEBS Mod6 (1 µM, 0.1 mol%), 20 hours, stationary, RT.

Workup and purification: Quenched with acetone (2x volume, 184 mL), placed in a -20 °C freezer for 1 hour and filtered through a celite plug. Remaining insoluble material was suspended in acetone and this solution was used to rinse the celite plug. Acetone was removed through rotary evaporation and the aqueous layer was saturated with NaCl and extracted 3x EtOAc. Combined organic layers were concentrated. **30** was purified directly by preparatory HPLC using a Phenomenex Luna 5u C18 250 x 21.2 mm column (serial 444304-4) monitoring at 250 nM. Method 9 mL/min, A: H$_2$O 0.1% formic acid, B: MeCN 0.1% formic acid, 5% B 0-5 min, 5-100% B linear gradient 5-45 min, 100% B 45-65 min, 5% 65-75 min.

28 from Pik hexaketide **27** (4.1 mg, 0.089 mmol, 12.6% yield).

^{1}H NMR (700 MHz, cd$_3$od) δ 7.11 (dd, J = 15.9, 4.5 Hz, 1H), 6.27 (dd, J = 15.9, 2.0 Hz, 1H), 5.10 (ddd, J = 9.4, 4.5, 1.8 Hz, 1H), 3.44 (dd, J = 9.0, 1.5 Hz, 1H), 3.41 (dd, J = 7.2, 1.3 Hz, 1H), 2.69 (ddd, J = 8.7, 4.3, 2.0 Hz, 1H), 2.63 (ddd, J = 10.8, 7.0, 4.0 Hz, 1H), 2.55 (m, 1H), 1.78 (m, 1H), 1.76 (ddd, J = 14.1, 7.0, 4.9 Hz, 1H), 1.63 (ddd, J = 14.0, 7.4, 4.6 Hz, 1H), 1.41 – 1.35 (m, 1H), 1.31 – 1.26 (m, 1H), 1.25 (s, 1H), 1.22 (d, J = 4.0 Hz, 1H), 1.20 (t, J = 4.4 Hz, 1H), 1.18 (d, J = 6.9 Hz, 3H), 1.15 (d, J = 6.9 Hz, 3H), 1.13 (d, J = 6.8 Hz, 3H), 1.03 (d, J = 6.7 Hz, 3H), 0.95 (m, 3H), 0.93 (d, J = 7.4 Hz, 2H).

13C NMR (700 MHz; CD$_3$OD) δ 206.07, 177.72, 152.84, 124.72, 78.40, 78.10, 75.00, 45.90, 45.30, 42.93, 40.80, 37.45, 35.71, 26.52, 19.41, 18.03, 15.61, 10.92, 10.34, 8.92.

HRMS: Calculated [M+Na]$^+$ 377.2298, found 377.2307

Incubation of **29** with Juv Mod6-Pik TE:

Reaction conditions: sodium phosphate buffer (400 mM, 20% v/v glycerol, 205 mL total, pH = 7.2), hexaketide **31** (42 mg, 0.1 mmol, 0.5 mM), MM-SNAC (20 equiv, 10 mM), NADP$^+$ (0.5 equiv, 0.25 mM), glucose-6-phosphate (2.5 equiv, 1.25 mM), glucose-6-phosphate dehydrogenase (2 units/mL), Juv Mod6-Pik TE (4 µM, 0.8 mol%), 20 hours, stationary, RT.

HRMS: Calculated [M+H]$^+$ 351.2529, found 351.2533

References

(1)   Weissman, K. J. *Nat. Prod. Rep.* **2016**, *33*, 203.

(2)   Kittendorf, J. D.; Sherman, D. H. *Curr. Opin. Biotechnol.* **2006**, *17*, 597.

(3)   Reeves, C. D. *Crit. Rev. Biotechnol.* **2003**, *23*, 95.

(4)   Wong, F. T.; Khosla, C. *Curr. Opin. Chem. Biol.* **2012**, *16*, 117.

(5)   McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Betlach, M.; Ashley, G. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 1846.

(6)   Xue, Q.; Ashley, G.; Hutchinson, C. R.; Santi, D. V. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 11740.

(7)   Menzella, H. G.; Reid, R.; Carney, J. R.; Chandran, S. S.; Reisinger, S. J.; Patel, K. G.; Hopwood, D. A.; Santi, D. V. *Nat. Biotechnol.* **2005**, *23*, 1171.

(8)   Jacobsen, J. R.; Keatinge-Clay, A. T.; Cane, D. E.; Khosla, C. *Bioorg. Med. Chem.* **1998**, *6*, 1171.

(9)   Leaf, T.; Cadapan, L.; Carreras, C.; Regentin, R.; Ou, S.; Woo, E.; Ashley, G.; Licari, P. *Biotechnol. Prog.* **2000**, *16*, 553.

(10)   Kinoshita, K.; G Williard, P.; Khosla, C.; Cane, D. E. *J. Am. Chem. Soc.* **2001**, *123*, 2495.

(11)   Harvey, C. J.; Puglisi, J. D.; Pande, V. S.; Cane, D. E.; Khosla, C. *J. Am. Chem. Soc.* **2012**, *134*, 12259.

(12)   Aldrich, C. C.; Beck, B. J.; Fecik, R. A.; Sherman, D. H. *J. Am. Chem. Soc.* **2005**, *127*, 8441.

(13)   Mortison, J. D.; Kittendorf, J. D.; Sherman, D. H. *J. Am. Chem. Soc.* **2009**, *131*, 15784.

(14)   Hansen, D. A.; Koch, A. A.; Sherman, D. H. *J. Am. Chem. Soc.* **2017**, *139*, 13450.

(15)   Yin, Y.; Lu, H.; Khosla, C.; Cane, D. E. *J. Am. Chem. Soc.* **2003**, *125*, 5671.

(16)   Pinto, A.; Wang, M.; Horsman, M.; Boddy, C. N. *Org. Lett.* **2012**, *14*, 2278.

(17)   Lu, H.; Tsai, S.-C.; Khosla, C.; Cane, D. E. *Biochemistry* **2002**, *41*, 12590.

(18)   Gokhale, R. S.; Hunziker, D.; Cane, D. E.; Khosla, C. *Chem. Biol.* **1999**, *6*, 117.

(19)   Watanabe, K.; Wang, C. C. C.; Boddy, C. N.; Cane, D. E.; Khosla, C. *J. Biol. Chem.* **2003**, *278*, 42020.

(20)   Notredame, C.; Higgins, D. G.; Heringa, J. *J. Mol. Biol.* **2000**, *302*, 205.

(21)   Robert, X.; Gouet, P. *Nucleic Acids Res.* **2014**, *42*, W320.

(22)   Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. A.; Müller, R.; Wohlleben, W.; Breitling, R.; Takano, E.; Medema, M. H. *Nucleic Acids Res.* **2015**, *43*, W237.

(23)   Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C. *Science* **2001**, *291*, 1790.

(24)   Hansen, D. A.; Rath, C. M.; Eisman, E. B.; Narayan, A. R.; Kittendorf, J. D.; Mortison, J. D.; Yoon, Y. J.; Sherman, D. H. *J. Am. Chem. Soc.* **2013**, *135*, 11232.

(25)   Lowell, A. N.; DeMars, M. D.; Slocum, S. T.; Yu, F.; Anand, K.; Chemler, J. A.; Korakavi, N.; Priessnitz, J. K.; Park, S. R.; Koch, A. A.; Schultz, P. J.; Sherman, D. H. *J. Am. Chem. Soc.* **2017**, *139*, 7913.

(26)   Hansen, D. A.; Koch, A. A.; Sherman, D. H. *J. Am. Chem. Soc.* **2015**, *137*, 3735.

(27)   Horsman, M. E.; Hari, T. P. A.; Boddy, C. N. *Nat. Prod. Rep.* **2016**, *33*, 183.

(28)   Beck, B. J.; Aldrich, C. C.; Fecik, R. A.; Reynolds, K. A.; Sherman, D. H. *J. Am. Chem. Soc.* **2003**, *125*, 4682.

Chapter III

Interrogating thioesterase domains directly

DH contributed to Figure 3.2, Scheme 3.1, and Table 3.3.

AK contributed to all others.

Introduction

Macrocycles are a common motif amongst natural product and natural product derived therapeutics, with >100 marketed drugs possessing a macrocyclic core.[1] The conformational preorganization of these large rings enables precise display of their functional groups to engage challenging targets such as protein-protein interactions, and, in the case of macrolides, ribosomal machinery.[2-4] The macrolactone core of macrolide[5] natural products is formed through two steps: first, the linear intermediate is transferred from the upstream acyl carrier protein (ACP) to a catalytic serine of the TE via a transesterification to form an acyl-enzyme complex. Next, the acyl intermediate is offloaded via intermolecular water hydrolysis or intramolecular nucleophilic attack to release the product as a linear acid or macrolactone, respectively[6] (Figure 3.1). The macrolactonization of natural product analogs is notoriously difficult to perform synthetically[7] and formation of the macrocycle is a key feature of many bioactive polyketide and nonribosomal peptide natural products.[8] Thus, we were intrigued by the prospect that the TE domain is serving as a gatekeeper domain in engineered PKS pathways.
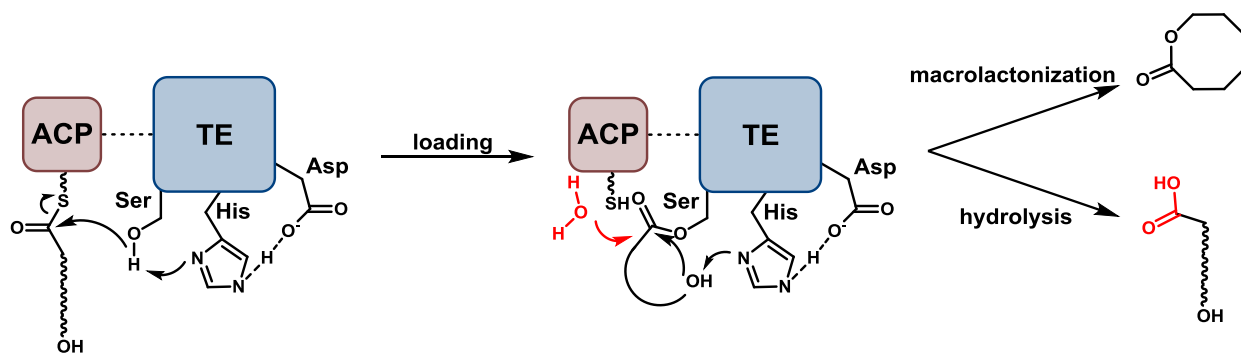
Figure 3.1 TE catalyzed macrolactonization or hydrolysis of an ACP-tethered polyketide intermediate.

Biochemical and structural studies of the Pik TE and related DEBS TE (erythromycin pathway) have provided initial insights into the mechanism of ring formation over hydrolysis.[9,10] While it is generally accepted that modular type I TE domains have high substrate flexibility for the initial acylation step, the second, macrolactone forming release step is far more stringent.[6] Since the formation of the macrocyclic core is essential for downstream tailoring and biological activity of a natural product, aberrant hydrolysis limits access to new macrocyclic analogs. In vivo

44

engineering of the DEBS biosynthetic pathway has shown that DEBS TE possesses some tolerance to modifications in both the length and functionality of the linear polyketide substrate. Modifications in the DEBS pathway have yielded a large number of 6-dEB analogs from differential extender unit incorporation and reductive processing[11-18] as well as macrolactones ranging in size from 6- to 16-membered rings.[19-22] However, the titers of these unnatural products are greatly diminished compared to wild type production levels[6], and the inherent complexity of in vivo biosynthesis has prevented identification of pathway bottlenecks, such as TE domains. Initial in vitro biochemical characterization of the DEBS TE[23,24] provided further evidence for the relatively high substrate tolerance of PKS TEs for acylation and hydrolysis, as terminally (omega) hydroxylated fatty acids and substrates resembling simplified DEBS heptaketides were all hydrolyzed by DEBS TE. However, the ability of PKS TEs to cyclize substrates other than their native linear intermediates has proved to be much more limited. In fact, TE mediated macrolactonization has only been observed in a select few studies.[25-34]

In addition to its native substrate, DEBS TE has been shown to catalyze macrolactonization of unnatural mimics of the DEBS heptaketide.[30,33] However, there are no reports of either the Pik or DEBS TEs catalyzing macrolactonization of a substrate containing a nucleophilic hydroxyl group with an unnatural, epimerized (*S*)-configuration. When probed for the ability to form an epimeric heptaketide mimic of 6-dEB , the DEBS TE displayed a high level of stereoselectivity for the natural (*R*)-configuration, and exclusively hydrolyzed the unnatural (*S*)-stereoisomer[33], adding to the observations in Chapter II indicating strict stereospecificity in TE catalyzed macrolactonization.

Stabilization of the Pik hexaketide

To further our understanding of the TE domain in the processing of unnatural substrates, we next set out to investigate the mechanistic parameters that govern macrolactonization directly by probing the TE as an excised domain.[34] However, the previously observed instability of the Pik TE native substrate (Pik hexaketide **4**) imposes

considerable experimental challenges[35,36], and required a practical solution prior to downstream study (Figure 3.2). While mature natural products typically possess adequate stability to survive isolation and purification from natural sources, polyketide intermediates often degrade rapidly through intramolecular hemiketalization and dehydration pathways presenting experimental bottlenecks in terms of synthetic accessibility and limited shelf-life.[35-38] Although the structural basis remains unclear, polyketide elongation intermediates that are covalently attached to the ACP domain during biosynthesis are likely stabilized through sequestration within the PKS polypeptide scaffold.[39,40]
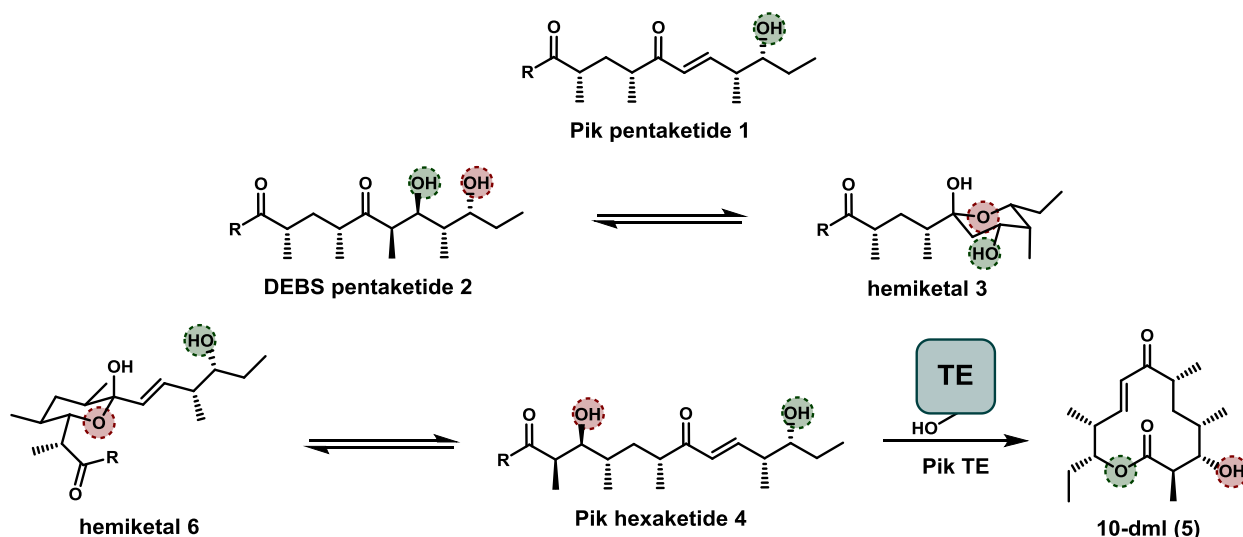
Figure 3.2 Examples of previously studied native PKS chain elongation intermediates. Hydroxyl groups highlighted in red form hemiketals, while those highlighted in green are unreactive.

Thus, to protect the pikromycin hexaketide for enhanced stability to enable downstream biochemical studies, we considered two distinct stabilization strategies: (i) a sterically undemanding protecting group that would remain attached throughout the catalytic cycle, and (ii) a protecting group that could be removed in a controlled manner to provide the native hexaketide immediately before use in enzymatic reactions. The corresponding seco-acids were produced starting from 10-dml (**5**) acquired by fermentation of an engineered strain of *Streptomyces venezuelae* ATCC 15439. Ultimately, a methyl ether protecting group was chosen to satisfy (i) and a photocleavable 2-nitrobenzyloxymethyl ether moiety (NBOM)[41] was explored to address objective (ii) (see reference 34 for synthetic details).

With C-3 protected hexaketides in hand, we next analyzed the ability of Pik TE to generate 10-dml (**5**) and methyl protected 10-dml **9** from NBOM protected thiophenol hexaketide **7** and methyl protected thiophenol hexaketide **8**, respectively. Initial reaction schemes were chosen based on previous work from Hansen et al. using thiophenol thioesters to load the substrate onto the PKS machinery in vitro.[42] HPLC quantification of reaction products using synthetic standards revealed that the in vitro reactions were sluggish resulting in conversions of 36% to **5** and only 10% to C-3 methyl protected **9** (Scheme 3.1), prompting further investigations into substrate optimization.
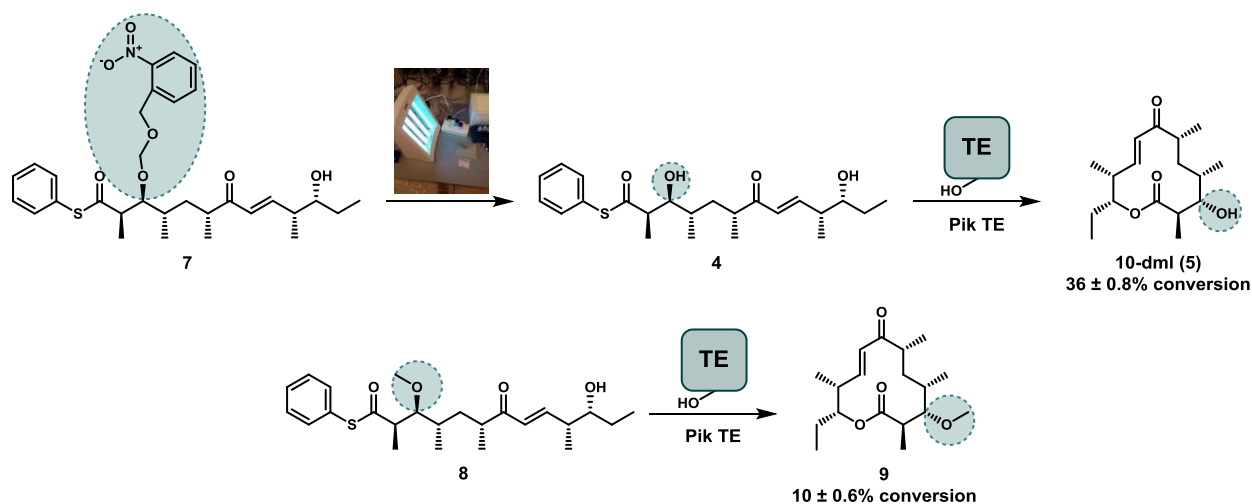
Scheme 3.1 Probing the Pik TE as an excised domain. Reactions containing Pik TE and the thiophenol thioester C-3 methyl and NBOM protected hexaketides **7** and **8** result in the production of the corresponding C-3 protected macrolactones, albeit with lower than desired efficiency.

Substrate controlled divergence in polyketide synthase catalysis

In order to study isolated PKS TE domains in vitro, investigations have relied upon electrophilic thioesters for diffusive loading in lieu of transfer from an upstream ACP. Historically, *N*-acetylcysteamine[43,44] (NAC) has been the thioester of choice as it mimics the terminal portion of the phosphopantetheine arm that tethers a growing polyketide chain.[45] In vitro studies of PikAIV with its native substrate have highlighted a key observation: Specifically, when incubated directly with *N*-acetylcysteamine Pik hexaketide **4**[46] PikAIV afforded a 4:1 ratio of macrolactones 10-dml (**5**) and narbonolide (**10**).[46] However, reaction schemes pairing PikAIII/PikAIV[35,47] with Pik pentaketide **1** where PikAIII performs an extension and delivers the hexaketide to PikAIV via an ACP_5

thioester favor narbonolide as the major product (Figure 3.3). Additionally, optimization of PikAIII (as an unnatural TE fusion[48] or when paired with the final module, PikAIV) demonstrated improved catalysis with thiophenol thioesters[49] over *N*-acetylcysteamine thioesters.[35] These results suggest that the choice of thioester employed by the substrate largely determines the enzymatic domain(s) that will be efficiently acylated within a PKS assembly line. Thus, we hypothesized the lower than expected conversions reported in Scheme 3.1 were due in part to suboptimal thioester usage and motivated us to further explore substrate engineering approaches for in vitro reactions with Pik TE.
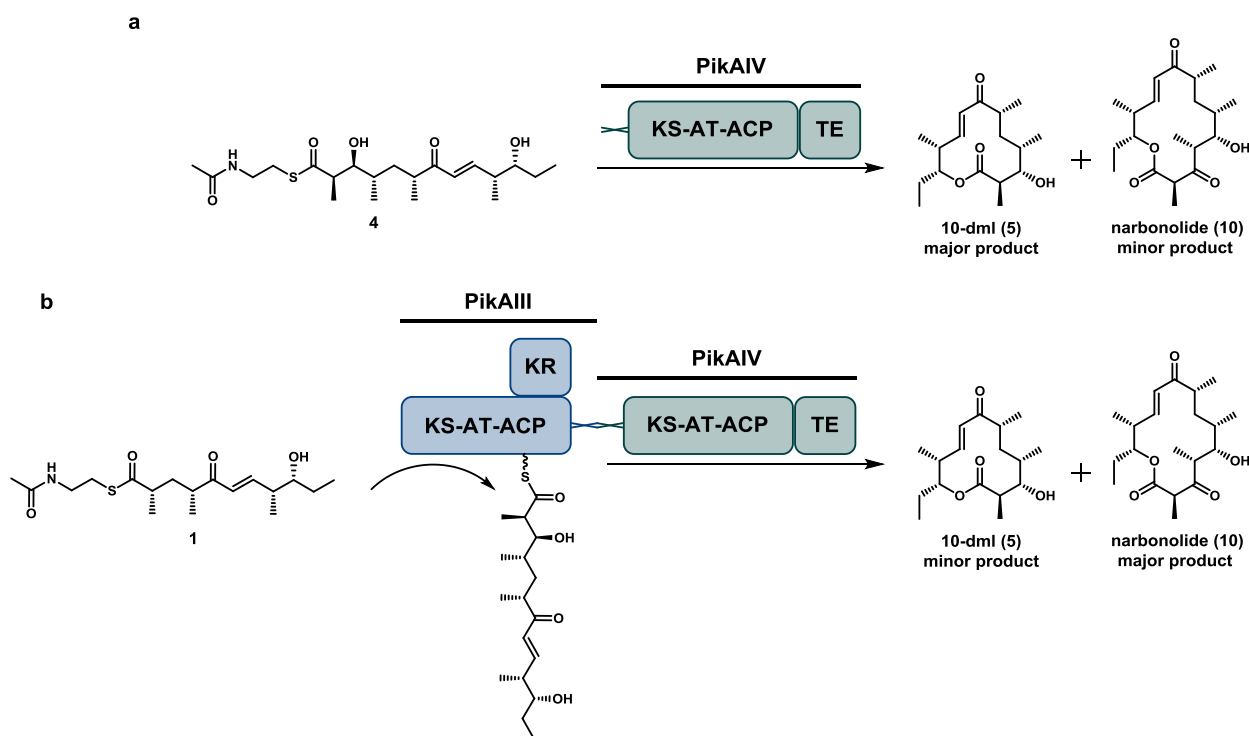
Figure 3.3 Product distributions from previous in vitro analysis of PikAIV. (a) Direct incubation of PikAIV with the *N*-acetylcysteamine thioester of the Pik hexaketide **4** results primarily in direct cyclization to 12-membered ring 10-dml. (b) Reaction schemes using the Pik pentaketide and PikAIII to present the hexaketide intermediate as an acyl-ACP result primarily in full-module processing to 14-membered ring narbonolide.

Starting with the C-3 protected seco-acids, we synthesized a series of hexaketides with a variety of thio- and oxoesters (see reference 28 for synthetic details) to test for loading onto Pik TE (Table 3.1). For NBOM protected substrates, 4-nitrophenol (entry 6) and *N*-hydroxysuccinimide (entry 8) substrates decomposed

Table 3.1 Evaluation of stabilized Pik hexaketides with Pik TE. Enzymatic reaction conditions: 1 mM Pik hexaketide, 8 mM 2-vinylpyridine, 1 mM sodium metabisulfite and 25 mM ascorbic acid with NBOM photolysis, 1 mol% Pik TE (10 µM), 4 hours, RT. Conversion to **5** (R_3 = H) or **9** (R_3 = Me) was monitored (HPLC) with data represented as the mean ± standard deviation where n = 3.

rapidly upon photolysis and subsequently gave low conversion to macrolactones. In contrast, the corresponding hexaketide thiophenol, benzyl mercaptan and *N*-acetylcysteamine thioesters photolyzed smoothly (entries 2, 4, and 10, respectively), with *N*-acetylcysteamine yielding the highest level of conversion to 10-dml (**5**). The methyl protected substrates followed a similar trend; however, the overall conversions were decreased in comparison to the photolyzed substrates, indicating the methyl ether protecting group negatively effects macrolactonization to a small extent. These experiments demonstrated the significant variation in macrolactonization efficiency to 10-dml (**5**) or methyl 10-dml **9** dictated by the ester employed. *N*-acetylcysteamine thioester (entries 9-10) gave the highest conversion to 10-dml (**5**) or methyl protected 10-dml (**9**) under all conditions tested, with *N*-hydroxysuccinimide esters providing moderate conversion to methyl protected 10-dml (**9**) (entry 7). These results indicate

that thiophenol thioesters are inefficient for direct macrolactonization utilizing the excised TE domain, and highlight the crucial nature of substrate optimization when studying PKS enyzmes in vitro. Surprisingly, a series of control reactions were conducted with NBOM protected hexaketides and Pik TE without photolysis, and yielded appreciable conversions to NBOM protected 10-dml **11** (Table 3.2). The same general trends were observed with *N*-acetylcysteamine giving the highest levels of conversion, followed by *N*-hydroxysuccinimide, with aryl and benzyl thio- and oxoesters giving uniformly low levels of product formation.

NBOM protected Pik hexaketide   **11**

Entry	R_1	$h\nu$	%Conv to 11
1	phenyl-S	-	3 ± 0.1
2	benzyl-S	-	trace
3	O_2N-phenyl-O	-	5 ± 0.15
4	N-hydroxysuccinimide-O	-	19 ± 3.7
5	acetyl-NH-CH2CH2-S	-	41 ± 2.8

Table 3.2 Evaluation of NBOM protected Pik hexaketides without photolysis. Enzymatic reaction conditions: 1 mM Pik hexaketide, 8 mM 2-vinylpyridine, 1 mol% Pik TE (10 µM), 4 hours, RT. Conversion to **11** was monitored (HPLC) with data represented as the mean ± standard deviation where n = 3. Trace=detected by LC-HRMS but below the detection limit of HPLC.

After observing the stark variations in TE catalysis with hexaketides bearing different thio- and oxoesters, we next pursued experiments utilizing this substrate panel

with the terminal module PikAIV to determine if the variable conversions remained in the context of full module catalysis. To accomplish this, we performed a series of in vitro reactions containing PikAIV paired with the Pik hexaketide and methylmalonyl *N*-acetylcysteamine (MM-NAC)[49,50] as the extender unit (Table 3.3). Again, NBOM protected substrates, 4-nitrophenol (entry 6) and *N*-hydroxysuccinimide (entry 8) substrates decomposed rapidly upon photolysis and subsequently gave low conversion to macrolactones. In contrast, the corresponding hexaketide thiophenol, benzyl mercaptan and *N*-acetylcysteamine thioesters photolyzed smoothly (entries 2, 4, and 10, respectively) though benzyl mercaptan thioesters (entry 4) gave lower overall conversion to either macrolactone. Remarkably, we observed significant selectivity in product formation depending on the type of ester employed, where the hexaketide thiophenol thioester (entry 2) demo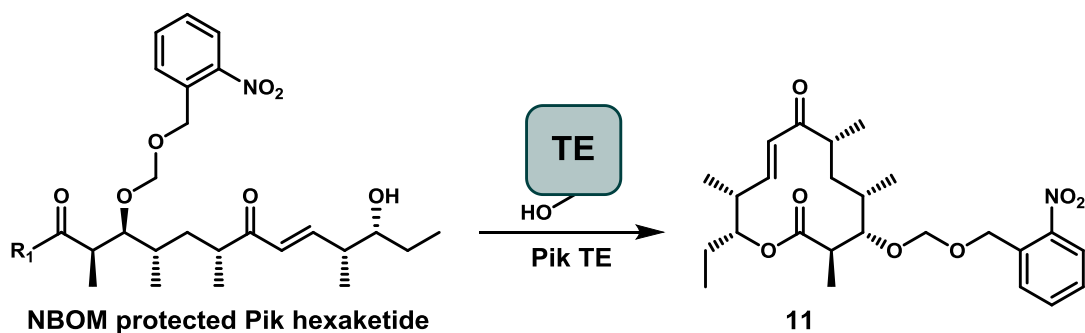nstrated greater than 10:1 selectivity for generating narbonolide (**10**), while the hexaketide *N*-acetylcysteamine thioester (entry 10) showed greater than 10:1 selectivity for 10-dml (**5**) production. In parallel experiments methylated substrates were converted to methyl protected 10-dml (**9**) or methyl protected narbonolide (**12**) albeit with selectivity shifted toward methyl 10-dml (**9**) and reduced overall conversions relative to native the substrates generated through initial NBOM photolysis.

The work described herein provides further insight into substrate loading parameters in modular PKS catalysis using thiophenol thioesters[49], indicating that product formation can be influenced by the type of substrate ester employed. While reactions with the native, upstream acyl-ACP likely offer the highest level of biosynthetic fidelity through docking domain mediated chain transfer[51], thiophenol thioesters appear to be highly effective for achieving full module catalysis. We propose that substrate control can be explained, at least in part, through the observation that thiophenol thioesters suffer diminished conversion when incubated with the excised Pik TE domain (entries 1 and 2, Table 3.1) relative to the corresponding NAC thioesters (entries 9 and 10, Table 3.1). Moreover, we[36] and others[23,52] have shown previously that PKS TE domains function as flexible hydrolases when substrate macrolactonization cannot be achieved. PikAIV is unique among terminal PKS modules as the TE domain is able to form both 12- and 14-membered rings. However, the likely result of errant TE acylation
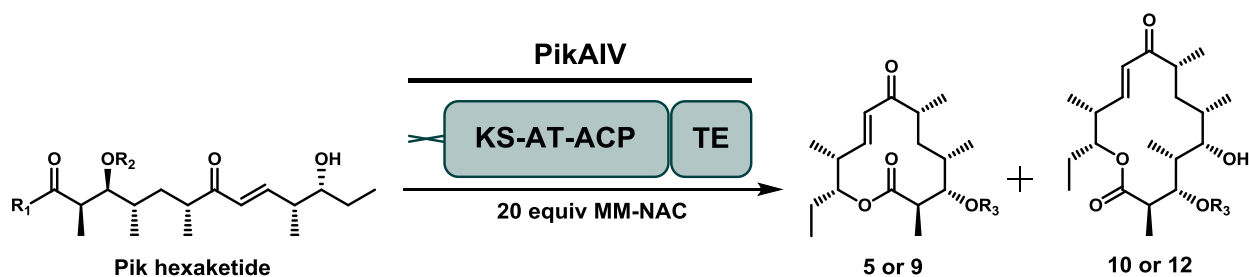
Entry	R$_1$	R$_2$	$h\nu$	%Conv to 5 or 9	%Conv to 10 or 12
1		Me	-	4 ± 0.1	6 ± 0.2
2		NBOM	+	3 ± 0.3	41 ± 0.5
3		Me	-	trace	ND
4		NBOM	+	trace	trace
5		Me	-	4 ± 0.2	ND
6		NBOM	+	trace	15 ± 0.6
7		Me	-	18 ± 3.3	3 ± 0.4
8		NBOM	+	3 ± 1.3	trace
9		Me	-	28 ± 2.1	ND
10		NBOM	+	52 ± 2.9	4 ± 0.3

Table 3.3 Evaluation of stabilized Pik hexaketides with PikAIV and MM-NAC extender unit. Enzymatic reaction conditions: 1 mM Pik hexaketide, 20 mM MM-NAC, 8 mM 2-vinylpyridine, 1 mM sodium metabisulfite and 25 mM ascorbic acid with NBOM photolysis, 0.25 mol% PikAIV (2.5 μM), 4 hours, RT. Conversion to **5** (R$_3$ = H) or **9** (R$_3$ = Me) and **10** (R$_3$ = H) or **12** (R$_3$ = Me) was monitored (HPLC) with data represented as the mean ± standard deviation where n = 3. Trace=detected by LC-HRMS but below the detection limit of HPLC. ND= not detected.

in typical PKS modules is substrate hydrolysis and decreased conversion to the desired macrocyclic product, which greatly limits the effectiveness of PKS modules as biocatalysts. As such, optimizing PKS substrates through substrate engineering may minimize hydrolysis and other undesired side reactions.

In summary, the results with PikAIV and Pik TE clearly demonstrate the critical importance of substrate engineering as a tool for studying PKS enzymes in vitro. Methyl ether and NBOM mediated stabilization proved to be a viable strategy for reactions with the excised Pik TE. Additionally, the NBOM protecting group proved to effectively stabilize the Pik hexaketide prior to in vitro reactions and should offer ready stabilization of a range of advanced polyketide chain elongation substrates for PKS functional

studies. Moreover, its ability to undergo rapid and efficient photo-induced deprotection demonstrates its role as an effective tool for in vitro studies of PKS enzymes.

Confirmation of a thioesterase bottleneck in the processing of unnatural substrates

With optimized in vitro conditions for probing the excised TE domain directly, we next performed experiments to directly confirm the TE bottleneck identified in Chapter II. To accomplish this, we began by synthesizing the *N*-acetylcysteamine thioester of the C-11-epimerized hexaketide, which is analogous to the C-9-epimerized pentaketide which failed to be processed by PikAIII-TE in Chapter II (see reference 47 for synthetic details), for probing the TE domain substrate stereospecificity directly. With C-11-epimerized methyl ether hexaketide **14** in hand, we next evaluated the ability of Pik TE to catalyze the macrolactonization of **14** to the corresponding epimeric macrolactone. HPLC analysis of analytical reactions containing the epimerized **14** resulted in complete hydrolysis to **15,** and failed to produce any macrocylic product (Scheme 3.2). The exclusive hydrolysis of epimerized **14** confirms the TE domain as the dominant catalytic bottleneck in the processing of diastereomeric pentaketides observed in Chapter II.
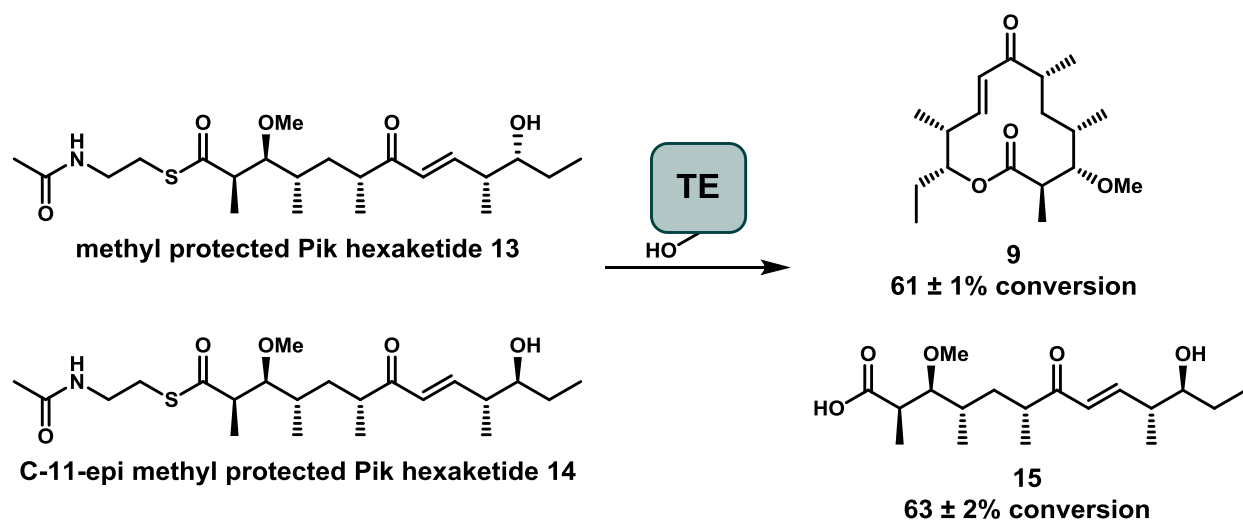
Scheme 3.2 Pik TE displays a high level of substrate stereospecificity. Enzymatic reaction conditions: sodium phosphate buffer (400 mM, pH 7.2), 1 mM hexaketide, 8 mM 2-vinylpyridine, purified Pik TE (10 µM), 4 hours, stationary, RT. Conversion **9** and **15** was monitored (HPLC) with data represented as the mean ± standard deviation where $n = 3$.

53

A single active site mutation in the pikromycin thioesterase generates a more effective macrocyclization catalyst

After confirming Pik TE as the catalytic bottleneck in our engineered PKS modules from Chapter II, we next sought to engineer the TE domain for increased substrate flexibility, and the ability to generate diastereomeric macrolactones. We first desired to increase the efficiency of TE domain loading using 4-nitrophenyl substrates as the release of 4-nitrophenol from these reactions provides a spectrophotometric handle that is useful in directed evolution engineering methods.[54] We expected that the KS over TE bias observed with thio- and oxoesters other than *N*-acetylcysteamine (entries 2 and 6, Table 3.3) was due in part to the cysteine active site nucleophile of the KS domain vs the corresponding serine nucleophile of the TE domain. Due to differences in atomic radii, electronegativity, and polarizability, sulfur atoms are significantly more nucleophilic than their oxygen congeners.[55] Because of the larger size of the atom, non-bonding electron-pairs on sulfur are softer (more polarizable) than those on oxygen and consequently, electron-pairs on sulfur are better nucleophiles. Therefore, we reasoned that substitution of the WT Pik TE nucleophile serine 148 to an analogous cysteine would increase the relative nucleophilicity of the protein and subsequently lead to increased acylation.

The cysteine mutant of Pik TE was generated from a pET28 plasmid encoding wild-type Pik TE by substitution of the wild type codon 148 TCC to TGT using the QuikChange site-directed mutagenesis (Stratagene) protocol. Expression and purification of Pik TE_{S148C} provided soluble protein in yields comparable to Pik TE_{WT}. A single time point assay comparing Pik TE_{WT} and TE_{S148C} provided a surprising result: Not only did Pik TE_{S148C} benefit from increased acylation with 4-nitrophenol esters, incubation with thiophenol and *N*-acetylcysteamine substrates also resulted in increased yields vs the WT enzyme (Figure 3.4). Encouraged by these initial results, we further investigated the catalytic performance of the S148C mutant by performing a reaction containing Pik TE_{S148C} with the epimerized **14.** Remarkably, Pik TE_{S148C} was able to

efficiently cyclize **14** to the corresponding epimerized macrolactone 11-epi-10-dml **16** (Scheme 3.3).
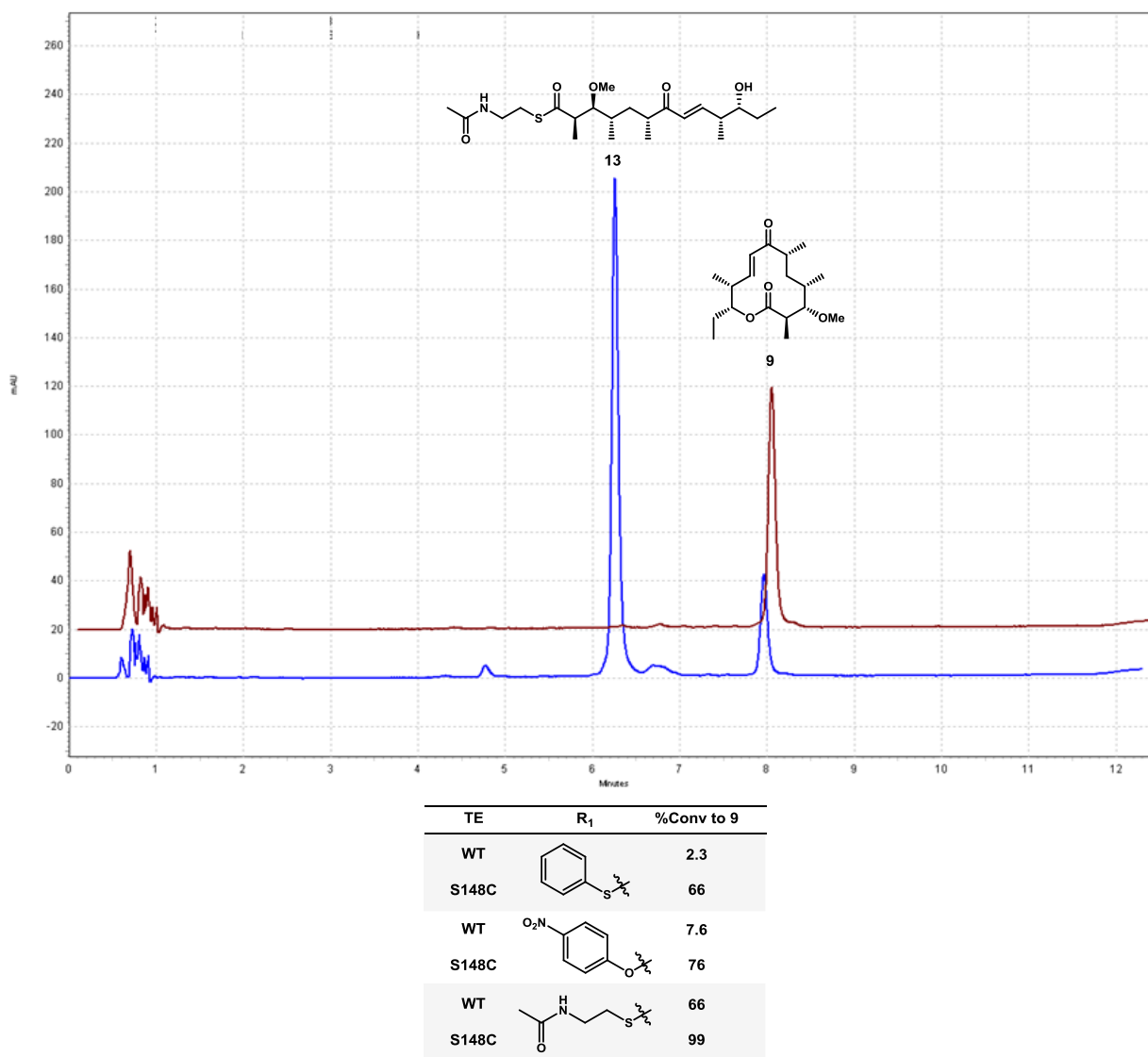
Figure 3.4 A single active site mutation in the pikromycin thioesterase generates a more effective macrocyclization catalyst. HPLC analysis of a single time point reaction demonstrates the stark catalytic advantage obtained with the S148C mutation (top) Evaluation of methyl protected Pik hexaketides with Pik TE_{S148C} (bottom) indicates this trend holds for thiophenol and 4-nitrophenol esters as well. Enzymatic reaction conditions: 1 mM Pik hexaketide, 8 mM 2-vinylpyridine, 1 mol% Pik TE_{S148C} (10 µM), 4 hours, RT. Conversion to **9** was monitored (HPLC) with data represented as the mean ± standard deviation where n = 3.

In addition to the gain-of-function macrolactonization of **14**, we also examined the kinetic effect of the S148C mutation on cyclization (Table 3.4). Steady state kinetic analysis of Pik TE_{WT} and TE_{S148C} was performed using both methyl protected Pik hexaketides **13** and **14**, as well as NBOM protected[34] native hexaketide **4,** following a
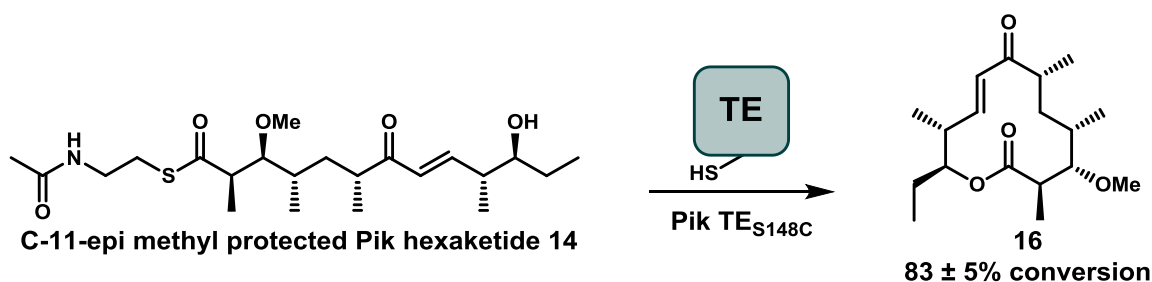
Scheme 3.3 Pik TE$_{S148C}$ displays increased substrate flexibility. Evaluation of methyl protected epimerized Pik Hexaketide with Pik TE$_{S148C}$. Enzymatic reaction conditions: 1 mM Pik hexaketide, 8 mM 2-vinylpyridine, 1 mol% Pik TE$_{S148C}$ (10 µM), 4 hours, RT. Conversion to **16** was monitored (HPLC) with data represented as the mean ± standard deviation where n = 3.

Pik hexaketide 4

methyl protected Pik hexaketide 13

C-11-epi methyl protected Pik hexaketide 14
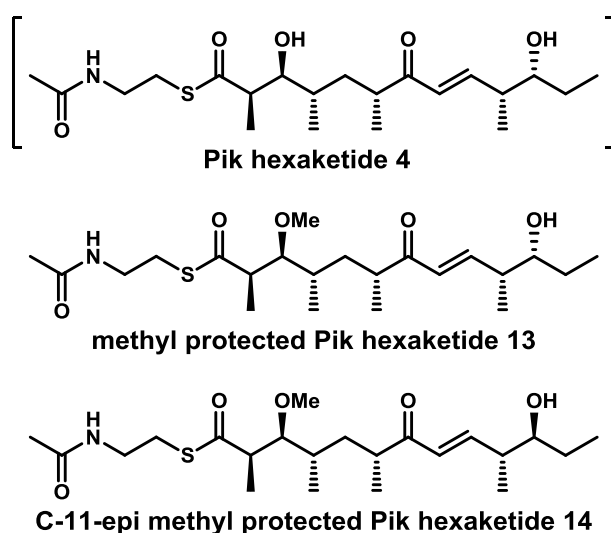
Substrate	TE	reaction	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m (mM^{-1}min^{-1})
4	WT	cyclization	101.7 ± 8.2	4.06 ± 0.68	25.1 ± 0.19
	S148C	cyclization	261.7 ± 19.6	2.40 ± 0.46	109 ± 0.21
13	WT	cyclization	11.61 ± 0.24	4.46 ± 0.18	2.60 ± 0.05
	S148C	cyclization	44.39 ± 2.01	1.42 ± 0.16	31.3 ± 0.12
14	WT	hydrolysis	8.58 ± 0.42	1.04 ± 0.16	8.25 ± 0.16
	S148C	cyclization	3.37 ± 0.09	0.07 ± 0.01	50 ± 0.4

Table 3.4 Steady state kinetic values for Pik TE$_{WT}$ and TE$_{S148C}$.

similar procedure used previously.[27] Kinetic analysis revealed that the serine to cysteine substitution afforded a superior cyclization catalyst for each substrate tested. Pik

TE$_{S148C}$ displayed a 4.3 and 12-fold increase in the k_{cat}/K_m for macrolactonization of both the native (**1**) and C-3 methoxy hexaketide **13** compared to Pik TE$_{WT}$, respectively. In reactions containing methyl protected epimerized **14**, not only did Pik TE$_{S148C}$ retain the ability to catalyze macrolactonization of the linear substrate, but notably the k_{cat}/K_m was 5.5-fold higher than Pik TE$_{WT}$ catalyzed hydrolysis. In addition to increased substrate flexibility of Pik TE$_{S148C}$, the gain-of-function mutation provided a 2.6-fold rate enhancement with the native hexaketide **4** over wild type. These kinetic parameters demonstrate that the S148C mutation of Pik TE produces a catalyst with both expanded substrate scope and increased catalytic efficiency

PikAIII-TE$_{S148C}$ with diastereomeric pentaketides

We next investigated whether the engineered Pik TE$_{S148C}$ was able to improve substrate flexibility in the context of full-module catalysis. To accomplish this, we generated PikAIII-TE$_{S148C}$ and incubated it with the diastereomeric series of Pik pentaketides reported in Chapter II. LC-HRMS analysis of reaction products revealed two new peaks in the chromatograms not observed in PikAIII-TE$_{WT}$ reactions (Figure 3.5), with masses and retention time corresponding to 12-membered ring macrolactones. As no authentic standards for the putative novel products were available, a 0.2 millimole scale reaction of **17** with PikAIII-TE$_{S148C}$ was performed and the reaction products were purified and characterized via MS and NMR. Structural determination of the reaction products confirmed that PikAIII-TE$_{S148C}$ was indeed able to generate 11-epi-10-dml **20** from **17**, as well as 3-keto-11-epi-10-dml **23**, due to failed reduction of the β-keto intermediate by the PikAIII KR domain prior to transfer to the terminal TE domain (Scheme 3.4).

Diminished relative KR activity has been previously observed in vitro by the production of both the predicted reduced and 3-keto macrolactones from reactions containing Ery5-TE with DEBS pentaketide.[56] Additionally, PikAIII-PikAIV chimeras[57] lacking a KR domain yielded exclusively 3-keto-10-dml when incubated with Pik pentaketide in vivo, indicating that the WT Pik TE domain is capable of cyclizing the C-3 keto intermediate. We next investigated if the KR-TE domain competition observed in
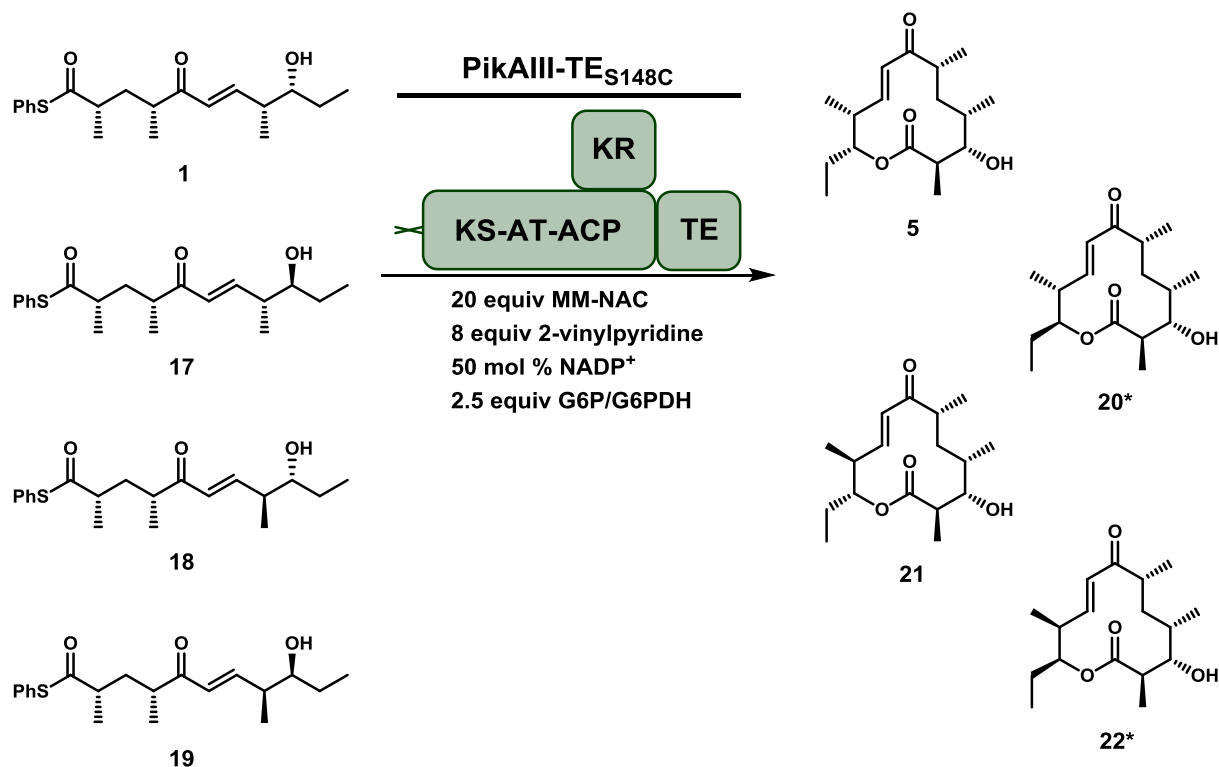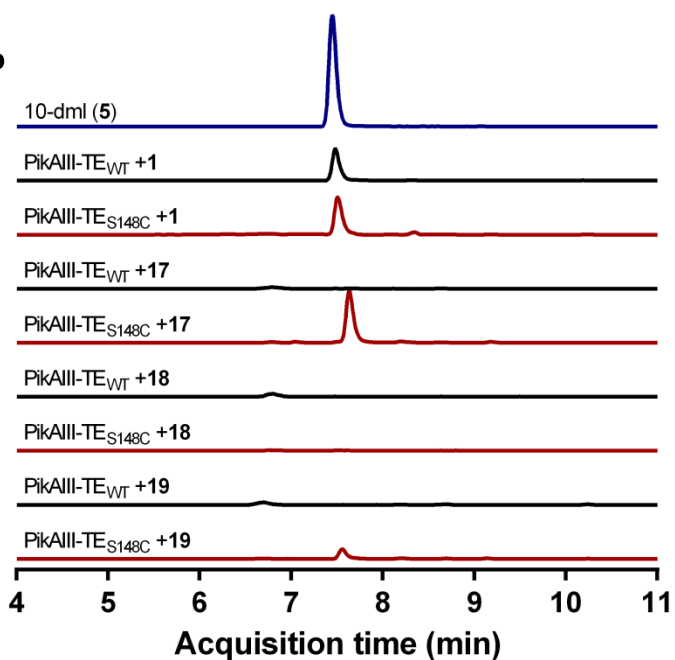
Figure 3.5 Incubation of diasteromeric pentaketides with PikAIII-TE$_{S148C}$. LC-HRMS analysis of reactions containing the stereoisomeric substrate panel. (a) Reaction scheme of stereoisomer pentaketides with predicted macrolactone products. The * indicates a species detected by LC-HRMS. (b) LC-HRMS chromatogram of reactions with PikAIII-TE$_{WT}$ (black traces) or PikAIII-TE$_{S148C}$ (red traces). Blue trace is authentic 10-dml (**5**) standard. Traces were generated by monitoring the extracted ion chromatogram of each reaction for the 12-membered macrolactone [M+H]$^+$ ion at 297.20 m/z.
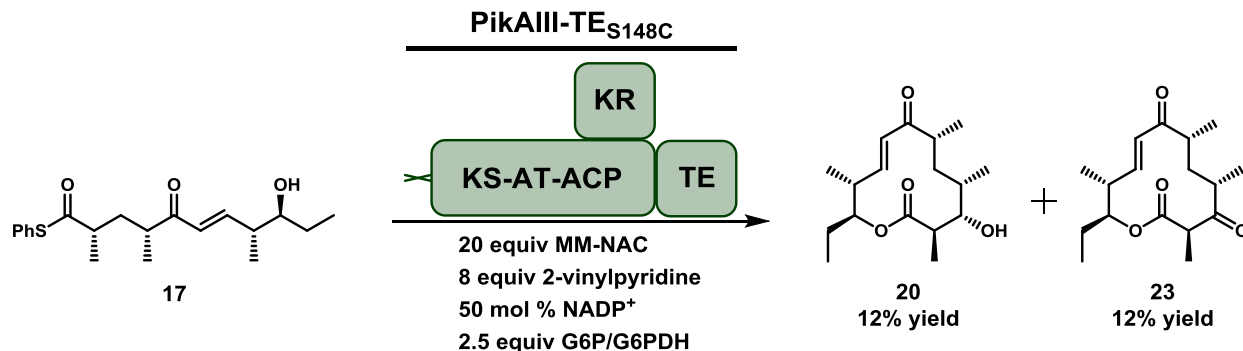
Scheme 3.4 Reaction of PikAIII-TE$_{S148C}$ with C-9-epimerized pentaketide **17**. Enzymatic reaction conditions: 1 mM Pik pentaketide, 20 mM (20 equiv) MM-NAC, 8 mM (8 equiv) 2-vinylpyridine, 0.5 mM (50 mol%) NADP+, 2.5 mM (2.5 equiv) glucose-6-phosphate, glucose-6-phosphate dehydrogenase (2 units/mL), 3 µM (0.3 mol%) PikAIII-TE, 8 hours, stationary, RT.

PikAIII-TE$_{S148C}$ reactions with **17** also occurred with the native Pik pentaketide (**1**). Accordingly, we performed analytical scale reactions containing PikAIII-TE$_{S148C}$ with its native substrate and analyzed the product distribution using synthetic standards for each product. HPLC quantification of the reaction products revealed the conversion of native pentaketide to 3-keto-10-dml **24** and 10-dml (**5**) to be 5.5% and 12.5%, respectively (Scheme 3.5). Generation of both the reduced and unreduced macrolactone products indicates competition for the linear hexaketide intermediate between the KR and TE domains even in the context of the native substrate. Isolation of **20** and **23** at 12% yields indicates that the identified thioesterase bottleneck had been alleviated by a single amino acid change (S148C) enabling substrate flux to generate novel epimerized macrolactones. The product distribution indicates competition between the KR mediated β-keto reduction and TE cyclization of the linear hexaketide
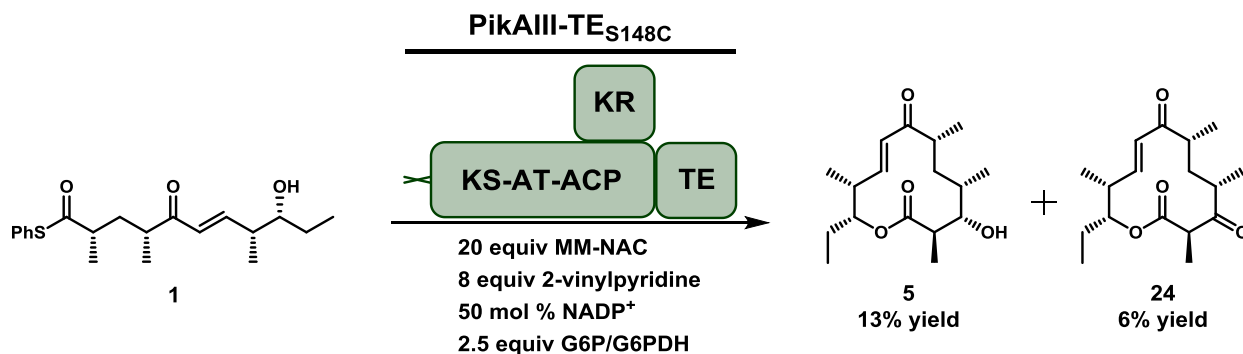
Scheme 3.5 Reaction of PikAIII-TE$_{S148C}$ with native pentaketide **1**. Enzymatic reaction conditions: 1 mM Pik pentaketide, 20 mM (20 equiv) MM-NAC, 8 mM (8 equiv) 2-vinylpyridine, 0.5 mM (50 mol%) NADP+, 2.5 mM (2.5 equiv) glucose-6-phosphate, glucose-6-phosphate dehydrogenase (2 units/mL), 3 µM (0.3 mol%) PikAIII-TE, 4 hours, stationary, RT.

59

intermediate. This competition is remarkable since the KR domain is fully competent toward processing the native substrate and no unreduced products are observed in reactions containing PikAIII-TE$_{WT}$, *suggesting that the engineered TE domain is able to outcompete the native catalytic sequence.*[58,59] While this domain competition for the hexaketide intermediate diminished the product yields for 11-epi-10-dml **20**, it provides insight into the effects of engineered domains on the sequence of catalytic events in PKS catalysis.

Experimental procedures

PKS and TE Biochemistry

Purified H_2O from a Millipore Milli-Q system with Millipore Q-Gard 2/Quantum Ex Ultrapure organex cartridges was used for all cell culture, protein purification, and enzymatic reactions. *E. coli* seed culture was grown in 15 mL sterile tubes, and subsequently grown in Corning Fernbach flasks (2.8 L) with 3x deep baffles. LB broth (Miller) and glycerol were obtained from EMD. Isopropyl-β-D-thiogalactopyranoside (IPTG) and Kanamycin (Kan) sulfate were obtained from Gold Biotechnology. Streptomycin sulfate (Strep) was obtained from AK scientific. NaCl, $CaCl_2$ and imidazole were obtained from Fisher Scientific. Lysozyme was purchased from RPI. Benzonase was purchased from Sigma Aldrich. PD-10 columns were purchased from GE scientific and equilibrated with 5 column volumes of storage buffer before use. Ni-NTA agarose resin was purchased from Qiagen and pre-equilibrated with five column volumes of lysis buffer before use.

A Symphony SB70P pH meter was calibrated according to the manufacturer's specifications and used to monitor the pH of all solutions during adjustment. Cells were lysed using a 550 Sonic Dismembrator purchased from Fisher Scientific. Optical density (OD_{600}) was determined using an Eppendorf Biophotometer. All solutions were autoclaved or sterile filtered through a 0.2 µm filter.

Buffers:
lysis: HEPES (50 mM), NaCl (300 mM), imidazole (10 mM), glycerol (10% v/v), pH 8.0.
wash: HEPES (50 mM), NaCl (300 mM), imidazole (30 mM), glycerol (10% v/v), pH 8.0.
elution HEPES (50 mM), NaCl (300 mM), imidazole (300 mM), glycerol (10% v/v), pH 8.0.
storage: HEPES (50 mM), NaCl (150 mM), EDTA (1 mM), glycerol (20% v/v), pH 7.2.
PikAIV reactions: sodium phosphate (400 mM), glycerol (20% v/v), 2-vinylpyridine (8 mM), pH 7.2.
Pik TE reactions: sodium phosphate (400 mM), 2-vinylpyridine (8 mM), pH 7.2.

Stock solutions: hexaketide substrates (50 mM in DMSO), 2-vinylpyridine (500 mM in DMSO), ascorbic acid (500 mM in H_2O), sodium metabisulfite (100 mM in H_2O), PikAIV reaction buffer [2x, sodium phosphate (800 mM), glycerol (40% v/v), pH 7.2], Pik TE reaction buffer [2x, sodium phosphate (800 mM), pH 7.2], MM-SNAC (500 mM in H_2O, neutralized to pH 7.2 with $NaHCO_3$).

Protein Expression

The cloning, expression and purification of PikAIV[60] and the Pik TE[61] has been reported previously. The cysteine mutant of Pik TE was generated from a pET28 plasmid encoding wild type Pik TE[34] by substitution of the native codon 148 TCC to TGT using the QuikChange site-directed mutagenesis (Stratagene) protocol. Identical protocols were used for the expression and purification of Pik TE_{S148C}.

A starter culture was generated by inoculating 5 mL of LB broth containing Kan (50 mg/L) with fresh transformants of E. coli (BAP1)[62] cells containing the corresponding plasmids for expression of the respective PKS proteins and grown overnight at 37 ˚C. Following the overnight growth, the entire starter culture was subsequently used to inoculate an expression culture of 1 L of TB containing Kan (50 mg/L) and grown at 37 ˚C to an OD_{600} of 0.3-0.4. The expression cultures were then cooled to 18 ˚C and growth was maintained until an OD_{600} of 0.7-0.8 was reached, at which point protein expression was induced via addition of IPTG (350 µM) and the cultures were incubated at 200 RPM at 18 ˚C for 20 hours.

Protein Purification

To retain maximum enzymatic activity, the following purification procedure was performed at 4 ˚C in less than 2 hours. Overexpression cultures were harvested by centrifugation (5,500 x g, 10 min, 4 ˚C) and cell pellets were suspended in 5 mL of lysis buffer per gram of cells via vortex. Cell lysis was accomplished by gentle agitation at 4 ˚C with 0.4 mg/ml lysozyme and 8 units/ml benzonase for 30 min followed by sonication on ice (6 x 10s with 50s rest periods). Cellular debris was pelleted by centrifugation (40,000 x g, 15 min, 4 ˚C), and the supernatant applied to 3 mL of Ni-NTA resin. After binding, the column was washed with 25 mL of wash buffer under gentle syringe

pressure and the target protein was eluted with 15 mL of <u>elution</u> buffer. Protein containing fractions were assessed via their absorption at 280 nm, pooled, and buffer exchanged into <u>storage</u> buffer using a PD-10 column. Finally, protein containing fractions were determined via their absorption at 280 nm, pooled, flash frozen in liquid N_2, and stored at -80 °C.

Enzymatic reactions

Analytical Enzymatic Reactions

All enzymatic reactions were performed in triplicate at a volume of 50 µL and were initiated via the addition of enzyme. 2-vinylpyridine (Sigma) was employed as a thiol scavenger (8 mM final concentration) in all reactions. After 4 hour stationary incubation at RT, the reactions were quenched with 3 volumes of MeOH (150 µL), clarified by centrifugation (17,000 x g, 15 min, 4 °C) and analyzed for macrolactone production. In all cases, the reactions were carried out in PikAIV or Pik TE reaction buffers.

<u>Methyl Protected Substrates:</u>

Reactions employing methylated substrates were performed as one-pot reactions containing phosphate buffer, methylated hexaketide (1 mM), with or without MM-SNAC (20 mM). Catalysis was initiated via the addition of enzyme, and incubated for 4 hours. Conversion to macrolactones was monitored by <u>Method A</u> (HPLC analysis section).

<u>NBOM Protected Substrates:</u>

Enzymatic reactions utilizing NBOM protected substrates were performed over two steps. First, a solution of ascorbic acid (25 mM final concentration), sodium metabisulfite (1 mM final concentration), NBOM protected substrate (1 mM final concentration), and H_2O (requisite dead volume) was irradiated under a consumer facial tanning lamp at a height of 14 cm (Verseo #AH129c) for 20 min to furnish the deprotected Pik hexaketide. NOTE: Irradiation through the side of the microtubes employed (Axygen #MCT-175-C) did not interfere with photolysis, and this process was reproducible over the course of this study. After photolysis, the solution was diluted with either PikAIV of Pik TE reaction buffer, MM-NAC (20 mM final concentration when

included). Catalysis was initiated via the addition of enzyme, and incubated for 4 hours. Conversion to macrolactones was monitored by <u>Method B</u> (HPLC analysis section).

HPLC analysis

Macrolactone production was monitored via analytical high performance liquid chromatography (HPLC) using a Beckman Coulter instrument (model 366 serial 385-1160) and a Zorbax SB-Phenyl 3.5 µM 4.6 x 150 mm column (part number 863953-912) at a wavelength of 250 nm.

<u>Method A:</u> For reactions employing methylated substrates, separation was accomplished by the following method: 3.0 mL/min, solvent A: H_2O, solvent B: MeCN, 20% B 0-1 min, 20-60% B linear gradient 1-10 min, 100% B 10-11 min, 20% B 11-12 min.

<u>Method B:</u> For reactions employing NBOM protected substrates, separation was accomplished by the following method: 3.0 mL/min, solvent A: H_2O, solvent B: MeCN, 10% B 0-1 min, 10-40% B linear gradient 1-10 min, 100% B 10-11 min, 10% B 11-12 min.

Samples were quantified by linear regression using equations derived from fitting the peak areas of the corresponding standard curves. Standard curves were generated by analyzing macrolactone standards in triplicate at a range of concentrations from 1.0-0.0156 mM, representing a range in percent conversions from 400-6.5%, and were linear in all cases. Conversions below 6.5% were quantified via extrapolation of the standard curve.

LC-HRMS analysis

Analytical liquid chromatography-mass spectrometry (LC-MS) was performed on an Agilent LC system (1290 series) coupled to an Agilent QTOF mass spectrometer (6500 series) using a Phenomenex Synergi 4µ Hydro RP 100 x 2 mm column (serial 48836-5) heated to 50 °C. Method: 0.4 mL/min, A: H_2O 0.1% formic acid, B: MeCN 0.1% formic acid, 0% B 0-2 min,0-100% B linear gradient 1-9 min, 100% 9-10 min, 100- 0% B 10-10.5 min linear gradient re-equilibration, 0-4 min were diverted to waste.

Large scale reactions for product characterization

Incubation of **14** with Pik TE$_{S148C}$:

Reaction conditions: sodium phosphate buffer (400 mM, pH = 7.2), hexaketide **14** (1mM), 2-vinylpyridine (8 mM), purified Pik TE (10 µM in reaction, 1 mol %), 18 hours, stationary, RT.

Workup and purification: The reaction was quenched with acetone (300 mL), filtered through a celite plug, concentrated by rotary evaporation, and extracted with EtOAc (3x). Flash chromatography employing EtOAc/hexanes (10:90) afforded compound **9** (0.03 g, 0.09 mmol, 90%).

^{1}H-NMR (599 MHz; CDCl$_3$): δ 6.59 (dd, J = 16.3, 8.0 Hz, 1H), 6.06 (d, J = 16.3 Hz, 1H), 4.48 (ddd, J = 9.6, 7.1, 5.0 Hz, 1H), 3.45 (s, 3H), 3.18 (dd, J = 8.9, 1.8 Hz, 1H), 2.77-2.65 (m, 3H), 2.08-2.02 (m, 1H), 1.80-1.75 (m, 2H), 1.62 (ddd, J = 14.3, 8.1, 5.9 Hz, 1H), 1.26 (d, J = 7.0 Hz, 3H), 1.16-1.12 (ovlp m, 1H), 1.13 (d, J = 6.7 Hz, 3H), 1.09 (d, J = 6.7 Hz, 3H), 1.07 (d, J = 6.9 Hz, 3H), 0.93 (t, J = 7.4 Hz, 3H).

^{13}C-NMR (151 MHz; CDCl$_3$): δ 204.6, 173.5, 146.8, 128.8, 86.2, 80.8, 60.2, 42.8, 42.4, 39.7, 36.6, 34.1, 25.0, 17.4, 16.2, 15.60, 15.45, 9.5

HRMS: Calculated [M+H]$^+$ 311.2217, found 311.2216.

Incubation of **17** with PikAIII-TE$_{S148C}$:

Reaction conditions: sodium phosphate buffer (300 mM, 15% v/v glycerol, 133 mL total, pH = 7.2), pentaketide **17** (48 mg, 0.13 mmol, 1 mM) MM-SNAC (20 equiv, 20 mM), NADP$^+$ (0.5 equiv, 0.5 mM), glucose-6-phosphate (2.5 equiv, 2.5 mM), glucose-6-phosphate dehydrogenase (2 units/mL), 2-vinylpyridine (8 mM), PikAIII-TE$_{S148C}$ (3 µM, 0.3 mol %), 12 hours, stationary, RT.

Workup and purification: Quenched with acetone (2x volume, 260mL), placed in a -20 °C freezer for 1 hour and filtered through a celite plug. Remaining insoluble material was suspended in acetone and this solution was used to rinse the celite plug. Acetone was removed through rotary evaporation and the aqueous layer was saturated with NaCl and extracted 3x EtOAc. Combined organic layers were washed with brine and filtered through a sodium sulfate plug then rinsed 2x with EtOAc and concentrated. 11

and 12 were purified directly by preparatory HPLC using a Phenomenex Luna C18, 5 um, 250 x 21.2 mm column (serial 444304-4) monitoring at 250 nM. Method 9 mL/min, A: H_2O 0.1% formic acid, B: MeCN 0.1% formic acid, 25% B 0-5 min, 25-40% B linear gradient 5-25 min, 60% B 25-30 min, 60-90% B linear gradient 30-50 min,100% B 50-65 min, 25% B 65-75 min.

20 from pentaketide **17** (4.8 mg, 0.016 mmol, 12% yield).
^{1}H NMR (700 MHz; CD$_3$OD): δ 6.39 (dd, J = 15.9, 9.2 Hz, 1H), 6.19 (d, J = 15.9 Hz, 1H), 4.44 -4.38 (m, 1H), 3.51 (dd, J = 9.5, 2.0 Hz, 1H), 2.73-2.64 (m, 2H), 2.53-2.47 (m, 1H), 2.19-2.14 (m, 1H), 1.93-1.91 (m, 2H), 1.83-1.75 (m, 2H), 1.49-1.43 (m, 1H), 1.27 (d, J = 7.0 Hz, 3H), 1.08 (d, J = 6.7 Hz, 3H), 1.06 – 1.03 (m, 6H), 0.93 (t, J = 7.4 Hz, 3H).
^{13}C NMR (700 MHz; CD$_3$OD): δ 207.3, 176.2, 147.9, 131.5, 83.2, 77.5, 46.4, 43.6, 42.7, 37.2, 36.3, 25.9, 17.4, 17.3, 17.0, 16.3, 10.4.
HRMS: Calculated [M+H]+ 297.2060, found 297.2046

23 from pentaketide **17** (4.8 mg, 0.016 mmol, 12% yield).
^{1}H NMR (700 MHz; CD$_3$OD): δ 6.43 (dd, J = 15.6, 9.1 Hz, 1H), 6.28 (d, J = 15.6 Hz, 1H), 4.74 (td, J = 8.9, 3.2 Hz, 1H), 3.87 (q, J = 6.8 Hz, 1H), 3.07-3.02 (m, 1H), 2.77-2.71 (m, 1H), 2.59-2.53 (m, 1H), 1.97–1.84 (m, 3H), 1.66-1.58 (m, 1H), 1.20 (d, J=6.8 Hz, 3H), 1.17 (d, J = 7.2 Hz, 3H), 1.08 (d, J = 6.8 Hz, 3H), 1.06 (d, J = 6.7 Hz, 3H), 0.94 (t, J = 7.4 Hz, 3H).
^{13}C NMR (700 MHz; CD$_3$OD): δ 209.1, 206.1, 171.3, 148.8, 132.1, 80.1, 52.6, 47.2, 45.1, 42.6, 38.7, 26.2, 18.1, 17.8, 16.1, 14.3, 9.6.
HRMS: Calculated [M+H]+ 295.1904, found 295.1902

Kinetic analysis of thioesterase catalysis
Product formation for all kinetic experiments was quantified by linear regression using equations derived from fitting the peak areas of the corresponding standard curves. Standard curves were generated by analyzing product standards in triplicate at a range of concentrations and were linear in all cases. Prior to performing kinetic analysis, the

pH dependence for TE macrolactonization was established by running a series of assays at pH 6, 6.5, 7, 7.2, 7.5, 7.8, 8. The reactions consisted of sodium phosphate buffer (50 mM), 1 µM Pik TE (WT or S148C), 1 mM 4 and 10% (v/v) DMSO in a total volume of 50 µL. The reactions were incubated at RT for 1 h and quenched by addition of MeOH (150 µL). All reactions were run in duplicate and following centrifugal clarification (17,000 x g, 15 min, 4°C) the samples were directly analyzed for product formation, pH 7.2 was found to be optimum.

Macrolactonization of 4:

Macrolactonization of the native hexaketide **4** was monitored via analytical high performance liquid chromatography (HPLC) using an Agilent 1100 Series HPLC and a Zorbax SB-Phenyl 3.5 µm, 4.6 x 150 mm (part number 863953-912) column at a wavelength of 250 nm. Products were separated using a linear gradient via the following method: 3.0 mL/min, solvent A: H2O, solvent B: MeCN, 10% B 0-1 min, 10-40% B linear gradient 1-10 min, 100% B 10-11 min, 10% B 11-12 min.

A time course for the cyclization reaction with each TE was accomplished by incubating assay mixtures containing 400 mM phosphate buffer (pH 7.2), 10% (v/v) DMSO, 0.125 mM **4**, and 0.3 µM TE_{S148C} or 1.0 µM TE_{WT} in a total reaction volume of 350 µL at room temperature for intervals of 5, 10, 20, and 30 min, at which point they were quenched and analyzed via HPLC. The NBOM protected substrate was deprotected and utilized in enzymatic reactions as previously reported[34], briefly: a solution of ascorbic acid (25 mM final concentration), sodium metabisulfite (1 mM final concentration), NBOM protected hexaketide (1 mM final concentration), and H_2O (requisite dead volume) was irradiated under a consumer facial tanning lamp at a height of 14 cm (Verseo #AH129c) for 20 min to furnish the deprotected Pik hexaketide **4**. After photolysis, the solution was diluted with enzymatic reaction buffer, and catalysis was initiated via the addition of enzyme. Reactions were run in duplicate and the velocities were determined to be linear by linear regression analysis of the data over a period of 20 min.

Steady-state kinetic analysis was performed by determining the initial velocities for the macrolactonization of **4** at concentrations of 0.125 mM, 0.25 mM, 0.5 mM, 1 mM, 2 mM, 3 mM, 4 mM, and 8 mM. The reactions consisted of 400 mM phosphate buffer

(pH 7.2), 10% (v/v) DMSO, variable concentrations of **4**, and 0.3 µM TE$_{S148C}$ or 1.0 µM TE$_{WT}$. The reactions were incubated at room temperature for 10 min at which point they were quenched and analyzed via HPLC for macrolactone production. All reactions were run in triplicate and the Initial velocities at different substrate concentrations were fit to the Michaelis-Menten equation by nonlinear least-squares regression to calculate k_{cat} and K_M.

Macrolactonization of 13:

Macrolactonization of SNAC-MeHK **13** to **9** was monitored via analytical high performance liquid chromatography (HPLC) using an Agilent 1100 Series HPLC and a Zorbax SB-Phenyl 3.5 µm, 4.6 x 150 mm (part number 863953-912) column at a wavelength of 250 nm. Products were separated using a linear gradient via the following method: 3.0 mL/min, solvent A: H$_2$O, solvent B: MeCN, 20% B 0-1 min, 20-60% B linear gradient 1-10 min, 100% B 10-11 min, 20% B 11-12 min.

A time course for the macrolactonization reaction of each TE at pH 7.2 was obtained by analyzing reaction mixtures containing 400 mM phosphate buffer (pH 7.2), 10% (v/v) DMSO, 0.2 mM **13**, and 0.3 µM TE$_{S148C}$ or 1.0 µM TE$_{WT}$ in a total reaction volume of 1500 µL incubated at room temperature for intervals of 5, 10, 20, and 30 min at which point they were quenched and analyzed via HPLC for macrolactone production. Reactions were run in duplicate and the velocities were determined to be linear by linear regression analysis of the data over a period of 20 min.

Steady-state kinetic analysis was performed by determining the initial velocities for the macrolactonization of **13** at concentrations of 0.25 mM, 0.5 mM, 1 mM, 2 mM, 3 mM, 4 mM, 6 mM, and 8 mM. The reactions consisted of 400 mM phosphate buffer (pH 7.2), 10% (v/v) DMSO, variable concentrations of **13**, and 0.3 µM TE$_{S148C}$ or 1.0 µM TE$_{WT}$. The reactions were incubated at room temperature for 10 min at which point they were quenched with MeOH (150 µL) and analyzed via HPLC for macrolactone production. All reactions were run in triplicate and the initial velocities at different substrate concentrations were fit to the Michaelis-Menten equation by nonlinear least-squares regression to calculate k_{cat} and K_M. The 6 mM and 8 mM concentration points were

68

omitted from the plots with TE_{S148C} as these concentrations resulted in significant substrate inhibition.

Hydrolysis and macrolactonization of 14:

The hydrolysis (TE_{WT}) and macrolactonization (TE_{S148C}) of **14** to either **9** or **15**, respectively, was monitored via HPLC using the same method as above for **13**. A time course for the hydrolysis and cyclization reaction for each TE was accomplished by incubating assay mixtures containing 400 mM phosphate buffer (pH 7.2), 10% (v/v) DMSO, 0.03 mM (TE_{S148C}) or 0.125 mM (TE_{WT}) **14**, and 0.3 µM TE_{S148C} or 1.0 µM TE_{WT} in a total reaction volume of 1500 µL at room temperature for intervals of 5, 10, 20, and 30 min, at which point they were quenched and analyzed via HPLC. Reactions were run in duplicate and the velocities were determined to be linear by linear regression analysis of the data over a period of 30 min.

Steady-state kinetic analysis was performed by determining the initial velocities for the macrolactonization and hydrolysis of **14** at concentrations of 0.25 mM, 0.5 mM, 1 mM, 2 mM, 3 mM, 4 mM, 6 mM, and 8 mM for TE_{WT} and 0.03 mM, 0.06 mM, 0.125 mM, 0.25 mM, 0.5 mM, 1 mM, 2 mM, 3 mM, 4 mM, 6 mM, and 8 mM for TE_{S148C}. The reactions consisted of 400 mM phosphate buffer (pH 7.2), 10% (v/v) DMSO, variable concentrations of **5**, and 0.5 µM TE_{S148C} or 1.0 µM TE_{WT}. The reactions were incubated at room temperature for 10 min at which point they were quenched and analyzed via HPLC for product formation. All reactions were run in triplicate and the Initial velocities at different substrate concentrations were fit to the Michaelis-Menten equation by nonlinear least-squares regression to calculate k_{cat} and K_M. The concentration points not represented were removed from the plot as they resulted in significant substrate inhibition.

References

(1)   Driggers, E. M.; Hale, S. P.; Lee, J.; Terrett, N. K. *Nat. Rev. Drug. Discov.* **2008**, *7*, 608.

(2)   Marne, G.; Alexander, S. M. *Curr. Top. Med. Chem.* **2003**, *3*, 949.

(3)   Mankin, A. S. *Curr. Opin. Microbiol.* **2008**, *11*, 414.

(4)   Bulkley, D.; Innis, C. A.; Blaha, G.; Steitz, T. A. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 17158.

(5)   Woodward, R. B. *Angew. Chem.* **1957**, *69*, 50.

(6)   Horsman, M. E.; Hari, T. P. A.; Boddy, C. N. *Nat. Prod. Rep.* **2016**, *33*, 183.

(7)   Parenty, A.; Moreau, X.; Niel, G.; Campagne, J. M. *Chem. Rev.* **2013**, *113*, PR1.

(8)   Kopp, F.; Marahiel, M. A. *Nat. Prod. Rep.* **2007**, *24*, 735.

(9)   Akey, D. L.; Kittendorf, J. D.; Giraldes, J. W.; Fecik, R. A.; Sherman, D. H.; Smith, J. L. *Nat. Chem. Biol.* **2006**, *2*, 537.

(10)   Giraldes, J. W.; Akey, D. L.; Kittendorf, J. D.; Sherman, D. H.; Smith, J. L.; Fecik, R. A. *Nat. Chem. Biol.* **2006**, *2*, 531.

(11)   McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Betlach, M.; Ashley, G. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 1846.

(12)   Tang, L.; Fu, H.; McDaniel, R. *Chem. Biol.* **2000**, *7*, 77.

(13)   Reeves, C. D.; Murli, S.; Ashley, G. W.; Piagentini, M.; Hutchinson, C. R.; McDaniel, R. *Biochemistry* **2001**, *40*, 15464.

(14)   Ruan, X.; Pereda, A.; Stassi, D. L.; Zeidner, D.; Summers, R. G.; Jackson, M.; Shivakumar, A.; Kakavas, S.; Staver, M. J.; Donadio, S.; Katz, L. *J. Bacteriol.* **1997**, *179*, 6416.

(15)   Liu, L.; Thamchaipenet, A.; Fu, H.; Betlach, M.; Ashley, G. *J. Am. Chem. Soc.* **1997**, *119*, 10553.

(16)   Stassi, D. L.; Kakavas, S. J.; Reynolds, K. A.; Gunawardana, G.; Swanson, S.; Zeidner, D.; Jackson, M.; Liu, H.; Buko, A.; Katz, L. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 7305.

(17)   Xue, Q.; Ashley, G.; Hutchinson, C. R.; Santi, D. V. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 11740.

(18)   Marsden, A. F. A.; Wilkinson, B.; Cortés, J.; Dunster, N. J.; Staunton, J.; Leadlay, P. F. *Science* **1998**, *279*, 199.

(19)   McDaniel, R.; Kao, C. M.; Fu, H.; Hevezi, P.; Gustafsson, C.; Betlach, M.; Ashley, G.; Cane, D. E.; Khosla, C. *J. Am. Chem. Soc.* **1997**, *119*, 4309.

(20)   Kao, C. M.; McPherson, M.; McDaniel, R. N.; Fu, H.; Cane, D. E.; Khosla, C. *J. Am. Chem. Soc.* **1997**, *119*, 11339.

(21)   Kao, C. M.; Luo, G.; Katz, L.; Cane, D. E.; Khosla, C. *J. Am. Chem. Soc.* **1995**, *117*, 9105.

(22)   Jacobsen, J. R.; Hutchinson, C. R.; Cane, D. E.; Khosla, C. *Science* **1997**, *277*, 367.

(23)   Gokhale, R. S.; Hunziker, D.; Cane, D. E.; Khosla, C. *Chem. Biol.* **1999**, *6*, 117.

(24)   Aggarwal, R.; Caffrey, P.; Leadlay, P. F.; Smith, C. J.; Staunton, J. *J. Chem. Soc., Chem. Commun.* **1995**, 1519.

(25)   Boddy, C. N.; Schneider, T. L.; Hotta, K.; Walsh, C. T.; Khosla, C. *J. Am. Chem. Soc.* **2003**, *125*, 3428.

(26)   Aldrich, C. C.; Venkatraman, L.; Sherman, D. H.; Fecik, R. A. *J. Am. Chem. Soc.* **2005**, *127*, 8910.

(27)   He, W.; Wu, J.; Khosla, C.; Cane, D. E. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 391.

(28)   Wang, M.; Boddy, C. N. *Biochemistry* **2008**, *47*, 11793.

(29)   Heberlig, G. W.; Wirz, M.; Wang, M.; Boddy, C. N. *Org. Lett.* **2014**, *16*, 5858.

(30)   Hari, T. P.; Labana, P.; Boileau, M.; Boddy, C. N. *Chembiochem* **2014**, *15*, 2656.

(31)   Kudo, F.; Kitayama, T.; Kakinuma, K.; Eguchi, T. *Tetrahedron Lett.* **2006**, *47*, 1529.

(32)   Wang, M.; Zhou, H.; Wirz, M.; Tang, Y.; Boddy, C. N. *Biochemistry* **2009**, *48*, 6288.

(33)   Pinto, A.; Wang, M.; Horsman, M.; Boddy, C. N. *Org. Lett.* **2012**, *14*, 2278.

(34)   Hansen, D. A.; Koch, A. A.; Sherman, D. H. *J. Am. Chem. Soc.* **2015**, *137*, 3735.

(35)   Aldrich, C. C.; Beck, B. J.; Fecik, R. A.; Sherman, D. H. *J. Am. Chem. Soc.* **2005**, *127*, 8441.

(36)   Aldrich, C. C.; Venkatraman, L.; Sherman, D. H.; Fecik, R. A. *J. Am. Chem. Soc.* **2005**, *127*, 8910.

(37)   Mortison, J. D.; Kittendorf, J. D.; Sherman, D. H. *J. Am. Chem. Soc.* **2009**, *131*, 15784.

(38)   Harris, T. M.; Harris, C. M. *Tetrahedron* **1977**, *33*, 2159.

(39)   Crosby, J.; Crump, M. P. *Nat Prod Rep* **2012**, *29*, 1111.

(40)   Whicher, J. R.; Dutta, S.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Dosey, A. M.; Narayan, A. R. H.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. *Nature* **2014**, *510*, 560.

(41)   Schwartz, M. E.; Breaker, R. R.; Asteriadis, G. T.; deBear, J. S.; Gough, G. R. *Bioorg. Med. Chem. Lett.* **1992**, *2*, 1019.

(42)   Hansen, D. A.; Rath, C. M.; Eisman, E. B.; Narayan, A. R.; Kittendorf, J. D.; Mortison, J. D.; Yoon, Y. J.; Sherman, D. H. *J. Am. Chem. Soc.* **2013**, *135*, 11232.

(43)   Cane, D. E.; Yang, C. C. *J. Am. Chem. Soc.* **1987**, *109*, 1255.

(44)   Yue, S.; Duncan, J. S.; Yamamoto, Y.; Hutchinson, C. R. *J. Am. Chem. Soc.* **1987**, *109*, 1253.

(45)   Nonhydrolyzable NAC and CoA analogs have been employed previously in type I PKS systems, see: a) Tosin, M.; Betancor, L.; Stephens, E.; Ariel Li, W. M.; Spencer, J. B.; Leadlay, P. F. ChemBioChem 2010, 11, 539. b) Spiteller, D.; Waterman, C. L.; Spencer, J. B. Angew. Chem., Int. Ed. 2005, 44, 7079.

(46)   Wu, J.; He, W.; Khosla, C.; Cane, D. E. *Angew. Chem., Int. Ed.* **2005**, *44*, 7557.

(47)   Kittendorf, J. D.; Beck, B. J.; Buchholz, T. J.; Seufert, W.; Sherman, D. H. *Chemistry and Biology* **2007**, *14*, 944.

(48)   Yin, Y.; Lu, H.; Khosla, C.; Cane, D. E. *J. Am. Chem. Soc.* **2003**, *125*, 5671.

(49)   Hansen, D. A.; Rath, C. M.; Eisman, E. B.; Narayan, A. R.; Kittendorf, J. D.; Mortison, J. D.; Yoon, Y. J.; Sherman, D. H. *J. Am. Chem. Soc.* **2013**, *135*, 11232.

(50)   Pohl, N. L.; Gokhale, R. S.; Cane, D. E.; Khosla, C. *J. Am. Chem. Soc.* **1998**, *120*, 11206.

(51)   Broadhurst, R. W.; Nietlispach, D.; Wheatcroft, M. P.; Leadlay, P. F.; Weissman, K. J. *Chem. Biol.* **2003**, *10*, 723.

(52)   Aggarwal, R.; Caffrey, P.; Leadlay, P. F.; Smith, C. J.; Staunton, J. *J. Chem. Soc. Chem. Commun.* **1995**, 1519.

(53)   Hansen, D. A.; Koch, A. A.; Sherman, D. H. *J. Am. Chem. Soc.* **2017**, *139*, 13450.

(54)   Farooqui, A. A.; Taylor, W. A.; Pendley, C. E.; Cox, J. W.; Horrocks, L. A. *J. Lipid Res.* **1984**, *25*, 1555.

(55)   Schaumann, E. In *Sulfur-Mediated Rearrangements I*; Schaumann, E., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2007, p 1.

(56)   Mortison, J. D.; Kittendorf, J. D.; Sherman, D. H. *J. Am. Chem. Soc.* **2009**, *131*, 15784.

(57)   Chemler, J. A.; Tripathi, A.; Hansen, D. A.; O'Neil-Johnson, M.; Williams, R. B.; Starks, C.; Park, S. R.; Sherman, D. H. *J. Am. Chem. Soc.* **2015**, *137*, 10603.

(58)   Dutta, S.; Whicher, J. R.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Congdon, G. R.; Narayan, A. R.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. *Nature* **2014**, *510*, 512.

(59)   Whicher, J. R.; Dutta, S.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Dosey, A. M.; Narayan, A. R.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. *Nature* **2014**, *510*, 560.

(60)   Beck, B. J.; Aldrich, C. C.; Fecik, R. A.; Reynolds, K. A.; Sherman, D. H. *J. Am. Chem. Soc.* **2003**, *125*, 4682.

(61)   Lu, H.; Tsai, S.-C.; Khosla, C.; Cane, D. E. *Biochemistry* **2002**, *41*, 12590.

(62)   Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C. *Science* **2001**, *291*, 1790.

Chapter IV

Structural and mechanistic insights into type I PKS thioesterase catalysis

Portions of this chapter have been published and are reproduced in part with permission from:

Koch, A. A.; Hansen, D. A.; Shende, V. V.; Furan, L. R.; Houk, K. N.; Jiménez-Osés, G.; Sherman, D. H. *J. Am. Chem. Soc.* **2017**, *139*, 13456.

Copyright 2017 American Chemical Society.

GJO contributed to Figure 4.11-4.13.

VS contributed to Scheme 4.1.

AK contributed to all others.

Introduction

Pik TE belongs to the α,β-hydrolase class of serine hydrolases and contains the characteristic Ser-His-Asp catalytic triad.[1] X-ray crystallographic studies have shown Pik TE functions as a dimer and the catalytic triad (Ser148, His268, Asp176) lies at the center of a long open substrate channel that spans the entire width of the enzyme[2] (Figure 4.1). In vivo, Pik TE receives its substrate as a thioester tethered to the phosphopantetheinyl arm of the ACP domain from the upstream module. It is believed that the substrate entrance site is located at the N-terminal portion of the protein, while the product exits the C-terminal side. Structural studies of Pik TE trapped as an acyl-enzyme with covalent phosphonate affinity labels have provided further insights into the Pik TE catalytic mechanism.[3,4] By solving the 1.8-Å crystal structure of Pik TE in complex with the substrate mimic, Akey et al. report the existence of a "hydrophilic barrier" of water molecules that form at the protein exit site and serve to direct the hydrophobic acyl chain back into the substrate channel toward the location of nucleophilic attack. [4] Of note, there were remarkably few direct enzyme-substrate polar contacts and since the apo- and affinity label bound structures were essentially unchanged, it is unlikely that the protein utilizes an induced-fit mechanism for substrate recognition. Thus while these studies provided significant insight into the Pik TE catalytic mechanism, many questions remain as to how Pik TE discriminates among candidate substrates and what are the structural parameters that govern macrolactonization over hydrolysis after formation of the acyl-enzyme intermediate.

Trapping the TE acyl-enzyme intermediate

Following the remarkable results with Pik TE$_{S148C}$ from Chapter III, we were motivated to explore the structural and mechanistic parameters that govern macrolactonization in Pik TE. Intrigued by the stark catalytic divergence from macrolactonization to hydrolysis in Pik TE reactions containing a hexaketide bearing the epimerized C-11 hydroxyl group (Scheme 3.2), we pursued x-ray crystallographic studies of Pik TE to better understand the structural parameters that produce this
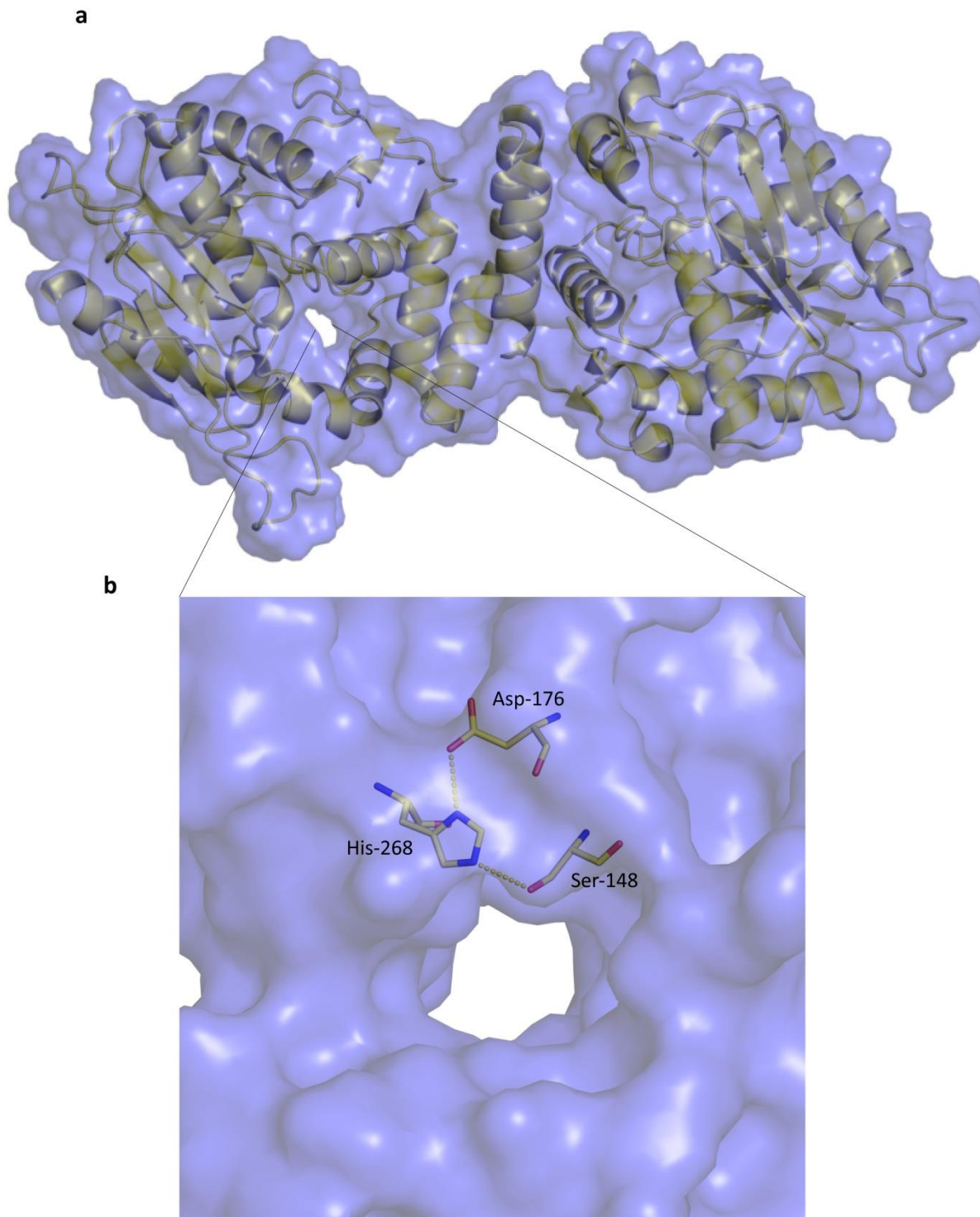
Figure 4.1 Overall structure of the Pik TE dimer. (a) Surface (blue) and cartoon (yellow) representation of the global Pik TE crystal structure. (b) Magnified view of the Pik TE substrate channel traversing the entirety of the protein, with the residues comprising the catalytic triad colored yellow. Structures were generated from Protein Data Bank entry 1MN6.[2]

divergent catalysis. At the initiation of this project, we reasoned that trapping Pik TE as an acyl-enzyme intermediate with its full length native substrate as opposed to affinity label mimics would provide the best insight into the mechanisms that govern productive catalysis.

A literature search of related protein co-crystal structures revealed a promising strategy for trapping substrates previously employed in fatty acyl-CoA thioesterase II.[5] In this scheme, Witkowski et al. successfully converted a hydrolase to an acyl-transferase through the mutation of two active site residues. The mutations performed included a Ser to Cys and a His to Arg substitution of the native residues within the catalytic triad. We speculated these mutations were able to stabilize the acyl-enzyme intermediate through the following general mechanism: Substitution of the native serine to a cysteine generates an active site residue with increased nucleophilicity and decreased pKa, resulting in a faster rate of acylation. Mutation of the His residue responsible for the activation of water during substrate hydrolysis is critical for stabilizing the thioester intermediate after formation of the acyl-enzyme intermediate (Figure 4.2). Since PKS TE domains share an analogous catalytic triad, we expected that this strategy could be successfully applied to Pik TE.

To pursue this strategy, we performed the analogous mutations in Pik TE (S148C and H268R) and expressed this mutant for in vitro analysis. Unfortunately, we found that Pik $TE_{S148CH268R}$ expressed predominantly in the insoluble protein fraction and was recalcitrant to purification. In order to facilitate in vitro analysis, we fused Pik $TE_{S148CH268R}$ with a solubilizing maltose binding protein (MBP) tag at the N-terminus. The MBP-tagged protein expressed strongly in the soluble cell lysate fraction and following purification with TEV mediated removal of the MBP tag, we obtained soluble Pik $TE_{S148CH268R}$ in quantities suitable for initial testing (Figure 4.3). Purified Pik $TE_{S148CH268R}$ was tested for enzymatic activity using conditions previously employed with Pik TE_{WT} (Chapter III) and found to be inactive (Figure 4.4), indicating that the mutation strategy successfully eliminated catalytic turnover of the enzyme.

Since the single mutant Pik TE_{S148C} behaved comparably to the wildtype protein, we reasoned that the cause for the Pik $TE_{S148CH268R}$ insoluble expression was due to the H268R substitution and thus generated a series of mutants at this position to test for
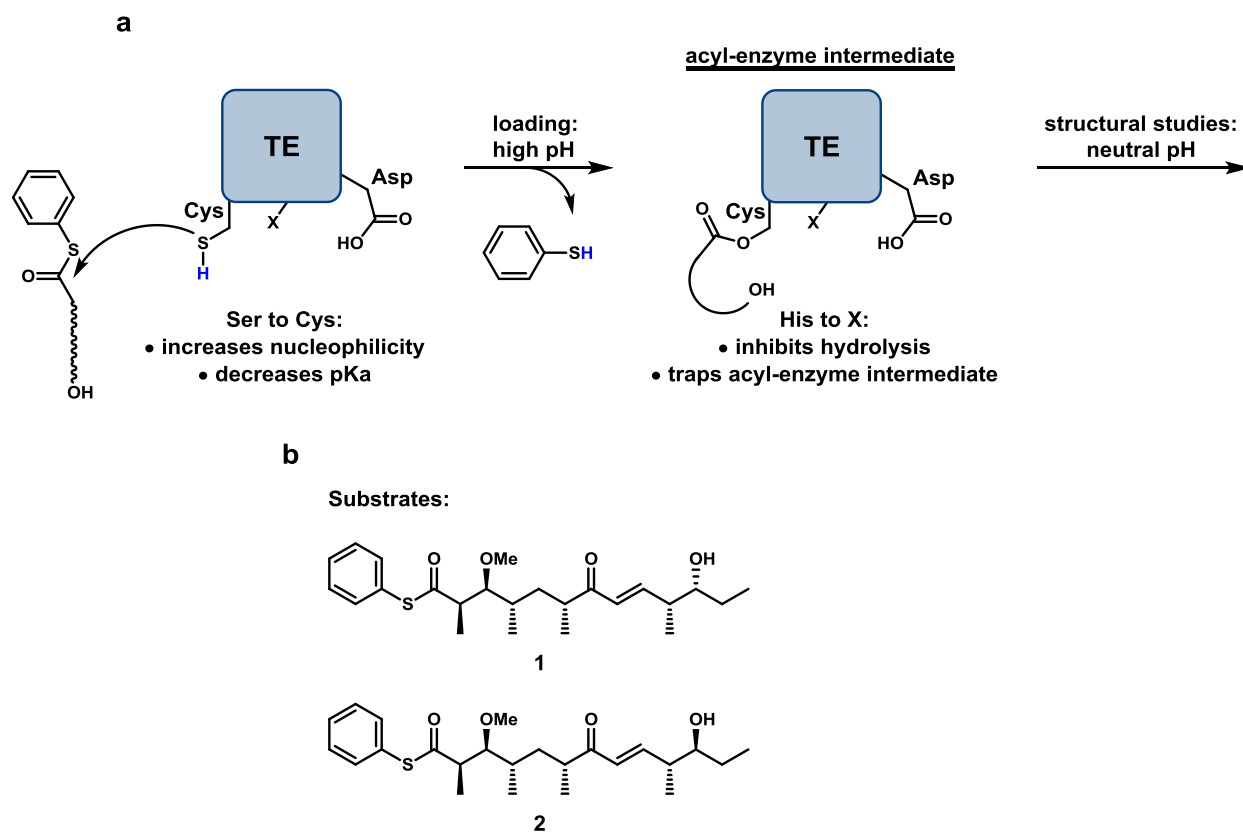
Figure 4.2 General mechanism for trapping the TE acyl-enzyme intermediate through mutation of the catalytic triad. (b) full length substrates for structural studies of Pik TE.

increased solubility. We performed mutations at position 268 to include Gly, Ala, Asn, and Gln and ultimately Pik $TE_{S148CH268Q}$ provided the highest soluble expression and was purified readily as a standalone protein (Figure 4.5). We next screened conditions suitable for performing the substrate labelling reactions with Pik $TE_{S148CH268Q}$. We anticipated a high pH buffer would increase the efficiency of the acylation reaction as the cysteine nucleophile would be in its more nucleophilic thiolate form. An extensive survey of buffers, additives, temperature, and time of incubation yielded optimized conditions suitable for achieving approximately 80% labelling after overnight incubation (Figure 4.6). The stable formation of Pik TE in complex with its native substrate demonstrates the success of this mutation strategy for trapping type I PKS TE domains as acyl-enzyme intermediates for downstream structural and functional studies. After optimizing reaction conditions, we performed large scale labeling experiments followed by gel filtration chromatography in efforts to provide material suitable for X-ray crystallography. Unfortunately, we have been unable to form crystals from the acyl-

enzyme intermediate to date and are actively pursuing broad screening conditions for crystallization. Current studies are also focused on applying these techniques to homologous TE domains with native and non-native substrates to generate material suitable for X-ray crystallography. If successful, these X-ray crystallography studies should provide significant insight into the structural and mechanistic parameters that govern macrolactonization in PKS TE domains.
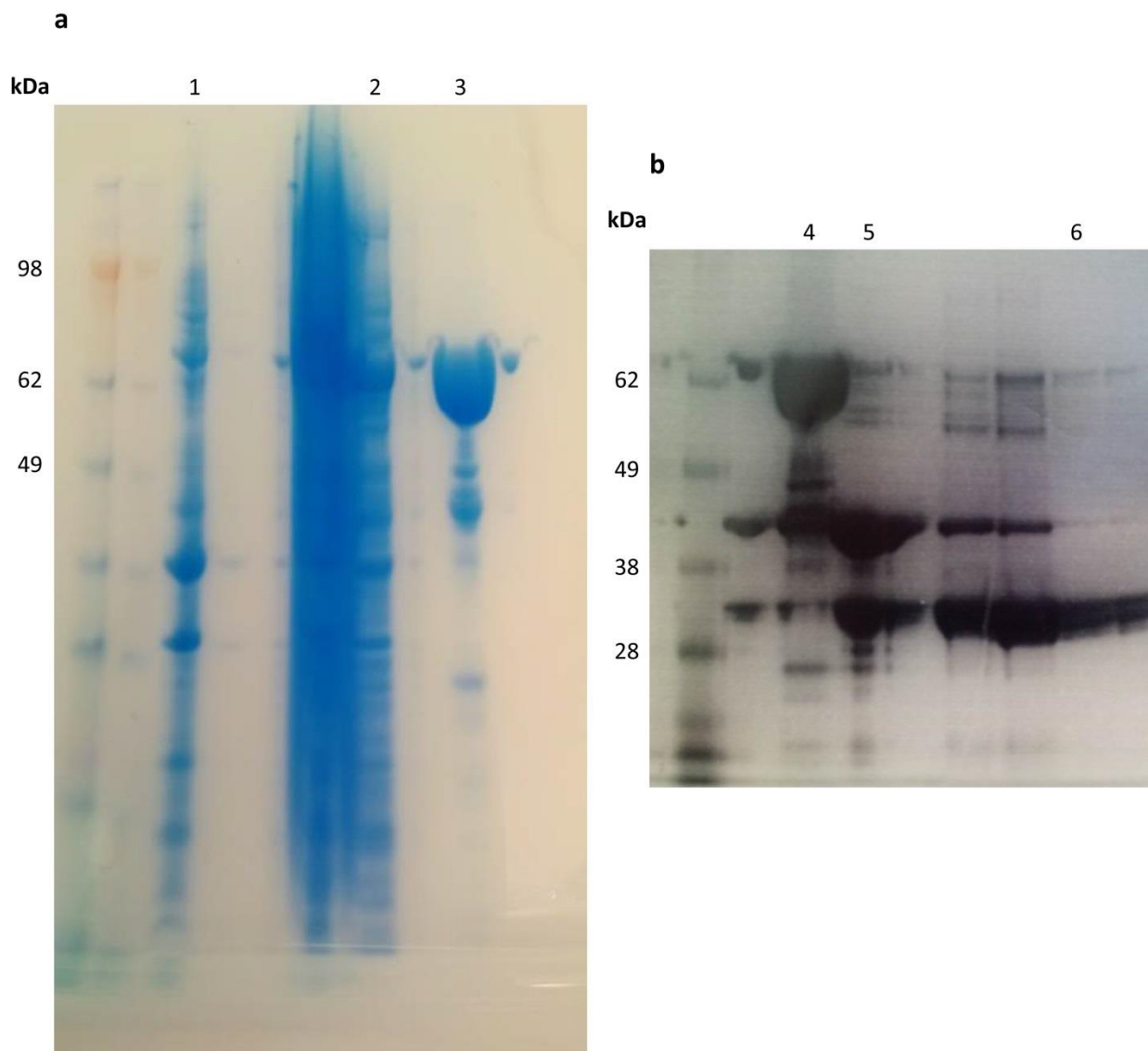
Figure 4.3 Expression and purification of MBP-Pik $TE_{S148CH268R}$ for substrate trapping studies. (a) Analysis of BL21 (DE3) E. coli expression cultures showing the soluble production of MBP-Pik $TE_{S148CH268R}$ (75 kDa). (b) TEV mediated MBP tag removal and subsequent isolation of untagged Pik $TE_{S148CH268R}$ (32 kDa). Lanes: 1. Insoluble cell lysate. 2. Soluble cell lysate. 3/4. Ni-NTA purification of the soluble cell lysate from Lane 2. 5. Purified protein after 8 hour TEV cleavage. 6. Removal of MBP tag and isolation of untagged Pik $TE_{S148CH268R}$.
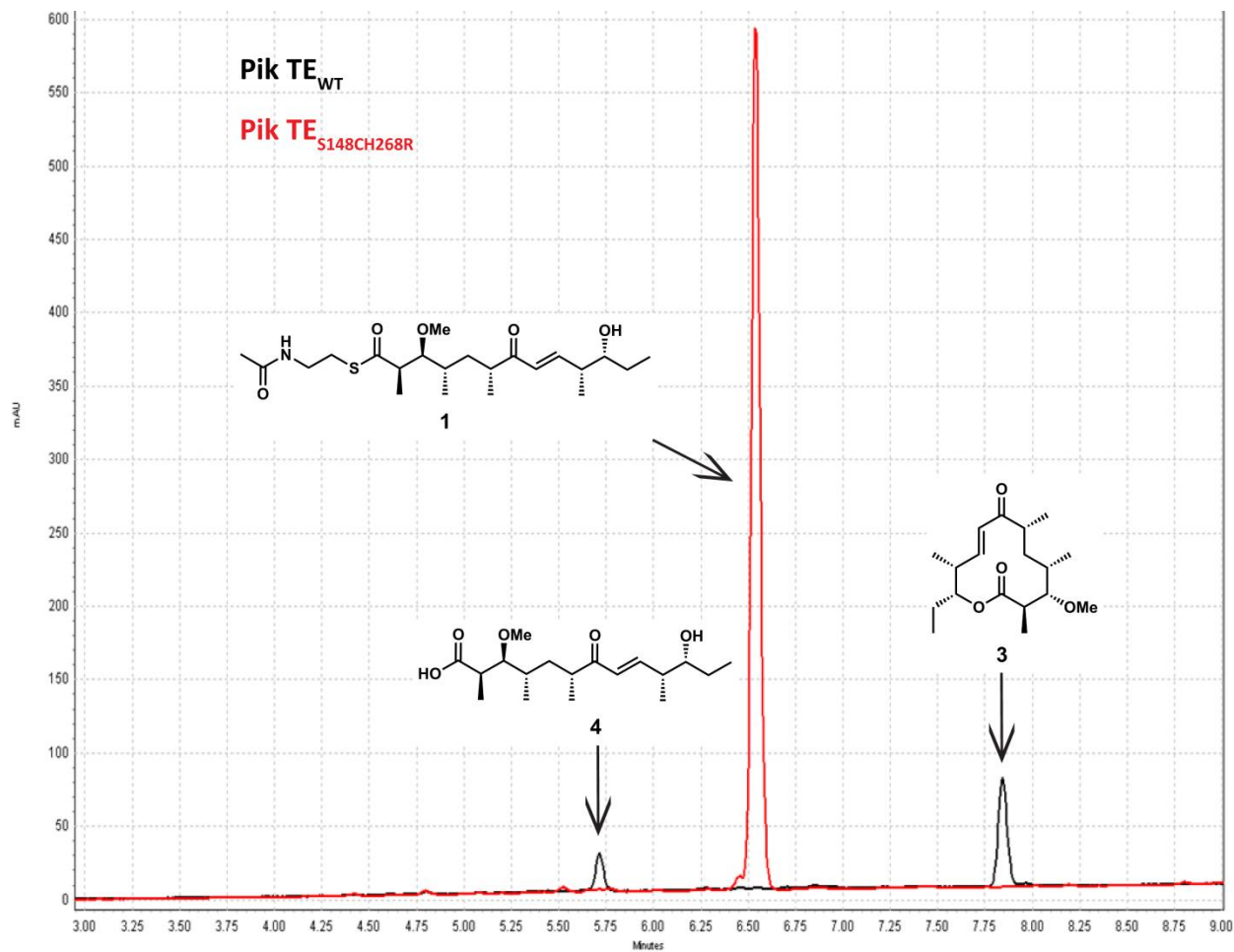
Figure 4.4 In vitro analysis confirms that Pik TE$_{S148CH268R}$ is catalytically inactive. Reactions containing **1** with Pik TE$_{WT}$ (black trace) produce the cyclization and hydrolysis products **3** and **4**, respectively. Incubation with Pik TE$_{S148CH268R}$ (red trace) fails to produce either product, indicating the protein is unable to turn over the substrate. Enzymatic reaction conditions: 1 mM Pik hexaketide, 8 mM 2-vinylpyridine, 1 mol% Pik TE (10 µM), 4 hours, RT.
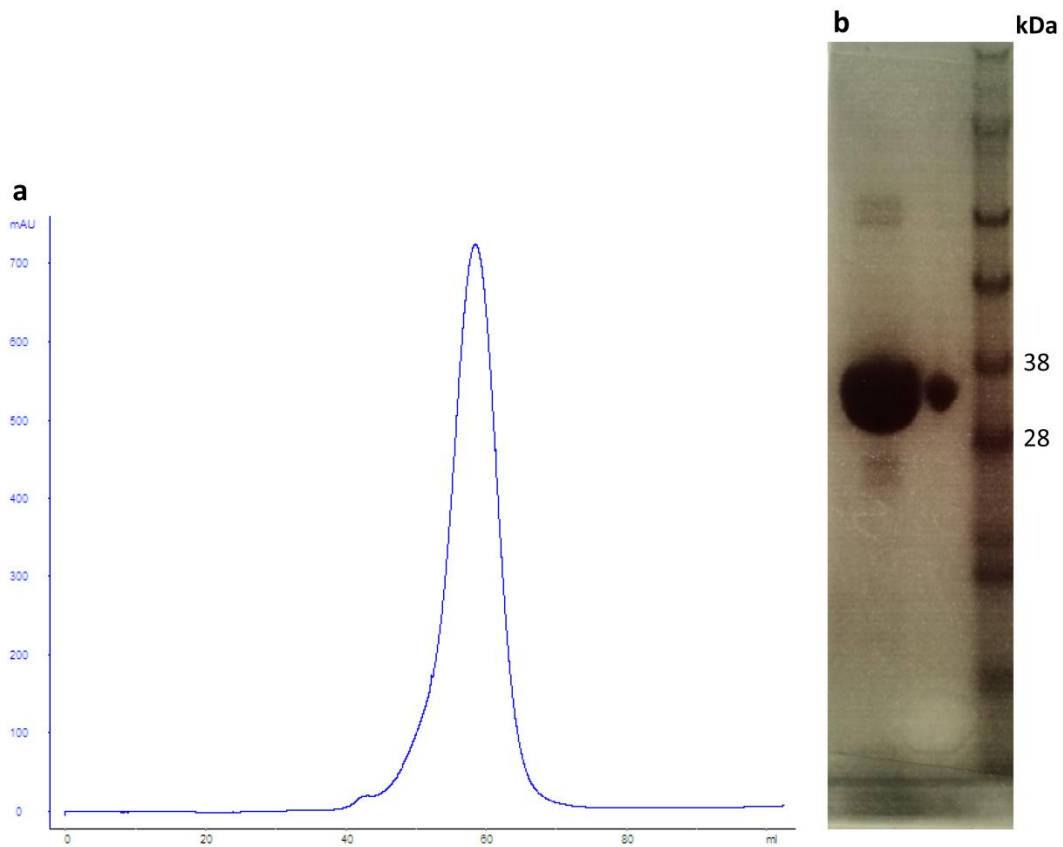
Figure 4.5 Pik TE$_{S148CH268Q}$ is more amenable to in vitro analysis than TE$_{S148CH268R}$. (a) Gel filtration chromatogram from the purification of Pik TE$_{S148CH268Q}$ showing a homogeneous, monodisperse protein. (b) SDS-polyacrylamide gel electrophoresis of the gel filtration purified TE$_{S148CH268Q}$ (34 kDa).
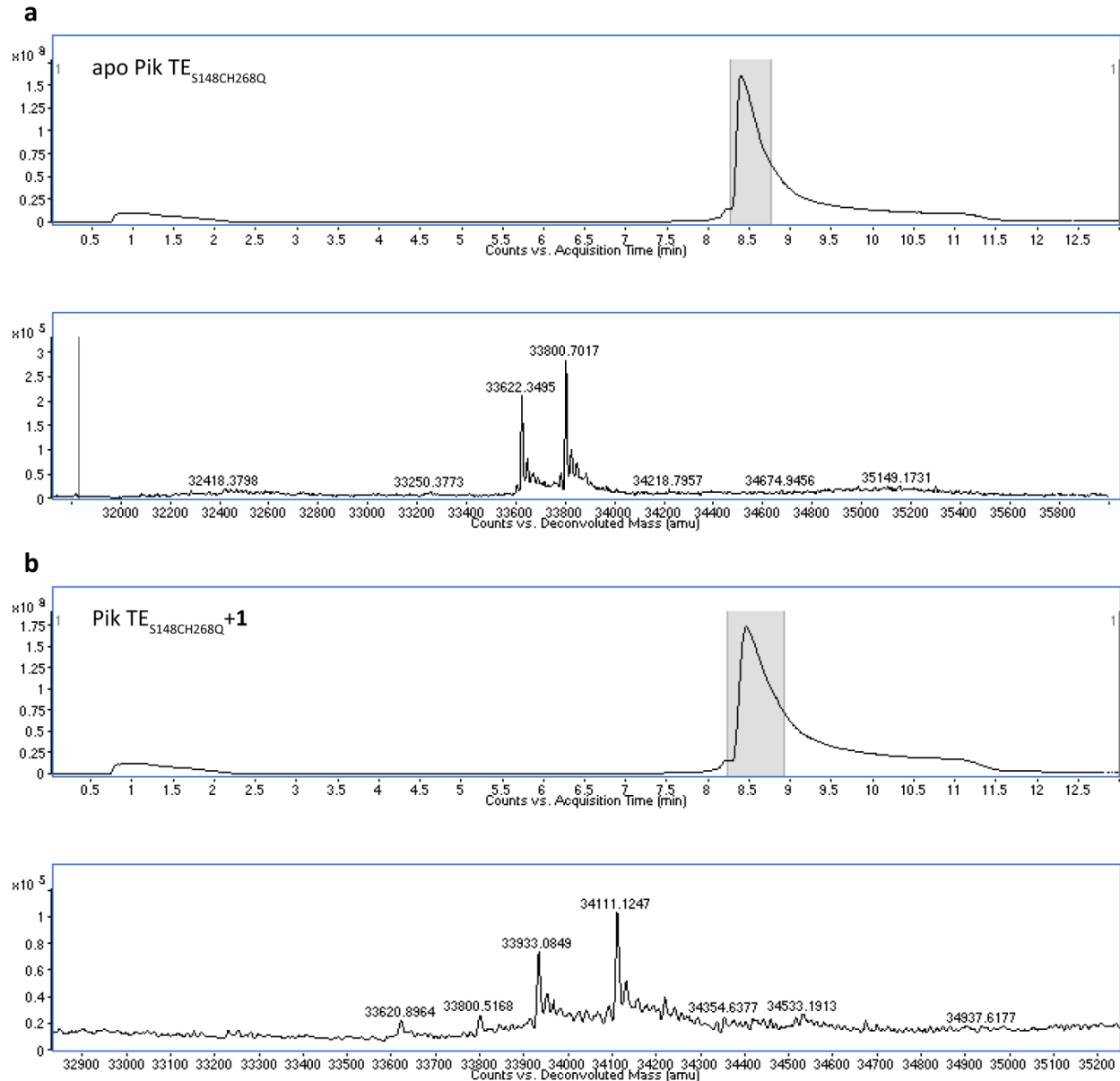
Figure 4.6 LC-HRMS analysis of Pik TE$_{S148CH268Q}$ substrate labelling reactions. The total ion chromatograms are displayed above and the deconvoluted mass below. (a) Overnight incubation with DMSO yields the apoprotein with calculated a mass of 33753 kDa observed with loss of the N-terminal methionine (33622 kDa) and phosphogluconylation[21] (33800 kDa). (b) Addition of the C-3 methyl protected Pik hexaketide **1** yields the acyl-enzyme intermediate with a calculated mass increase of 311 Da at approximately 80% incorporation.

MD simulations

In addition to pursuing X-ray crystallography, we envisioned that performing molecular dynamics (MD) simulations of Pik TE acylated with hexaketides **1** and **2** could provide structural insight into our experimentally observed catalytic divergence. Accordingly, we performed MD simulations of Pik TE_{WT} modeled as acyl-enzyme intermediates bound to the C-3 methyl protected hexaketides (denoted Pik TE-**1** and Pik TE-**2**, respectively, Figure 4.7), using the conformation of the derivatized products[6] as the starting arrangement for the individual substrates. MD simulations were initiated with the hexaketide C-11 alcohol constrained in a reactive conformation by imposing a maximum distance restraint of 2.3 Å between the hexaketide nucleophilic hydroxyl hydrogen and the Nε nitrogen of His268. After 50 ns, the distance restraint was removed and the simulations were allowed to continue for 500 ns.
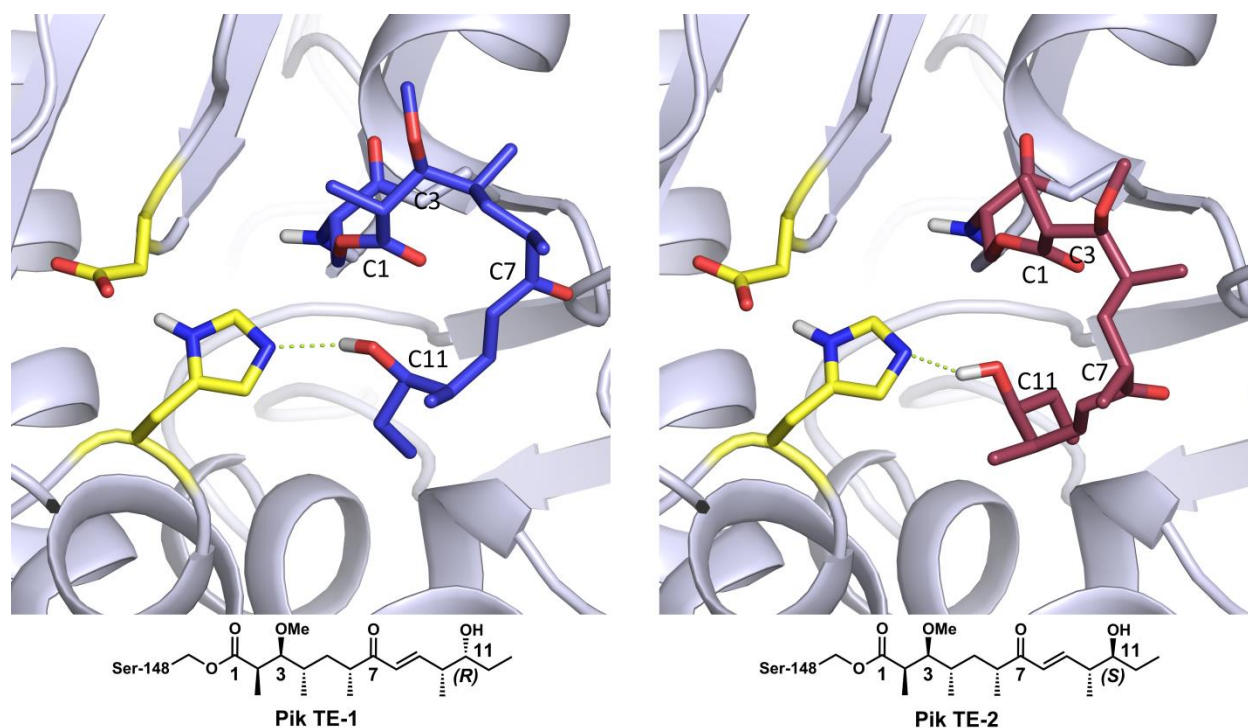
Figure 4.7 Acyl-enzyme starting structures for the MD simulations. Input structures for Pik TE modeled with **1** (Pik TE-**1**) and C-11-epimerized **2** (Pik TE-**2**).

Clustering analysis of the resulting simulations based on the conformation of the acyl-enzyme intermediate was used to bin each frame into representative clusters of structures. The clusters were subsequently examined for hydrogen-bond and hydrophobic interactions involving the substrate and TE using LigPlot+.[7] Examination of

the overall protein root-mean-square deviation (RMSD) and per-residue root-mean-square fluctuation (RMSF) values for the trajectories of each TE complex demonstrated stable dynamic behavior throughout the course of the simulations, with Pik TE-**2** maintaining a slightly increased level of flexibility (Figure 4.8). Investigation of the Pik TE-**2** residues that displayed higher RMSF values revealed these residues were located predominantly in loops and linker regions with higher intrinsic flexibility and had limited impact on the conformation of the hexaketide.
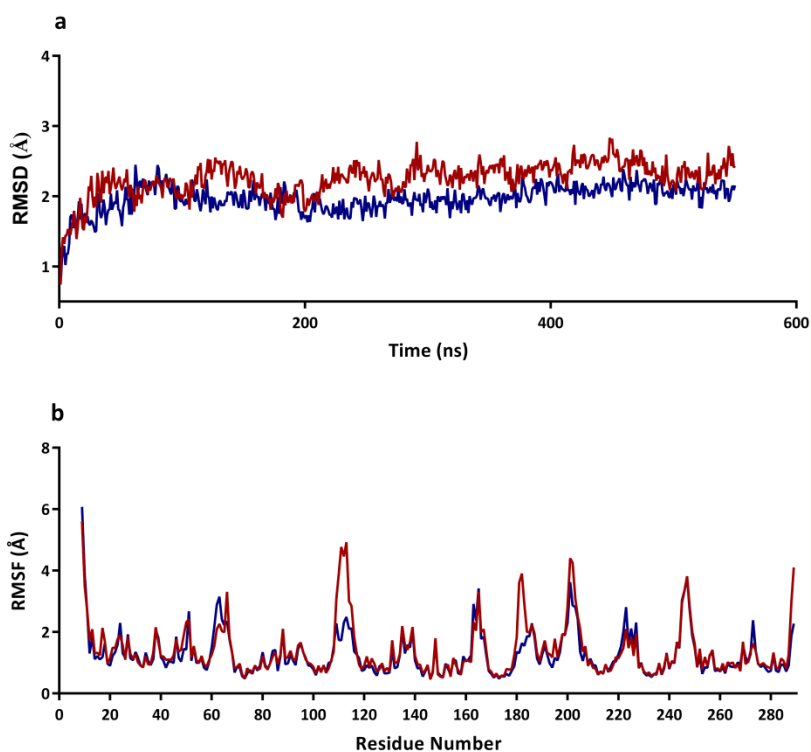
Figure 4.8 Examination of the overall structural flexibility of the TE domain during the MD simulations. The RMSD (a) and RMSF (b) values for the 550 ns production MD simulations of Pik TE-**1** (blue) and Pik TE-**2** (red).

Simulations of Pik TE-**1** revealed that while the initial catalytic restraint was in place, the native hexaketide readily adopts two main conformations favorable towards macrolactonization (denoted as I and II, Figure 4.9a). At the initial stages of the simulation, conformation I predominates, accompanied with a high level of shape complementarity with the TE active site. The main interactions constituting conformation I are hydrophobic contacts between the hexaketide and the side chains of residues lining the active site. Additionally, a hydrogen bond between the hexaketide C-7
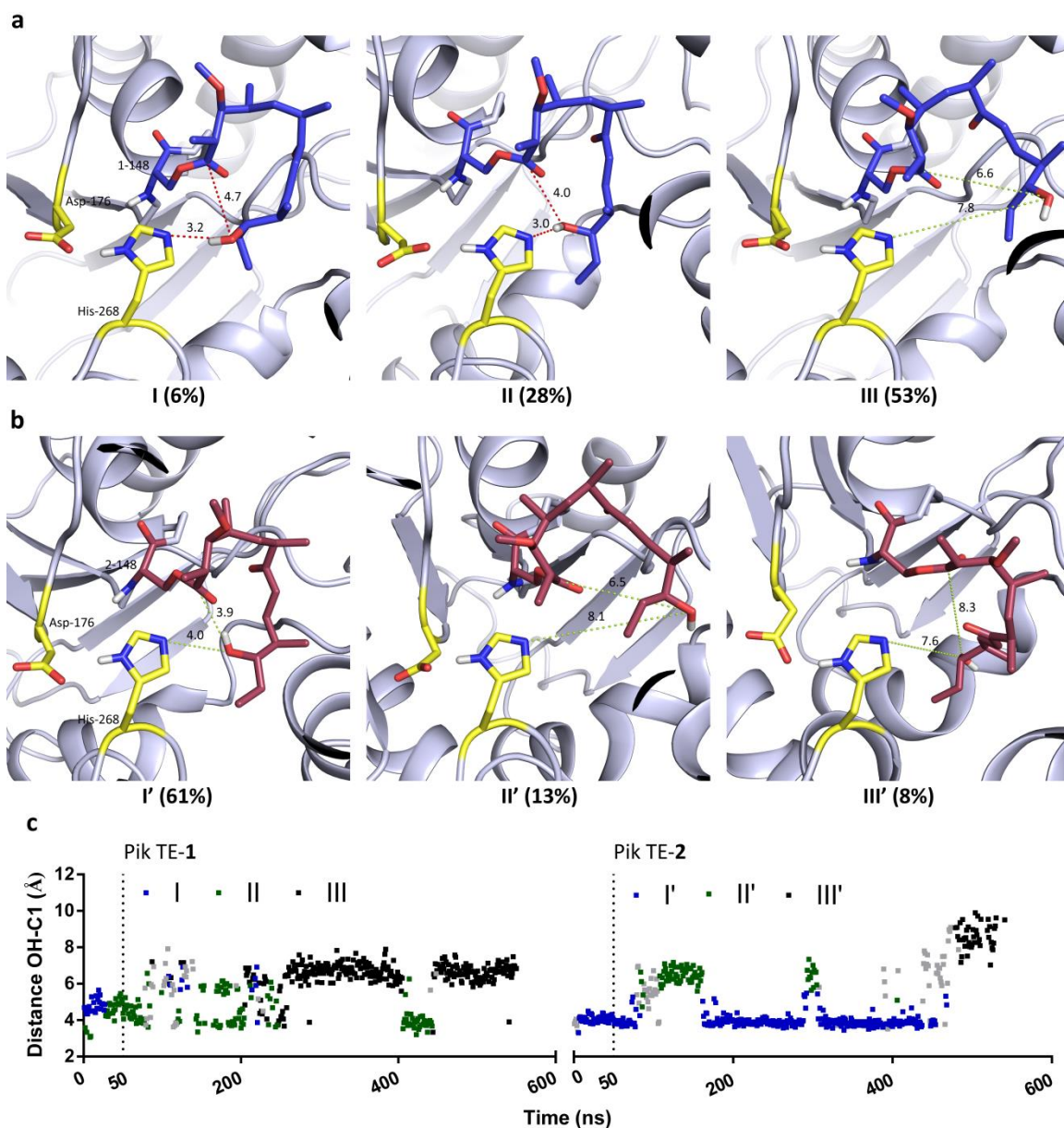
Figure 4.9 Comparison of the reactive conformations for each acyl-enzyme intermediate obtained from clustering analysis of MD simulations with Pik TE$_{WT}$. (a) Pik TE-**1** and (b) Pik TE-**2.** Pik TE-**1** conformations **I** and **II** contain a hexaketide orientation most conducive to macrolactonization with the C-11 OH in close proximity to both His268 and the C-1 carbonyl. The corresponding conformation (cluster **I'**) in the Pik TE-**2** simulation likely represents a larger barrier to macrolactonization as the distance between the C-11 OH and His268 has increased and the resulting geometry hinders deprotonation. The Pik TE-**2** hexaketide continues to evolve towards a linear conformation till the final cluster **III'** is reached which places the C-11 OH distal to both His268 and the C-1 carbonyl and in an orientation susceptible to hydrolysis. The catalytic triad His268 and Asp176 residues are colored yellow. For each conformation the distance in angstroms from the nucleophilic hydroxyl oxygen to the Nε nitrogen of His268 and the ester C-1 carbonyl is displayed above the dashed lines. Clusters containing catalytically productive conformations contain red dashed lines. (c) The distance of the nucleophilic hydroxyl oxygen and the ester C-1 carbonyl plotted for each frame of the MD simulation with each data point colored according to the corresponding clustered conformation. The vertical dashed line at 50 ns indicates when the distance constraints were released.

carbonyl and the side chain of Thr77 facilitates stabilization of the cyclic, reactive conformation (Figure 4.10). After ~30 ns the hexaketide evolves to conformation II through a substrate tail rotation that places the C-11 alcohol 0.2 Å and 0.7 Å closer to His268 and the C-1 carbonyl, respectively. This orientation resembles the macrolactonization transition state (*vide infra*) even closer than conformation I and is likely catalytically productive.

Following formation of conformation **II**, the hexaketide displays a large amount of conformational freedom within the TE active site. In the most prevalent conformation (**III**, Figure 4.9a) hydrophobic packing is largely reduced and non-productive hydrogen bonds between the C-11 alcohol and Tyr25/Leu193 are formed (Figure 4.10). Loss of the key hydrophobic interactions is detrimental to macrolactonization as it increases the conformational space accessible to the hexaketide, which adopts a more linear, unreactive conformation. Although conformation **III** accounts for the majority of the Pik TE-**1** simulation, conformation **II** developed again later in the trajectory (406 to 443 ns, Figure 4.9c) suggesting that productive conformations for macrocyclization are frequently sampled despite the high substrate flexibility.

The simulations of Pik TE-**2** involving the C-11-epimerized hexaketide revealed a significantly different scenario. While the initial distance restraint was in place, the hexaketide adopted a cyclic conformation stabilized by hydrophobic packing with the TE active site (**I'**, Figure 4.9b). However, in contrast with the native hexaketide, the C-11 alcohol of the epimerized hexaketide is prone to intramolecular hydrogen bonding with the C-1 carbonyl oxygen. Despite maintaining the substrate in a cyclic conformation with the nucleophilic hydroxyl in close proximity to the C-1 carbonyl, the resulting geometry impedes macrolactonization. Furthermore, His268 remains 4.0 Å away and positioned at an angle suboptimal for deprotonation, making this conformation catalytically unproductive. After ~80 ns, unrestricted Pik TE-**2** transitions to the second most prevalent conformation that is primarily defined by interruption of the intramolecular hydrogen-bond and rotation of the hexaketide tail away from His268 resulting in poor shape complementarity with the TE active site and a more linear conformation of the hexaketide (**II'**, Figure 4.9). These two unproductive conformations are alternately sampled for the majority of the simulation (~470 ns). Afterwards, the hexaketide chain
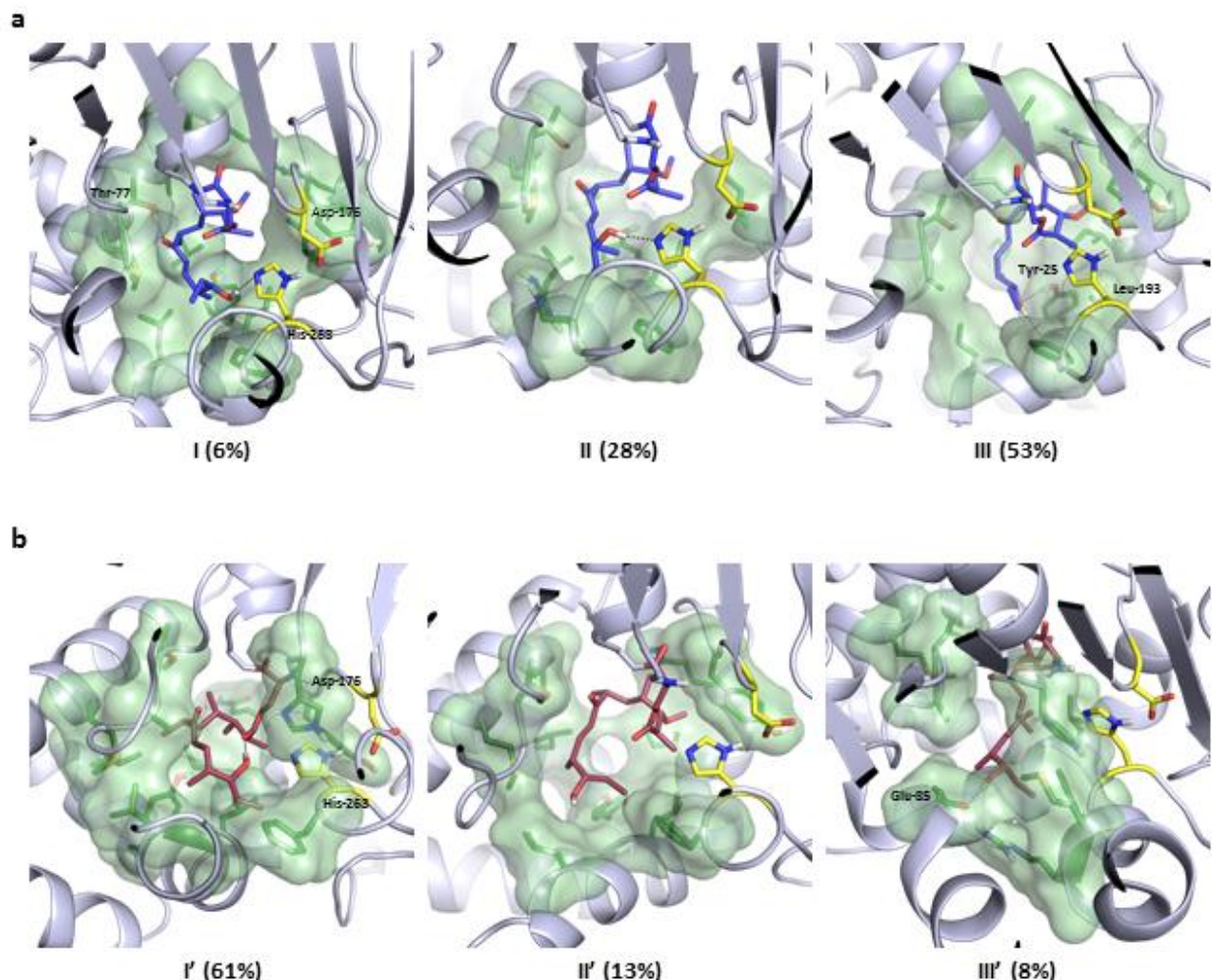
Figure 4.10 The TE-hexaketide contacts identified from the clustering analysis of the Pik TE$_{WT}$ MD simulations. (a) Pik TE-**1** and (b) Pik TE-**2**. Pik TE-**1** clusters I and II stabilize the hexaketide in a catalytically productive conformation mainly by hydrophobic packing leading to a high level of shape-complementarity and hydrogen bonding between the nucleophilic hydroxyl and the catalytic histidine. Conversely, cluster III shows decreased hydrophobic contacts and the hexaketide adopts a more linear conformation making it unreactive towards macrolactonization. Pik TE-**5** Cluster I' stabilizes the hexaketide in a cyclic conformation via hydrophobic packing; however the hexaketide adopts a conformation that impedes macrocyclization due to formation of an intramolecular hydrogen bond. Clusters II' and III' show a gradual loss of hydrophobic packing as the hexaketide adopts a linear conformation that makes it vulnerable to the experimentally observed hydrolysis. The catalytic triad His268 and Asp176 residues are colored yellow while the active site residues found to interact with the acyl-intermediate are colored green. Residues involved in hydrophobic interactions are displayed in surface representation; dashed lines indicated hydrogen bonds to the labeled residues.

evolves towards even more extended conformations (**III'**, Figure 4.9) characterized by a gradual loss of hydrophobic interactions with the protein, and formation of unproductive hydrogen bonds between the C-11 alcohol and Glu85 (Figure 4.10). This completely impedes macrolactonization and increases the likelihood competing hydrolysis of the acyl-enzyme intermediate will occur.

We next compared the frequency of productive conformations sampled in the MD simulations by calculating quantum mechanically (QM) optimized macrolactonization transition structures (TS) for each hexaketide (*vide infra*). We used these models to describe the precise angle of nucleophilic attack (O11-C1-O1) and distance (O11-C1) for macrolactonization of each hexaketide. The resulting geometric values were then compared to those extracted from each frame of the corresponding MD simulations and the entire trajectories were plotted according to their deviation from the ideal TS values (Figure 4.11). Consistent with our clustering analysis of the hexaketide conformations, comparison of the geometric deviations from ideal within each simulation revealed that
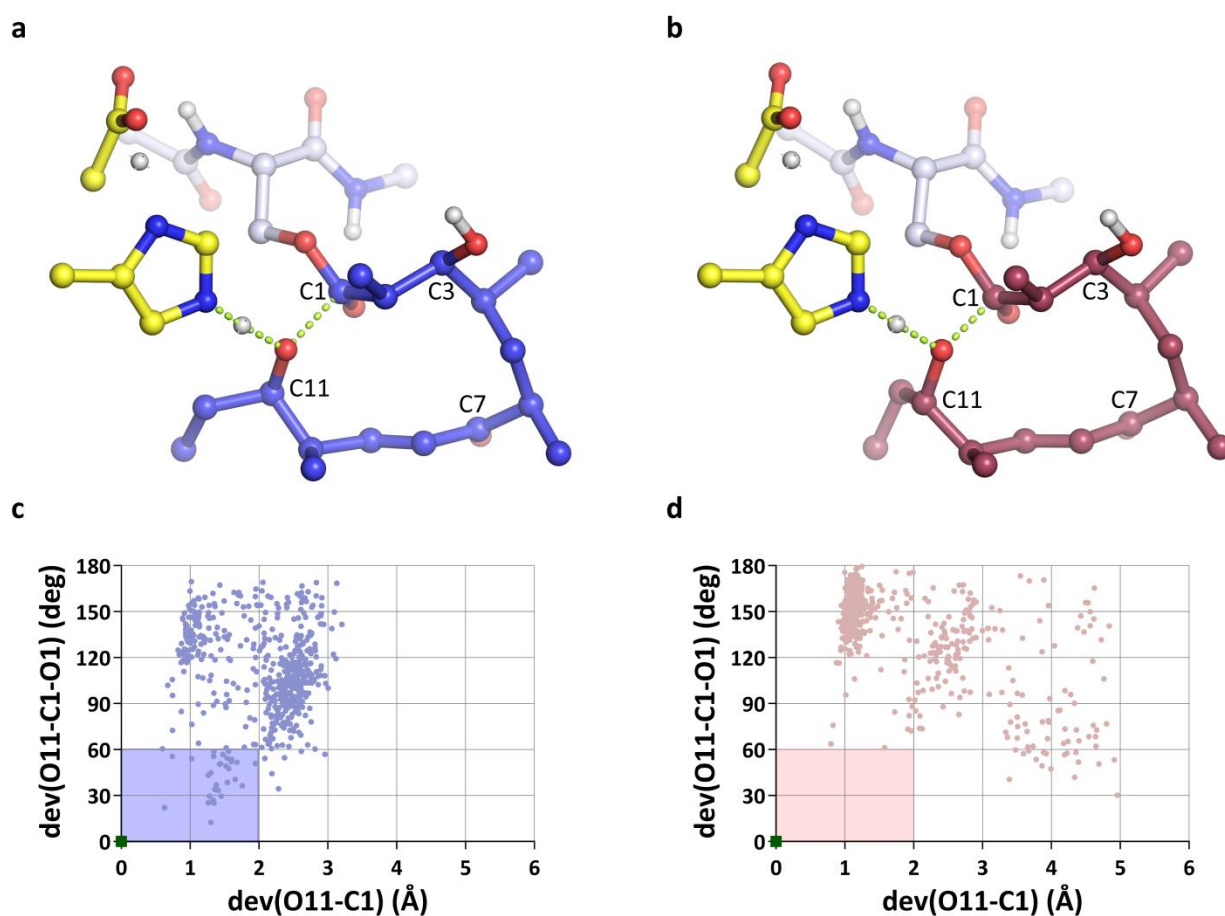
Figure 4.11 Procatalytic sampling of Pik TE during MD simulations. (a,b) Low energy QM optimized transition states for macrolactonization of Pik hexaketides (a) Pik TE-**1** and (b) Pik TE-**2**. Non-polar hydrogens have been removed for clarity. (c,d) Deviations of the key catalytic distances (x axis) and angles (y axis) in the MD simulations of (c) Pik TE-**1** and (d) Pik TE-**2** from their respective optimized transition structure (green square at the origin of coordinates). Each point represents a single frame from the 550 ns simulation while the shaded rectangles represent frames from the MD that are likely in a catalytically productive state.

Pik TE-**1** contained more frames with the hexaketide in a position favorable to macrolactonization (Figure 4.11c). In contrast, Pik TE-**2** presented a larger distribution of geometries with very few catalytically productive structures (Figure 4.11d). Taken together, the MD results on the covalent acyl-enzyme intermediates indicate that hydrophobic interactions accompanying a high level of substrate-TE shape complementarity are critical for maintaining each hexaketide in a catalytically productive conformation. These contacts are, to a lesser extent, maintained in simulations containing the C-11-epimerized hexaketide **2**, indicating that the Pik TE active site has sufficient flexibility to accommodate both epimers. The intrinsic conformational preferences of each hexaketide further influence catalysis. The acyl intermediates must reach conformations that are matched for deprotonation by His268 and subsequent nucleophilic attack in order to achieve macrolactonization. When these structural preferences are perturbed, as in the case with Pik TE-**2,** the energetic barrier to macrolactonization is increased above that of competing hydrolysis.

Chemical lactonization

As the MD simulations containing covalently bound hexaketides **1** and **2** revealed that both substrates are accommodated by the Pik TE active site, we next investigated if macrolactonization of the C-11-epimerized **2** suffered in bulk solution outside of enzymatic constraints. Starting with conditions employed for similar seco-acids[8,9] we screened contemporary lactonization methodologies capable of cyclizing **4** to **3**. Optimization of Yamaguchi conditions originally reported from the total synthesis of spinosyn analogs[10], enabled direct comparison of lactonization efficiency of **4** and **5**. Employing identical reaction conditions resulted in 73% and 6% conversions to generate **3** and **6**, respectively (Scheme 4.1). This result confirmed the intrinsically poor reactivity of **5** toward macrocyclization compared to **4**, corroborating the MD findings that the lack of macrolactone formation with the C-11-epimerized hexaketide is not due to steric clashes within the Pik TE active site.

Synthetic, biomimetic macrolactonizations of macrolide cores predate biochemical studies of TE domains, with the total synthesis of erythronolide B first

reported in 1978.[11] In the subsequent total synthesis of erythromycin A, Woodward and co-workers successfully generated the 9-dihydroerythronolide macrolactone only after screening 17 differentially protected seco-acids.[12] The need for specific protecting group arrays to promote macrolactonization was rationalized through conformational biasing of the seco-acid.[13] This conjecture was reevaluated using a C-9 reduced tetra-methylated seco-acid of 6-dEB lacking the extraneous hydroxyl groups at C-6 and C-12, and the authors found no biasing elements to be necessary.[8] Of note, 9-dihydroerythronolide is a dihydroxylated analog of 6-dEB and not the native substrate for the DEBS TE. Difficulties associated with macrolactonization of 9-dihydroerythronolide are potentially due to unproductive conformations imparted by two additional hydroxyl groups at C-6 and C-12, which, during biosynthesis, would be introduced after TE catalyzed macrolactonization by p450 hydroxylase tailoring enzymes. This discussion is meant to highlight the difficulties in biomimetic ring closure with analogs of polyketide natural products, where the conformation(s) necessary to enable macrolactonization may be perturbed from the natural, "competent conformation." The C-11-epimerized hexaketide **5** represents one such analog where the stability of the cyclized product **6** is decreased. Unsurprisingly, chemical lactonization suffers relative to the methyl ether of native hexaketide **4** (6% and 73% conversion, respectively), though Pik TE$_{S148C}$ is able to
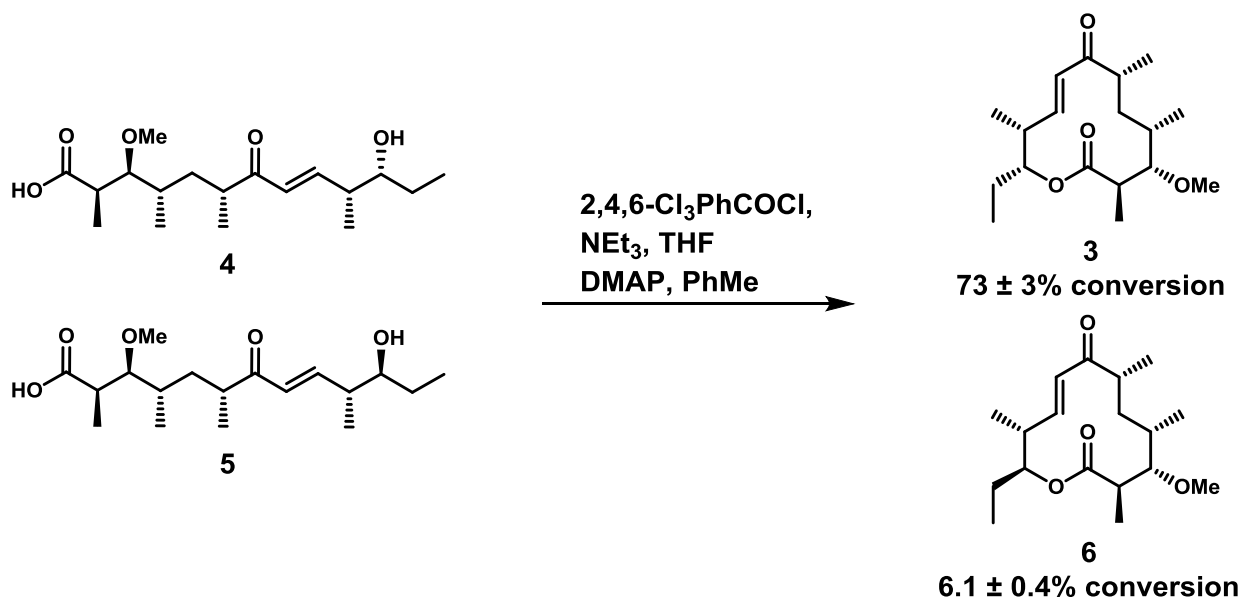
Scheme 4.1 Yamaguchi macrolactonization of methyl protected hexaketides. Conversion of **4** to **3** and **5** to **6** was monitored by HPLC with data represented as the mean ± standard deviation where $n = 3$.

91

overcome the structural penalty imposed by C-11 epimerization thus improving conversion (94% and 83%, respectively). Of note, Pik TE$_{S148C}$ catalysis operates using 1 mol% enzyme in water at room temperature, avoiding large volumes of dry organic solvents, heating, and expensive Yamaguchi or Shiina reagents.

QM calculations

Since MD simulations revealed the Pik TE$_{WT}$ active site to be competent in binding the unnatural substrate **5**, we hypothesized the catalytic advantages imparted by the S148C mutation were kinetic in nature as opposed to structural. To further our understanding of the energetics of these processes and how Pik TE$_{S148C}$ is able to overcome the barrier to macrolactone formation with **5**, we turned to QM modeling of the catalytic steps comprising macrolactonization after formation of the acyl-enzyme intermediate. To accomplish this, we generated abbreviated active site models (*theozymes*[14]) for Pik TE$_{WT}$ and Pik TE$_{S148C}$ containing the native C-3-unprotected hexaketide (Pik TE$_{WT}$-**7** and Pik TE$_{S148C}$-**7**) and the C-11-epi hexaketide (Pik TE$_{WT}$-**8** and Pik TE$_{S148C}$-**8**). Analysis of the resulting free energy landscapes revealed that macrolactonization of the linear hexaketide intermediates is an exergonic process in all four systems after product release (Figure 4.12). Cyclization of the native hexaketide was more thermodynamically favorable compared to C-11-epimerized hexaketide, particularly with the wild-type enzyme. Moreover, the macrolactonization mechanism with each TE was calculated to change from a stepwise addition-elimination with existence of a tetrahedral intermediate in Pik TE$_{WT}$ to a concerted acyl substitution[15] upon the S148C mutation. This change in mechanism was evidenced by the flat potential energy surface which precludes the formation of a tetrahedral intermediate.

Figure 4.13 shows the lowest energy transition structures and associated activation barriers ($\Delta G^{\ddagger}$) for the macrolactonization of the acyl intermediates of both the natural **7** and C-11-epimerized **8** substrates with the Pik TE$_{WT}$ and TE$_{S148C}$ protein models. The rate-limiting activation barriers for the TE$_{S148C}$ catalyzed reactions are significantly lower than those for the TE$_{WT}$ reactions, which correspond to the final Ser148-acyl cleavage. Thus, Pik TE$_{S148C}$ performs the macrolactonization of the native

and C-11-epimerized hexaketides with $\Delta G^{\ddagger}$ values of 14.8 and 16.0 kcal mol^{-1}, respectively, while Pik TE$_{WT}$ displays $\Delta G^{\ddagger}$ values of 18.2 and 22.4 kcal mol^{-1} for the same substrates (Figure 4.13). These values predict that the rate of reactions containing the native hexaketide after the formation of the acyl-enzyme intermediate is faster in Pik TE$_{S148C}$ compared to TE$_{WT}$.
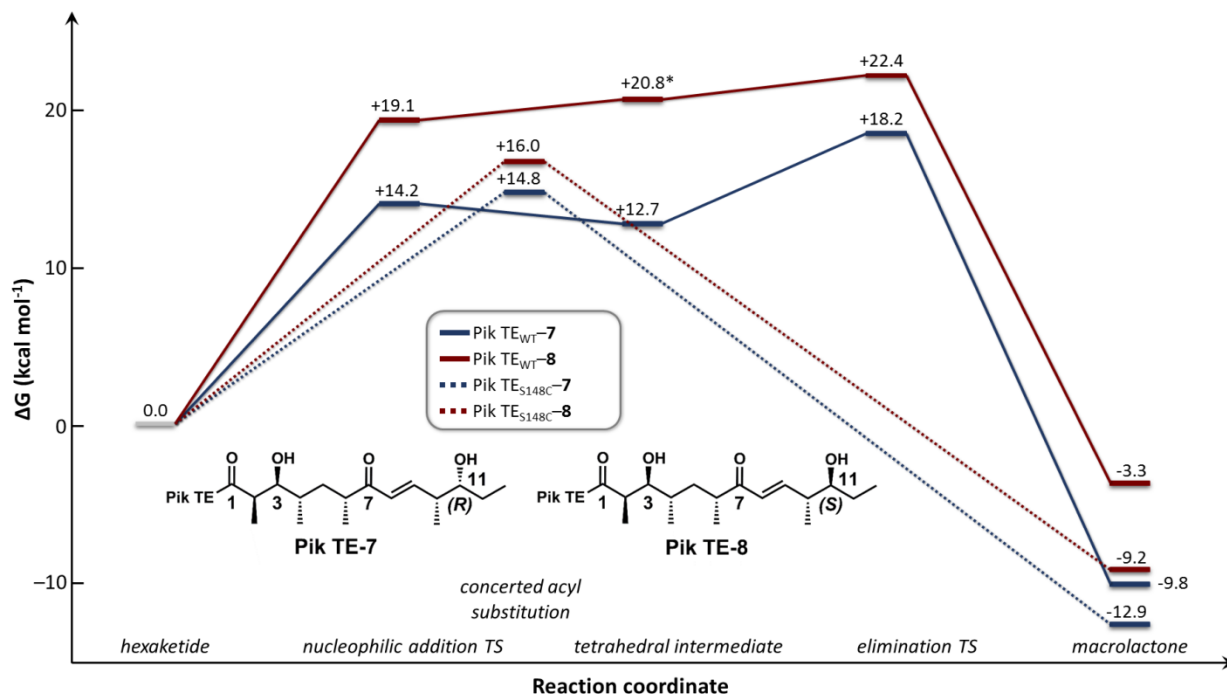
Figure 4.12 Reaction coordinate diagram representing the relative free energies for Pik TE catalyzed macrolactonization of the native and C-11-epimerized hexaketides **7** and **8**. Calculations were performed at the PCM/M06-2X/6-31+G(d,p) level using reduced models that define the enzymatic active site (*theozymes*[19]). Relative free energies are in kcal mol^{-1}. Only the lowest energy conformers are represented. *This intermediate is higher in energy than its preceding TS due to conformational differences between both stationary points.

The calculated increase of 4.2 kcal mol^{-1} in the activation barrier of the macrolactonization of Pik TE$_{WT}$-**7** *vs* Pik TE$_{WT}$-**8**, agrees well with the experimentally observed lack of epimerized-macrolactone formation. This higher macrolactonization energy barrier in the TE$_{WT}$ system increases the difficulty in proceeding from the acyl-enzyme intermediate, which is vulnerable to water hydrolysis.[16] While we did not observe significant water hydrolysis in reactions containing Pik TE$_{WT}$ and the native substrate, incubation with epimerized methyl protected **5** resulted exclusively in hydrolyzed product, the result of an inability to form the epimerized macrolactone **6**. In

contrast, the 6.4 kcal mol^{-1} decrease in activation barrier for Pik TE$_{S148C}$-**7** macrolactonization is consistent with our experimentally observed product formation.
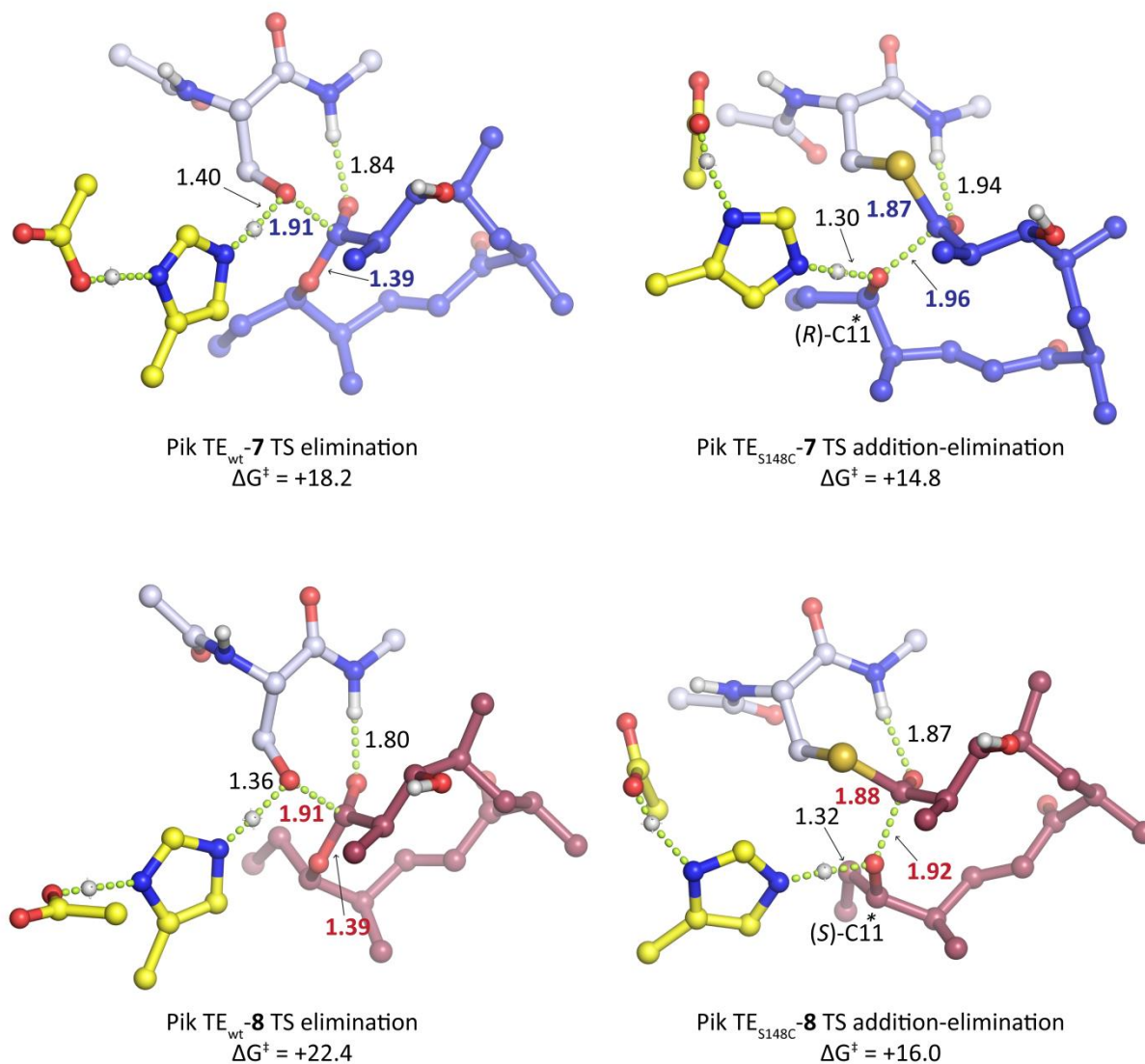
Figure 4.13 Lowest energy rate-limiting transition structures calculated with PCM/M06-2X/6-31+G(d,p) for the abbreviated active site models (*theozymes*[14]) of the Pik TE$_{WT}$ (left) and Pik TE$_{S148C}$ (right) catalyzed macrolactonization of native hexaketide **7** (top, in blue) and C-11-epi hexaketide **8** (bottom, in red). Activation free energies ($\Delta G^{\ddagger}$) calculated form the corresponding Pik TE hexaketides are given in kcal mol^{-1} and distances in angstroms. Relevant breaking/forming C–O and C–S bonds are shown in boldface. Non-polar hydrogens have been removed for clarity.

Computational investigation of the Pik TE domain catalysis

Our investigation revealed mutual recognition, TE-substrate shape complementarity, and intrinsic substrate structural preferences to be critical for the hexaketide to reach a conformation productive towards macrolactonization. MD simulations revealed hydrophobic interactions between the substrate and TE residues lining the active site as being vital for guiding the hexaketide to a catalytically competent *pro*-cyclic conformation. These findings are consistent with mutational analysis of DEBS TE where exchange of potential hydrogen-bonding residues did not substantively affect the specificity constant for hydrolysis of four unnatural thioester substrates. In this case, Wang and Boddy suggested that hydrophobic interactions between the active site and substrate are the main driving force of substrate specificity.[17] Additionally, our results are in agreement with the available crystal structures of Pik TE bound with phosphonate substrate mimics, which displayed a lack of specific TE-substrate polar contacts.[3,4]

Recently, Chen et al.[18] reported theoretical investigations of the thioesterase domain from the erythromycin biosynthetic pathway (DEBS TE) to describe the mechanistic parameters that determine the catalytic partitioning of substrates to either macrocyclic or linear hydrolysis products. MD simulations coupled with QM calculations were performed on systems of DEBS TE modeled with the native DEBS heptaketide and Pik hexaketide, which both led to macrolactonization, as well as two diastereomers of a reduced C-7 hydroxyl analog of Pik hexaketide that result exclusively in substrate hydrolysis. Analysis of the resulting MD simulations provided findings consistent with those in this report, particularly highlighting the importance of the formation of a substrate pre-reaction state through induced-fit mutual recognition between the enzyme and the substrate for macrolactonization to occur. Consistent with the present study, the authors found a hydrogen bond between the lactonizing hydroxyl group of the substrate and the catalytic histidine as well as hydrophobic interactions to be critical for formation of a catalytically competent pre-reaction state.

To understand the energetic consequences of the S148C mutation during the macrolactonization process, we performed DFT analysis of Pik TE_{WT} and TE_{S148C} modeled as acyl-enzyme intermediates with the native and C-11-epimerized

hexaketides. The results of our QM calculations revealed a significant kinetic advantage in the reactions catalyzed by Pik TE_{S148C}. The S148C mutation provides a mechanistic change during the macrolactonization step from a two-step transesterification in the TE_{WT} reaction (i.e. addition-elimination) to a lower energy single concerted step in the TE_{S148C} pathway.

Overall, our combined computational method for investigating Pik TE catalysis using MD simulations in concert with QM calculations provides a plausible explanation for improved substrate flexibility and catalytic efficiency of Pik TE_{S148C}. According to the results from our MD simulations, epimerization of the hexaketide C-11 stereocenter generates a substrate with a reduced propensity for acquiring a catalytically competent conformation within the TE active site. However, the ability of the hexaketides to reach a conformation viable for catalysis is not the only factor affecting macrocyclicization, especially if this step is not rate-limiting.[18] Hence, even if a substrate has a poor propensity to arrange in a productive conformation and/or lactonization is structurally hindered (as with the C-11-epimerized hexaketide) a significant acceleration through a key single mutation can overcome these structural limitations, boosting reactivity even with unnatural substrates.

Engineering TE domains

Encouraged by the remarkable results achieved with the Pik TE S148C mutation, we next set out to investigate whether similar benefits could be obtained with additional PKS TE domains. For initial testing we chose the homologous DEBS and Juv TE domains (Figure 2.3 and 2.4) and introduced the analogous serine to cysteine substitution in each. The cysteine mutant of each TE was generated from a pET28 plasmid encoding the corresponding wild type TE by substitution of the native serine codon using the QuikChange site-directed mutagenesis (Stratagene) protocol. Expression, purification, and analysis of DEBS TE_{WT} and TE_{S139C} demonstrated that the Ser to Cys mutation imparted increased substrate flexibility consistent with the results from Pik TE. HPLC analysis of a single time point reaction containing the methyl protected *N*-acetylcysteamine Pik hexaketide **1** revealed that although DEBS TE_{WT} was incompatible with the Pik substrate, DEBS TE_{S139C} was able to effectively generate

methyl protected 10-dml **3** (Figure 4.14). Analysis of Juv TE$_{WT}$ and TE$_{S142C}$ is currently ongoing and future efforts will build upon our initial results with DEBS TE$_{S139C}$ using an expanded the substrate scope to better understand the macrolactone analogs achievable through this TE engineering approach.
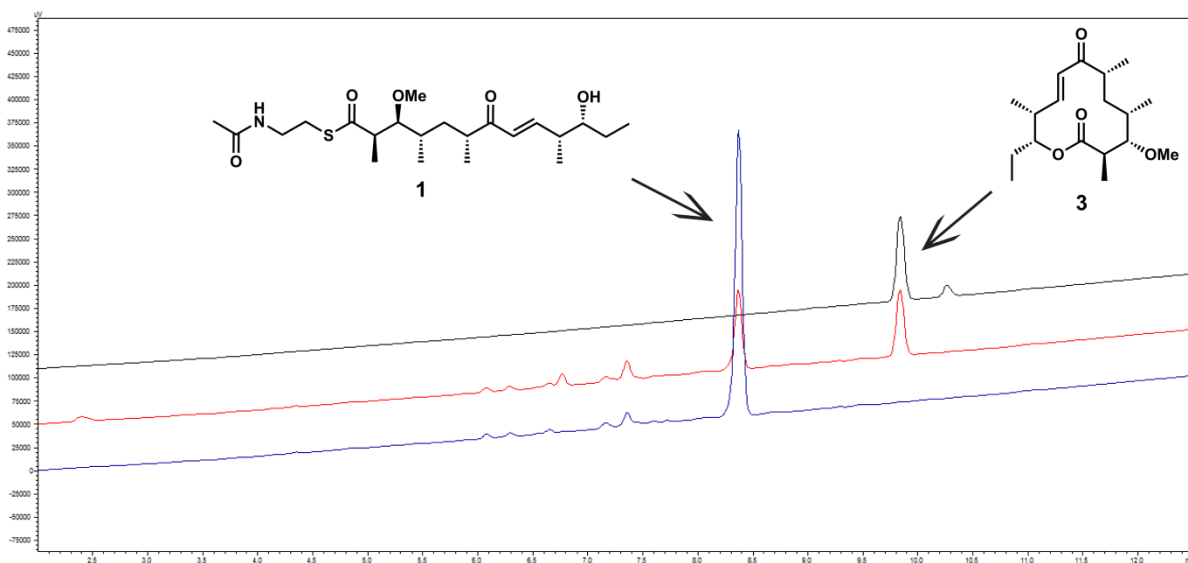
Figure 4.14 DEBS TE$_{S139C}$ possesses increased substrate flexibility. Comparison of reactions containing DEBS TE$_{WT}$ (blue trace) and DEBS TE$_{S139C}$ (red trace) confirms the gain-of-function processing of the methyl protected Pik hexaketide **1** to methyl 10-dml **3** (black trace) resulting from the Ser to Cys mutation. Enzymatic reaction conditions: 1 mM Pik hexaketide, 8 mM 2-vinylpyridine, 1 mol% DEBS TE (10 µM), 4 hours, RT. Conversion to **3** was monitored by HPLC.

Experimental procedures

Chemistry

Reactions were performed in dried glassware furnished through flame drying evacuated (<0.05 torr) flasks fitted with rubber septa containing PFTE coated magnetic stir bars backfilled with dry N_2 and run under a positive pressure of dry N_2 provided by mineral oil bubblers unless stated otherwise (open flask). Analytical thin-layer chromatography (TLC) was performed with EMD 60 F_{254} pre-coated glass plates (0.25 mm) and visualized using a combination of UV, *p*-anisaldehyde, $KMnO_4$ stains. Room temperature (RT) reactions were conducted at ~23 °C, reactions run cooler than room temperature were performed in a cold room (4 °C), an ice bath (0 °C), or a dry ice/acetone bath (-78 °C). Commercial solvent purification system MBraun-MB-SPS # 08-113 provided all dry solvents unless stated otherwise (technical grade). Flash column chromatography was performed using EMD 60 Gerduran® (particle size 0.04-0.063) silica gel.

TE Biochemistry

Purified H_2O from a Millipore Milli-Q system with Millipore Q-Gard 2/Quantum Ex Ultrapure organex cartridges was used for all cell culture, protein purification, and enzymatic reactions. *E. coli* seed culture was grown in 15 mL sterile tubes, and subsequently grown in Corning Fernbach flasks (2.8 L) with 3x deep baffles. LB broth (Miller) and glycerol were obtained from EMD. Isopropyl-β-D-thiogalactopyranoside (IPTG) and Kanamycin (Kan) sulfate were obtained from Gold Biotechnology. NaCl, $CaCl_2$ and imidazole were obtained from Fisher Scientific. Lysozyme was purchased from RPI. Benzonase was purchased from Sigma Aldrich. PD-10 columns were purchased from GE scientific and equilibrated with 5 column volumes of storage buffer before use. Ni-NTA agarose resin was purchased from Qiagen and pre-equilibrated with five column volumes of lysis buffer before use. A Symphony SB70P pH meter was calibrated according to the manufacturer's specifications and used to monitor the pH of all solutions during adjustment. Cells were lysed using a 550 Sonic Dismembrator purchased from Fisher Scientific. Optical

98

density (OD_{600}) was determined using an Eppendorf Biophotometer. All solutions were autoclaved or sterile filtered through a 0.2 μm filter.

Buffers:

lysis: HEPES (50 mM), NaCl (300 mM), imidazole (10 mM), glycerol (10% v/v), pH 8.0.
wash: HEPES (50 mM), NaCl (300 mM), imidazole (30 mM), glycerol (10% v/v), pH 8.0.
elution HEPES (50 mM), NaCl (300 mM), imidazole (300 mM), glycerol (10% v/v), pH 8.0.
storage: HEPES (50 mM), NaCl (150 mM), EDTA (1 mM), glycerol (20% v/v), pH 7.2.
gel filtration: Tris (20 mM), NaCl (150 mM), glycerol (10% v/v), pH7.5.

Protein Expression

The cloning, expression and purification of Pik TE_{WT}[19] and DEBS TE_{WT}[20] have been previously reported and served as the templates for generating the corresponding cysteine mutants. The creation of Pik TE_{S148C} and Juv TE_{WT} are reported in Chapter III of this Thesis. Pik $TE_{S148CH268X}$ mutants were generated from a pET28 plasmid encoding Pik TE_{S148C} by substitution of the wild type codon 268 CAC using the QuikChange site-directed mutagenesis (Stratagene) protocol. MBP-Pik $TE_{S148CH268R}$ was generated by insertion of Pik $TE_{S148CH268R}$ into a pET28 vector containing an N-terminal 6x His tag followed by the MBP protein and a TEV cleavage site. Identical protocols were used for the expression and purification of all Pik TE mutants.

A starter culture was generated by inoculating 5 mL of LB broth containing Kan (50 mg/L) with fresh transformants of E. coli (BAP1)[21] cells containing the corresponding plasmids for expression of the respective PKS proteins and grown overnight at 37 °C. Following the overnight growth, the entire starter culture was subsequently used to inoculate an expression culture of 1 L of TB containing Kan (50 mg/L) and grown at 37 °C to an OD_{600} of 0.3-0.4. The expression cultures were then cooled to 18 °C and growth was maintained until an OD_{600} of 0.7-0.8 was reached, at which point protein expression was induced via addition of IPTG (350 μM) and the cultures were incubated at 200 RPM at 18 °C for 20 hours.

Protein Purification

All proteins were subjected to the following first round of purification:

To retain maximum enzymatic activity, the following purification procedure was performed at 4 °C in less than 2 hours. Overexpression cultures were harvested by centrifugation (5,500 x g, 10 min, 4 °C) and cell pellets were suspended in 5 mL of <u>lysis</u> buffer per gram of cells via vortex. Cell lysis was accomplished by gentle agitation at 4 °C with 0.4 mg/ml lysozyme and 8 units/ml benzonase for 30 min followed by sonication on ice (1 x 10s with 50s rest periods). Cellular debris was pelleted by centrifugation (40,000 x g, 15 min, 4 °C), and the supernatant applied to 3 mL of Ni-NTA resin. After binding, the column was washed with 25 mL of <u>wash</u> buffer under gentle syringe pressure and the target protein was eluted with 15 mL of <u>elution</u> buffer. Protein containing fractions were assessed via their absorption at 280 nm, pooled, and buffer exchanged into <u>storage</u> buffer using a PD-10 column. Finally, protein containing fractions were determined via their absorption at 280 nm, pooled, flash frozen in liquid N_2, and stored at -80 °C.

Following purification by Ni-NTA chromatography, the relevant proteins were further processed for in vitro analysis:

TEV cleavage of the MBP-tagged Pik TE$_{S148CH268R}$ was accomplished by incubation of the fusion protein with the TEV protease (1:30 molar ratio of Pik TE to TEV protease) in <u>storage buffer</u> at 4 °C for 18 hours. After cleavage, the reaction mixture was passed over a small bed of Ni-NTA resin to remove the His-tagged MBP.

Gel filtration chromatography of Pik TE$_{S148CH268Q}$ was accomplished using HiPrep 26/60 Sephacryl S-300 HR column from GE Healthcare Life Sciences in <u>gel filtration</u> buffer set to a flow rate of 0.5 mL/min. Protein containing fractions were assessed via their absorption at 280 nm, pooled, concentrated with a Amicon Ultra-15 Centrifugal Filter (10 kDa cut-off), flash frozen in liquid N_2, and stored at -80 °C.

Pik TE acyl-enzyme formation

To determine the ability of Pik TE$_{S148CH268R}$ to be stabilized as the acyl-enzyme, we explored reaction conditions for optimum substrate loading. We anticipated alkaline

100

conditions favoring the transthioesterification reaction would provide the best results, ultimately a borate buffered solution at pH 10.2 provided the most consistent substrate incorporation. Final reaction conditions: 1:1 molar ratio of Pik TE to **1**, 1.7% DMSO, 150 mM borate pH 10.2 18 hours, RT. Analytical labelling reactions were assessed via LC-HRMS was performed on an Agilent LC system (1290 series) coupled to an Agilent QTOF mass spectrometer (6500 series) using a Phenomenex Aeris Widepore 3.6μ C4 50 x 2.10 mm column (serial 687169-3). Method: 0.4 mL/min, solvent A: H_2O 0.1% formic acid, solvent B: MeCN 0.1% formic acid, 0% B 0-2 min,0-100% B linear gradient 2-10 min, 100% 10-11 min, 0% B 11-12 min, 0-1 min were diverted to waste.

Chemical lactonizations

Optimum conditions for comparison of head-to-head lactonization efficiency of **4** and **5** were based on total synthesis of structurally related 12-membered macrocycles[10] and were performed on 12 μmol scale as follows. Seco-acid **4** or **5** (4 mg, 0.012 mmol, 1.0 equiv) and reaction vial were azeotroped with PhMe (3x). THF was added (1mL, 0.012 M) and cooled to 0 °C. Et_3N (10 μL, 0.048 mmol, 6.0 equiv) and 2,4,6-trichlorobenzoyl chloride (8 μL, 0.048 mmol, 4 equiv) were added sequentially and the reaction was allowed to stir at RT for 1.5 hour. In a separate flask, 4-dimethylaminopyridine (14.7 mg, 0.12 mmol, 10 equiv) was dissolved in PhMe (6 mL) and heated to 60 °C. After 1.5 hours at RT, activated seco-acid was diluted in additional PhMe (1 mL) and this solution was added dropwise to the heated DMAP solution over 2.5 hour via syringe pump. After addition was complete, reaction was allowed to continue at 60 °C for another 0.5 hour before being cooled to RT and quenched with aq. $NaHCO_3$ (sat). Organic and aqueous layers were separated, and organic layer was washed with brine and dried over $MgSO_4$ before concentration and HPLC analysis.

Computational analysis of TE catalysis

Pik TE structure preparation

The wild type apo Pik TE structure was obtained from the RCSB Protein Data Bank (PDB: 1MNA).[2] Structures for the wild type and mutant enzymes were prepared from WT Pik TE using PyMOL 1.3.[22] Chain B of PDB file 1MNA and the crystallographic

water were removed, and 5 missing protein residues (T109-T113) were built into chain A using PyMOL 1.3. Parameters for the apo TE were generated using the *antechamber* module of AMBER 14[23] with the widely tested Stony Brook modification of the Amber 99 force field (*ff99SB*)[24] force field. The initial substrate conformations were generated by modeling the linear hexaketide chain as neutral acyl serine or cysteine amino acids with N-terminal (-CO-CH$_3$) and C-terminal (-NH-CH$_3$) caps in an approximately cyclic orientation using the solved crystal structures of the derivatized macrolactone products as templates. The structure reported in reference 6 served as the template for Pik TE-**1** and Pik TE-**2**. The initial hexaketide confirmations were then optimized by performing a geometry optimization at the M06-2X functional[25] level using the Gaussian 09 package.[26] The charges and parameters for the hexaketide substrates were then determined using the *general AMBER force field* (GAFF)[27] with partial charges set to fit the electrostatic potential generated at the HF/6-31G(d) level by the RESP model.[28] The charges were calculated according to the Merz–Singh–Kollman scheme[29,30], using Gaussian 09. The N and C-terminal caps were then removed and the parameterized hexaketides were used to replace Ser148 of the TE to generate Pik TE modeled as acyl-enzyme intermediates. Parameter, topology, and coordinate files for the complex were compiled for molecular dynamics (MD) simulation using the *tleap*[23] module. Following parameterization, the Pik TE models with bound hexaketides were immersed in a pre-equilibrated truncated cuboid box with a 8 Å buffer of TIP3P[31] water molecules using *tleap* and the systems were neutralized by addition of explicit counter ions (Na$^+$ and Cl$^-$).

Molecular Dynamics simulations

MD simulations were performed using the AMBER 14 package.[23] A cutoff of 8 Å was used for all Lennard-Jones and electrostatic interactions. Water molecules were treated with the SHAKE algorithm such that the angle between the hydrogen atoms was kept fixed. Periodic boundary conditions simulated the effects of a larger system size. Long-range electrostatic effects were modelled using the particle-mesh-Ewald method.[32] Prior to performing the production MD simulations, the solvated systems were subjected to two steps each of minimization and equilibration to relax and correct any possibly

unrealistic arrangements: two minimizations were performed first over a total of 5000 cycles each, with the method being switched from the steepest descent algorithm to conjugate gradient after 2500 cycles. The first minimization relaxed the positions of the solvent molecules and ions by imposing positional restraints on the solute by a harmonic potential with a force constant of 500 kcal mol^{-1} Å^{-2}. Following the first minimization, a distance restraint of 2.3 Å between the hexaketide nucleophilic hydroxyl hydrogen and the Nε nitrogen of His268 was imposed and maintained throughout the second minimization, both equilibrations, and first production simulation, the restraint was then released for the second production simulation. This restraint maintains the hexaketide and the catalytic histidine residue –His268 acts as the general base–in an arrangement close to that calculated for the macrolactonization transition state, while allowing the hexaketide backbone to explore both reactive and non-reactive conformations. The restraint also served to minimize any potential biases resulting from the initial substrate positioning. After the distance restraint was imposed, a second stage minimization was performed on all the atoms in the simulation cell without cartesian atomic restraints. Post minimization, two equilibrations were performed. In the first equilibration step, a 0.1 ns MD simulation was performed with a constant pressure of 1 atm and a constraint of 10 kcal mol^{-1} Å^{-2} on the protein and substrate. The Andersen thermostat equilibration scheme was used to control and equalize the temperature by gradually increasing the system temperature from 0 to 300 K over the course of the first equilibration run. The system volume was then fixed and a second equilibration step was achieved with 2 ns MD simulation conducted without cartesian atomic restraints with a temperature maintained at 300 K using the Andersen thermostat.

Post equilibration, two additional MD simulations were carried out utilizing the above conditions: Initially, a 50 ns simulation was performed with the aforementioned distance restraint intact to ensure proper shape complementarity of the TE active site for the acyl intermediate forced in a pre-catalytic state. Next, the substrate-protein distance restraint was removed and a final production simulation was run for an additional 500 ns. This portion of the simulation represents the conformation of both the substrate and the protein environment in the covalently bound Michaelis-Menten complex prior to reaction.

An integration time step of 2 fs was utilized with structural snapshots being extracted every 50 steps during both production simulations. The resulting production trajectories were analyzed using the *cpptraj* program in Ambertools14 employing the default options to describe hydrogen-bonding, hydrophobic interactions, and clustering analysis.

QM Computational Details.

Full geometry optimizations were carried out with Gaussian 09[26] using the M06-2X hybrid functional[25] and 6-31+G(d,p) basis set. Solvation was considered implicitly through the PCM polarizable continuum model[33] in order to better reproduce the electrostatic environment inside the enzyme active site ($\varepsilon=4$). The possibility of different conformations was taken into account for all structures. All stationary points were characterized by a frequency analysis performed at the same level used in the geometry optimizations from which thermal corrections were obtained at 298.15 K. Local minima and first order saddle points were identified by the number of imaginary vibrational frequencies. The quasiharmonic approximation reported by Truhlar *et al.* was used to replace the harmonic oscillator approximation for the calculation of the vibrational contribution to enthalpy and entropy.[34] Scaled frequencies were not considered. Mass-weighted intrinsic reaction coordinate (IRC) calculations were carried out by using the Gonzalez and Schlegel scheme[35,36] in order to ensure that the TSs indeed connected the appropriate reactants and products. Gibbs free energies (ΔG) were used for the discussion on the relative stabilities of the considered structures. Cartesian coordinates, electronic energies, entropies, enthalpies, Gibbs free energies, and lowest frequencies of the calculated structures are available below.

Attempts to locate transition states for the initial covalent binding step (i.e. Ser/Cys148 acylation) from thioester models of Pik TE$_{WT}$ and Pik TE$_{S148C}$ were unsuccessful, due to the very flat topology found for the corresponding potential energy surfaces. Hence, very low activation barriers are expected for the initial substrate binding, and macrocyclization was assumed to be the rate-limiting reaction.

References

(1)   Lu, H.; Tsai, S.-C.; Khosla, C.; Cane, D. E. *Biochemistry* **2002**, *41*, 12590.

(2)   Tsai, S.-C.; Lu, H.; Cane, D. E.; Khosla, C.; Stroud, R. M. *Biochemistry* **2002**, *41*, 12598.

(3)   Giraldes, J. W.; Akey, D. L.; Kittendorf, J. D.; Sherman, D. H.; Smith, J. L.; Fecik, R. A. *Nat. Chem. Biol.* **2006**, *2*, 531.

(4)   Akey, D. L.; Kittendorf, J. D.; Giraldes, J. W.; Fecik, R. A.; Sherman, D. H.; Smith, J. L. *Nat. Chem. Biol.* **2006**, *2*, 537.

(5)   Witkowski, A.; Witkowska, H. E.; Smith, S. *J. Biol. Chem.* **1994**, *269*, 379.

(6)   Hansen, D. A.; Koch, A. A.; Sherman, D. H. *J. Am. Chem. Soc.* **2015**, *137*, 3735.

(7)   Laskowski, R. A.; Swindells, M. B. *J. Chem. Inf. Model.* **2011**, *51*, 2778.

(8)   Stang, E. M.; White, M. C. *Angew. Chem. Int. Ed.* **2011**, *50*, 2094.

(9)   Junji, I.; Kuniko, H.; Hiroko, S.; Tsutomu, K.; Masaru, Y. *Bull. Chem. Soc. Jpn.* **1979**, *52*, 1989.

(10)   Tietze, L. F.; Schützenmeister, N.; Grube, A.; Scheffer, T.; Baag, M. M.; Granitzka, M.; Stalke, D. *Eur. J. Org. Chem.* **2012**, *2012*, 5748.

(11)   Corey, E. J.; Kim, S.; Yoo, S.-E.; Nicolaou, K. C.; Melvin, L. S.; Brunelle, D. J.; Falck, J. R.; Trybulski, E. J.; Lett, R.; Sheldrake, P. W. *J. Am. Chem. Soc.* **1978**, *100*, 4620.

(12)   Woodward, R. B.; Au-Yeung, B. W.; Balaram, P.; Browne, L. J.; Ward, D. E.; Au-Yeung, B. W.; Balaram, P.; Browne, L. J.; Card, P. J.; Chen, C. H. *J. Am. Chem. Soc.* **1981**, *103*, 3213.

(13)   Stork, G.; Rychnovsky, S. D. *J. Am. Chem. Soc.* **1987**, *109*, 1565.

(14)   Tantillo, D. J.; Jiangang, C.; Houk, K. N. *Curr. Opin. Chem. Biol.* **1998**, *2*, 743.

(15)   Yang, W.; Drueckhammer, D. G. *J. Am. Chem. Soc.* **2001**, *123*, 11004.

(16)   Horsman, M. E.; Hari, T. P. A.; Boddy, C. N. *Nat. Prod. Rep.* **2016**, *33*, 183.

(17)   Wang, M.; Boddy, C. N. *Biochemistry* **2008**, *47*, 11793.

(18)   Chen, X.-P.; Shi, T.; Wang, X.-L.; Wang, J.; Chen, Q.; Bai, L.; Zhao, Y.-L. *ACS Catal.* **2016**, *6*, 4369.

(19)   Lu, H.; Tsai, S.-C.; Khosla, C.; Cane, D. E. *Biochemistry* **2002**, *41*, 12590.

(20)   Gokhale, R. S.; Hunziker, D.; Cane, D. E.; Khosla, C. *Chem. Biol.* **1999**, *6*, 117.

(21)   Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C. *Science* **2001**, *291*, 1790.

(22)   The PyMOL Molecular Graphics System, Version 1.3 Schrödinger, LLC.

(23)   Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; T.S. Lee, S. L.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; W. Smith, J. S.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. AMBER 14 (UCSF, 2014).

(24)   Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.

(25)   Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.

(26)   Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.;

Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09 (Gaussian, 2009).

(27)   Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157.

(28)   Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.

(29)   Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431.

(30)   Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.

(31)   Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(32)   Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

(33)   Scalmani, G.; Frisch, M. J. *J. Chem. Phys.* **2010**, *132*, 114110.

(34)   Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B.* **2011**, *115*, 14556.

(35)   Gonzalez, C.; Schlegel, H. B. *J. Chem. Phys.* **1989**, *90*, 2154.

(36)   Gonzalez, C.; Schlegel, H. B. *J. Phys. Chem.* **1990**, *94*, 5523.

Chapter V

Discussion

Summary and insights

Due to advances in molecular techniques and the subsequent elucidation of the biosynthetic logic of polyketide diversification[1,2], the past two decades witnessed an intense interest in the biosynthetic production of macrocylic polyketide analogs.[3-7] During this time, many research groups reported the production of large libraries of macrolactone analogs achieved through PKS engineering strategies.[4-6] These PKS engineering efforts mainly employed canonical combinatorial biosynthesis techniques and targeted their modifications to the individual domain level to instill new functionality into the macrolactone ring. Three prominent examples were reported by Kosan Biosciences, highlighting the ability to generate libraries of polyketide analogs through modular PKS engineering.[4-6] In these studies, AT and KR domain substitutions were engineered into modules 2, 5, and 6 of the 6-deoxyerythronolide B (DEBS) biosynthetic pathway and the engineered PKSs were expressed in heterologous *Streptomyces* or *E. coli* hosts to generate a suite of 6-dEB analogs. However, as evidenced in this seminal work, these engineering attempts employing combinatorial biosynthesis techniques often resulted in large decreases in product titers.[8] This unintended consequence was exemplified by the <0.1 mg L^{-1} product yields from the double and triple mutant pathways versus the 20 mg L^{-1} from that of the wild type system.[4] Thus, although these historical examples successfully generated inspiringly diverse arrays of macrolactone analogs, the associated decreased product yields renders these strategies as impractical for providing quantities sufficient for the drug discovery pipeline.[8]

Convoluting the mechanistic interpretation of these decreased yields is the in vivo context in which these experiments were performed. Often times, the engineered

PKS system was expressed in a heterologous host and product confirmation was determined by LC-MS analysis of the resulting growth media[1,2]. Associated with this experimental design are two significant limitations: First, the application of LC-MS analysis for product confirmation, in contrast to isolated and structurally characterized products, provides ambiguity for the regio- and stereoselectivity of the reactions from the engineered catalytic domains as the reaction outcomes would be isobaric and inseparable from their mass spectrum. Secondly, the biochemical basis for the low yields or failed production of metabolites from engineered PKSs remains elusive. Typically, genetic modifications have been performed in early pathway PKS modules[1-6], and it has remained unclear whether attenuated product formation stems from the engineered module or the downstream modules (or both) that must accept and process the resulting unnatural intermediates. Therefore, even if the engineered module or domain successfully performs its non-native functionality, the nascent polyketide chain must then be productively acted upon by the downstream core, β-keto processing, and thioesterase domains in order to produce the anticipated natural product analog. If this downstream processing fails, then the assembly line tethered elongation intermediate may be offloaded hydrolytically thus producing an unstable shunt product difficult to detect by LC-MS or isolate from the complex fermentation medium. Notably, we[7-9] and others[10,11] have previously shown that PKS TE domains function as flexible hydrolases when substrate macrolactonization cannot be achieved, resulting in substrate flux to intermediates not readily detected by LC-MS. Consequently, interpretation of in vivo PKS engineering studies using a binary system of analysis based on LC-MS detection of the fully mature product fails to provide insight into the catalytic point of failure if the anticipated product is not detected. Motivated by this gap of knowledge, the studies reported in this thesis seek to provide further insight into the structural and mechanistic underpinnings that govern catalysis in engineered PKSs in efforts to improve their potential for the biosynthetic production of polyketide analogs.

Towards this goal, we have performed experiments focused on providing a more general framework for rational engineering of efficient PKSs through targeted investigations designed to probe the catalytic details of individual domains. Interrogating the substrate flexibility of the late stage PKS module PikAIII-TE[12-14] with

unnatural pentaketide analogs designed to simulate "combinatorial biosynthesis" provided key insights into the catalytic bottlenecks encountered within engineered PKSs. Using a targeted in vitro approach, we reported a strict substrate stereoselectivity inherent to PikAIII-TE, resulting in the failed processing of an epimeric pentaketide analog containing an (S)-configured nucleophilic hydroxyl group. As the only change from the natural Pik pentaketide is epimerization of the C-9 hydroxyl group that serves as the nucleophile during macrolactonization, we reasoned, a priori, that the substrate would detrimentally affect the TE domain responsible for ring formation. TE stereoselectivity has been observed previously from in vitro reactions of DEBS TE with a series of unnatural DEBS heptaketide substrate mimics. The DEBS TE displayed strict stereoselectivity for the natural (R)-configuration of the substrate nucleophilic alcohol and exclusively hydrolyzed the unnatural (S)-stereoisomer[15]. NMR and MS characterization of the reaction products from our panel of stereoisomeric Pik pentaketides showed that the intermediates had been extended; therefore the KS domain was at least partially competent in processing these analogs. Additionally, as the Pik TE is able to cyclize β-keto analogs of the natural hexaketide to yield 3-keto-10-dml[13,16] we expected to observe 3-keto macrolactones if the KR was compromised[17-19], but KS and TE domains remained functional.

We hypothesized that the isolated shunt products resulted from substrate stalling and premature off-loading from the PKS via hydrolysis and subsequent decarboxylation. Analysis of these extended but linear off-loaded intermediates suggests that the TE is acting as the dominant catalytic bottleneck to the formation of macrocylic products. The inability to attain a total mass balance for the stereoisomer substrate panel and the low recovery of shunt products from each reaction is attributed to the additional facile degradative pathway possible through the TE domain. PKS TE domains serve as flexible hydrolases when cyclization is impaired.[7,10] Thus, we reasoned that a portion of the pentaketides were extended and off-loaded as the unstable β-hydroxy hexaketide, which subsequently degrades through intramolecular hemiketalization and dehydration pathways[14,20].

Initial in vitro biochemical characterization of the DEBS TE[10,11] provided further evidence for the relatively high substrate tolerance of PKS TEs for acylation and

hydrolysis, as terminally (omega) hydroxylated fatty acids and substrates resembling simplified DEBS heptaketides were all hydrolyzed by DEBS TE. However, the ability of PKS TEs to cyclize substrates other than their native linear intermediates has proved to be much more limited. In fact, TE mediated macrolactonization has only been observed in a select few studies.[7,15,20-27] In addition to their native substrates, Pik TE has been shown to catalyze macrolactonization of C-3 methyl and NBOM protected derivatives[20], though not C-7 reduced analogs.[7,22] DEBS TE has been shown to catalyze macrolactonization of unnatural mimics of the DEBS heptaketide.[15,25] However, there are no reports of either the Pik or DEBS TEs catalyzing macrolactonization of a substrate containing a nucleophilic hydroxyl group with an unnatural, epimerized (*S*)-configuration. When probed for the ability to form an epimeric heptaketide mimic of 6-dEB , the DEBS TE displayed a high level of stereoselectivity for the natural (*R*)-configuration, and exclusively hydrolyzed the unnatural (*S*)-stereoisomer[15], adding to the observations in this study for strict stereospecificity in TE catalyzed macrolactonization. Indeed, synthesis of the corresponding C-11-epimerized hexaketide and subsequent probing of the Pik TE domain provides direct evidence for the strict stereoselectivity of Pik TE. This stereoselectivity confirms the suspected TE bottleneck for macrolactone formation and explains the loss of macrolactone formation in PikAIII-TE reactions containing the diasteromeric pentaketides.

Having demonstrated the critical nature of the TE domain in the processing of unnatural substrates, we hypothesized that the TE domain is acting as a dominant catalytic bottleneck in PKS engineering strategies that utilize the rearrangement of upstream catalytic domains as these result in intermediates that may fail to be processed by the downstream incompatible TE domain. To explore this concept, we generated a series of hybrid type I PKS modules with TE domains from three related biosynthetic pathways and tested each for catalytic activity with full length substrates to simulate the final steps in engineered PKS catalysis. Remarkably, we were able to achieve robust catalysis with non-native module/substrate pairs when the hybrid module was paired with a competent TE domain. These results expand upon our previous findings and demonstrate the key role of the TE domain in the processing of unnatural intermediates. Of note, the results reported in this study are in agreement with similar

work performed in fungal iterative polyketide synthases (iPKSs) which identified the TE domain as a key gatekeeper in the production of fungal polyketides.[28-31] Furthermore, our results with the Pik hexaketide indicate that the identity of the TE domain may also alter the sequence of catalytic events that occur in engineered PKS modules. This is evidenced by the product distribution in reactions pairing the Pik hexaketide with Ery Mod6-Pik TE and Ery Mod6-DEBS TE to generate predominantly narbonolide or 3-hydro-narbonolide, respectively.

The mechanism for this unexpected consequence from exchanging of TE domains is unclear at this point. It is tempting to speculate that the identity of the TE domain can interfere with the canonical vectorial processing of the PKS assembly line through premature loading of the elongation intermediate that most closely resembles its native substrate. Although the cryo-EM studies of PikAIII indicated that the sequence of interactions taking place between the PKS catalytic domains and the ACP-tethered intermediate occurs in an orchestrated and linear fashion[32,33], I note that these studies were performed using wildtype PikAIII with its native substrate. Perhaps PKS assembly line processing follows a more diffuse path in which the loaded ACP samples all available catalytic domains in search of the interaction which provides the highest degree of affinity. Notably, this paradigm indicates the molecular identity of the ACP-tethered elongation intermediate could impact the binding affinity, and thus subsequently, the length of time the elongation intermediate spends within each catalytic domain. In the context of native assembly line processing, this dynamic sampling would be finely tuned for proceeding in the catalytic sequence that leads to the fully mature natural product through variations in the individual domain and elongation intermediate binding affinities. However, in the processing of unnatural substrates, the native sequence could be perturbed if the order of binding affinities is rearranged such that the ACP-tethered elongation intermediate prematurely resides in an alternative catalytic domain. Thus since narbonolide is the native product of Pik TE and 3-hydro-narbolide more closely resembles the native product of DEBS TE (6-dEB contains a C-3 hydroxyl group), reactions containing Ery Mod6-DEBS TE follow the anticipated catalytic sequence to include KR mediated reduction while the reactions with Ery Mod6-

112

Pik TE result in the premature transfer of the ACP-tethered elongation intermediate to the Pik TE domain prior to processing by the KR domain.

Stabilization of the Pik hexaketide allowed for direct probing of the excised Pik TE domain and the following confirmation of the TE catalytic bottleneck due to strict substrate stereoselectivity. Remarkably, initial engineering attempts identified a single active site mutation within the Pik TE that imparts both increased substrate flexibility and catalytic efficiency, resulting in a macrolactonization catalyst capable of forming diastereomeric macrolactones. Investigating the potential of TE engineering within the context of full module processing through direct comparison of PikAIII-TE$_{WT}$ and PikAIII-TE$_{S148C}$ with a panel of stereoisomer Pik pentaketides demonstrated the success of this approach as a 0.2 millimole scale reaction resulted in the production of two new diastereomeric macrolactone products, 11-epi-10-dml and 3-keto-11-epi-10-dml. The 1:1 isolation of the reduced and unreduced macrolactones from reactions with PikAIII-TE$_{S148C}$ indicates that the TE was indeed responsible for the failed catalysis using WT PikAIII-TE, as the KR domain retained a sufficient level of activity towards the unnatural intermediate. This demonstration of TE engineering offers a potential means for increasing the substrate flexibility of biosynthetic pathways for the production of natural product analogs.

Unfortunately, one hurdle in the general application of Pik TE$_{S148C}$ in full module catalysis is the increased levels of substrate hydrolysis observed with PikAIII-TE$_{S148C}$. We had originally found that thio- or oxoester choice determined the catalytic route of the hexaketide in the presence of PikAIV in vitro, with thiophenol and *N*-acetylcysteamine thioesters having a 10:1 preference for either full-module processing or direct cyclization, respectively. Thus, we reasoned that the low yields from the PikAIII-TE$_{S148C}$ reactions with the C-9-epimerized pentaketide are likely due to direct loading of the S148C TE domain and subsequent hydrolysis. Indeed, control reactions confirmed (i) the S148C mutation overrides the KS preference previously enjoyed by thiophenol thioesters (ii) enzymatic reactions performed without extender unit or reductive cofactors showed a much higher hydrolysis with PikAIII-TE$_{S148C}$ relative to wild type.

Future directions

The results from the studies contained within this thesis serve to highlight the critical nature of performing targeted, in vitro investigations into the catalytic functioning of engineered PKS modules. Moving forward, it is of paramount importance to expand our growing understanding of the substrate flexibility of each particular PKS catalytic domain in order to provide a clearer blueprint for future PKS engineering techniques. Particularly productive will be the continued application of synthetic chemistry to provide full length native substrates, and analogs thereof, to probe the flexibility of late stage PKS modules. This strategy was successfully employed to reveal the unintuitive role of the TE domain as a dominant catalytic bottleneck in the processing of unnatural substrates. Identification of these gatekeeper domains should enable future efforts to focus on the engineering of each bottleneck in order to increase the efficiency of engineered PKS pathways for the production of polyketide analogs.

Although we achieved promising results by employing engineered TE domains within PKS modules, it is clear that the other core catalytic PKS domains play a role in substrate gating as well. This was particularly noted in the hybrid TE PKS module reactions pairing Juv Mod6 TE hybrids with the Pik hexaketide, as no choice in TE resulted in significant product formation. Since all three TE domains were able to produce at least one of the anticipated products when paired with DEBS Mod6, we interpret the failed Juv Mod6 TE hybrid reactions to indicate an additional catalytic bottleneck located elsewhere in the module. Thus, future efforts will need to focus on identifying these additional catalytic roadblocks through the use of full length substrates to probe individual modules and excised domains directly. Application of this strategy should provide additional insights into the structural and mechanistic parameters that govern PKS function in a manner similar to the identification of the TE domain bottleneck from the studies reported in this thesis.

Our results lay the foundation for the application of TE engineering to increase the substrate flexibility and catalytic efficiency of both excised TE domains and engineered PKS modules. The remarkable results achieved with Pik TE_{S148C} suggest that Ser to Cys mutations of the catalytic triad within type I PKS TE domains may

provide macrolactonization catalysts capable of producing difficult to form macrocycles. This would have significant impacts on the field of natural products synthesis as current chemical macrolactonization methodology is often ineffective in the key ring-closing step of linear natural product intermediates.[34] Although many efficient macrocyclization methods have been developed, the lactonization of seco-acids is still a frequently utilized approach to obtaining macrocycles[35]. Unfortunately, these classical approaches to the ring-closing step in the synthesis of natural products often require significant experimental optimization and suffer from complex synthetic sequences, high dilution conditions, etc.[34] In contrast, Pik TE_{S148C} catalysis operates using 1 mol % enzyme in water at room temperature, avoiding large volumes of dry organic solvents, heating, and expensive Yamaguchi or Shiina reagents. Furthermore, cryptic substrate conformational and stereochemical requirements are critical for successful chemical macrolactonization. These experimental limitations are highlighted from the pioneering total synthesis of erythromycin A in which Woodward and co-workers successfully generated the 9-dihydroerythronolide macrolactone only after screening 17 differentially protected seco-acids.[36] The need for specific protecting group arrays to promote macrolactonization was rationalized through conformational biasing of the seco-acid[37], and thus suggests that any substrate analogs would require their own extensive optimization. Consequently, the implementation of engineered TE domains as macrolactonization catalysts offers great potential value for the chemoenzymatic synthesis of natural products. Future efforts will focus on performing analogous Ser to Cys mutations within a diverse array of TE domains responsible for forming both macrolactones as well as macrolactams and testing them for the ability to more efficiently generate valuable macrocyles versus the current chemical macrocyclization routes.

A second promising implementation of TE domain engineering is in the incorporation of optimized TE domains within engineered PKS pathways as a means for increasing the efficiency of the biosynthetic production of polyketide analogs. It is tempting to speculate on the potential benefits an optimized TE domain would impart on the 6-dEB analog combinatorial biosynthesis studies mentioned previously. If the low product titers reported from this work were in fact due to incompatibilities of the native

115

DEBS TE with the unnatural elongation intermediates, then it is possible that a contemporaneous engineering of the TE domain may alleviate these detrimental effects. This concept was explored through the generation and analysis of hybrid TE PKS modules in Chapter II, and the current findings could be advanced by performing these TE domain swaps in vivo for comparison with the historical experimental design. Introduction of TE domains with cysteine nucleophiles into engineered PKS pathways also offers potential for increasing the product diversity achievable with these engineered systems. However, if the tendency towards premature loading of elongation intermediates observed from the in vitro reactions with PikAIII-TE$_{S148C}$ is recapitulated in vivo, then additional TE engineering with the goal of tuning the reactivity for the acylation step will be required.

Recently, Schaffer et al. reported a TE domain with a native cysteine nucleophile critical to the cyclization of a strained β-lactone ring in the obafluorin (Obi) nonribosomal peptide pathway.[38] Of note, an analogous cysteine to serine mutation in the Obi TE resulted solely in the hydrolysis product, further implicating the catalytic advantage of a cysteine TE in the cyclization of strained ring systems. In this study the authors evaluated the prevalence of Cys containing TE domains by analyzing the complete annotated Pfam database and the top 20,000 hits from a standard protein BLAST search. Survey of the TE protein family (PF00975) in Pfam indicated that 6.8% of members possessed a cysteine nucleophile while the remainder contained the more prevalent serine. Of the BLAST results, 13,944 possessed a canonical serine. Examples of modular type I polyketide TE domains containing cysteine active site nucleophiles are notably rare.[39] Furthermore, at the present time we are unaware of any naturally occurring macrolactone forming TE domains that contain a cysteine nucleophile. We note that the TE domain from the type I PKS fluvirucin B$_1$ naturally contains a cysteine nucleophile, however, Flu TE catalyzes the formation of a *macrolactam* since the acyclic precursor contains an amine intramolecular nucleophile arising from the β-alanine derived starter unit.[40]

We hypothesize that the lack of naturally occurring type I PKS TE domains possessing a cysteine nucleophile can be attributed at least in part to the potential negative effects the more nucleophilic cysteine would impart on the vectorial channeling

116

of intermediates within polyketide synthase assembly lines. As shown in Chapter III, Pik TE$_{S148C}$ is less sensitive to the choice in thio- and oxoesters for substrate loading, indicating a higher rate of acylation from the more reactive cysteine nucleophile. Furthermore, PikAIII-TE$_{S148C}$ when incubated with the native Pik pentaketide produces both 10-dml and the unreduced 3-keto-10-dml suggesting that the cysteine TE domain is able to outcompete the KR domain in the native catalytic sequence. The precise reactivity of the active site residues within each individual catalytic domain has been finely tuned in order for the assembly line to function in a cohesive manner to produce the fully mature natural product. Therefore, incorporation of a cysteine active site TE domain into a native type I PKS pathway may interfere with the productive catalytic sequence through inappropriate and/or premature loading of biosynthetic intermediates. Given the significant results obtained with Pik TE$_{S148C}$, we are currently in the process of curating all annotated TE domains involved in macrolactone formation in search of those that may contain a cysteine nucleophile. Further studies will focus on delineating the biosynthetic parameters that select for cysteine or serine active site residues.

Moving forward, investigations employing x-ray crystallography and cryo-electron microscopy focused on full modules and excised domains in conjunction with full length substrates will begin to unveil the structural parameters that govern substrate processing within the context of engineered PKS systems. Supplementing these experimental techniques with computational modelling will provide further insights into the conformational and thermodynamic effects resulting from specific engineered catalytic domains. These advanced structural studies should shed light on the parameters that determine the sequence of catalytic events in the processing of unnatural substrates and thus advance our understanding of the vectorial channeling of intermediates within polyketide synthase assembly lines in both engineered and native biosynthetic pathways. This foundational knowledge will in turn better equip the PKS community for designing de novo pathways capable of efficiently generating high value polyketide analogs. Due to the critical nature of the TE domain discovered by the experiments performed in this thesis, current efforts in the Sherman lab are focused on capturing the Pik TE in complex with its native substrate, as well as with full length analogs, as a co-crystal structure for x-ray crystallography. We anticipate that analysis

117

of the bound acyl-enzyme complex will provide structural insights useful for the precise engineering of non-native functionality into PKS TE domains. Optimization of crystallization conditions suitable for the acyl-enzyme intermediates generated in Chapter IV is currently ongoing.

While additional biochemical studies of excised domains and full-modules with full length analogs of native substrates are necessary to support and expand our understanding of PKS function, it is becoming apparent that efficient production of specific, designer macrolide analogs will require significant pathway engineering. However, production of a desired natural product analog should be obtainable through the following workflow: (i) a targeted domain is engineered to perform an unnatural function (ii) downstream modules are biochemically characterized in vitro with the resulting unnatural polyketide in order to identify bottlenecks (iii) catalytically inefficient domain(s) are engineered with the goal of restoring effective processing to generate unnatural products. Indeed, as noted in a recent review by Weissman[41], it is perhaps more realistic to envision the future application of PKS engineering as a synthetic biology tool for producing specific, high-value natural product derivatives through targeted reprogramming of modular type I polyketide pathways than generation of natural product libraries.

References

(1)   McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Betlach, M.; Ashley, G. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 1846.

(2)   Xue, Q.; Ashley, G.; Hutchinson, C. R.; Santi, D. V. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 11740.

(3)   Jacobsen, J. R.; Keatinge-Clay, A. T.; Cane, D. E.; Khosla, C. *Bioorg. Med. Chem.* **1998**, *6*, 1171.

(4)   Leaf, T.; Cadapan, L.; Carreras, C.; Regentin, R.; Ou, S.; Woo, E.; Ashley, G.; Licari, P. *Biotechnol. Prog.* **2000**, *16*, 553.

(5)   Kinoshita, K.; G Williard, P.; Khosla, C.; Cane, D. E. *J. Am. Chem. Soc.* **2001**, *123*, 2495.

(6)   Harvey, C. J.; Puglisi, J. D.; Pande, V. S.; Cane, D. E.; Khosla, C. *J. Am. Chem. Soc.* **2012**, *134*, 12259.

(7)   Aldrich, C. C.; Venkatraman, L.; Sherman, D. H.; Fecik, R. A. *J. Am. Chem. Soc.* **2005**, *127*, 8910.

(8)   Hansen, D. A.; Koch, A. A.; Sherman, D. H. *J. Am. Chem. Soc.* **2017**, *139*, 13450.

(9)   Koch, A. A.; Hansen, D. A.; Shende, V. V.; Furan, L. R.; Houk, K. N.; Jiménez-Osés, G.; Sherman, D. H. *J. Am. Chem. Soc.* **2017**, *139*, 13456.

(10)   Gokhale, R. S.; Hunziker, D.; Cane, D. E.; Khosla, C. *Chem. Biol.* **1999**, *6*, 117.

(11)   Aggarwal, R.; Caffrey, P.; Leadlay, P. F.; Smith, C. J.; Staunton, J. *J. Chem. Soc., Chem. Commun.* **1995**, 1519.

(12)   Yin, Y.; Lu, H.; Khosla, C.; Cane, D. E. *J. Am. Chem. Soc.* **2003**, *125*, 5671.

(13)   Aldrich, C. C.; Beck, B. J.; Fecik, R. A.; Sherman, D. H. *J. Am. Chem. Soc.* **2005**, *127*, 8441.

(14)   Mortison, J. D.; Kittendorf, J. D.; Sherman, D. H. *J. Am. Chem. Soc.* **2009**, *131*, 15784.

(15)   Pinto, A.; Wang, M.; Horsman, M.; Boddy, C. N. *Org. Lett.* **2012**, *14*, 2278.

(16)   Chemler, J. A.; Tripathi, A.; Hansen, D. A.; O'Neil-Johnson, M.; Williams, R. B.; Starks, C.; Park, S. R.; Sherman, D. H. *J. Am. Chem. Soc.* **2015**, *137*, 10603.

(17)   Fiers, W. D.; Dodge, G. J.; Li, Y.; Smith, J. L.; Fecik, R. A.; Aldrich, C. C. *Chem. Sci.* **2015**, *6*, 5027.

(18)   Li, Y.; Fiers, W. D.; Bernard, S. M.; Smith, J. L.; Aldrich, C. C.; Fecik, R. A. *ACS Chem. Biol.* **2014**, *9*, 2914.

(19)   Li, Y.; Dodge, G. J.; Fiers, W. D.; Fecik, R. A.; Smith, J. L.; Aldrich, C. C. *J. Am. Chem. Soc.* **2015**, *137*, 7003.

(20)   Hansen, D. A.; Koch, A. A.; Sherman, D. H. *J. Am. Chem. Soc.* **2015**, *137*, 3735.

(21)   Boddy, C. N.; Schneider, T. L.; Hotta, K.; Walsh, C. T.; Khosla, C. *J. Am. Chem. Soc.* **2003**, *125*, 3428.

(22)   He, W.; Wu, J.; Khosla, C.; Cane, D. E. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 391.

(23)   Wang, M.; Boddy, C. N. *Biochemistry* **2008**, *47*, 11793.

(24)   Heberlig, G. W.; Wirz, M.; Wang, M.; Boddy, C. N. *Org. Lett.* **2014**, *16*, 5858.

(25)   Hari, T. P.; Labana, P.; Boileau, M.; Boddy, C. N. *Chembiochem* **2014**, *15*, 2656.

(26)   Kudo, F.; Kitayama, T.; Kakinuma, K.; Eguchi, T. *Tetrahedron Lett.* **2006**, *47*, 1529.

(27) Wang, M.; Zhou, H.; Wirz, M.; Tang, Y.; Boddy, C. N. *Biochemistry* **2009**, *48*, 6288.

(28) Xu, Y.; Zhou, T.; Zhang, S.; Xuan, L.-J.; Zhan, J.; Molnár, I. *J. Am. Chem. Soc.* **2013**, *135*, 10783.

(29) Xu, Y.; Zhou, T.; Zhang, S.; Espinosa-Artiles, P.; Wang, L.; Zhang, W.; Lin, M.; Gunatilaka, A. A. L.; Zhan, J.; Molnár, I. *Proceedings of the National Academy of Sciences* **2014**, *111*, 12354.

(30) Newman, A. G.; Vagstad, A. L.; Belecki, K.; Scheerer, J. R.; Townsend, C. A. *Chem. Commun. (Camb.)* **2012**, *48*, 11772.

(31) Newman, A. G.; Vagstad, A. L.; Storm, P. A.; Townsend, C. A. *J. Am. Chem. Soc.* **2014**, *136*, 7348.

(32) Dutta, S.; Whicher, J. R.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Congdon, G. R.; Narayan, A. R.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. *Nature* **2014**, *510*, 512.

(33) Whicher, J. R.; Dutta, S.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Dosey, A. M.; Narayan, A. R.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. *Nature* **2014**, *510*, 560.

(34) Marti-Centelles, V.; Pandey, M. D.; Burguete, M. I.; Luis, S. V. *Chem. Rev.* **2015**, *115*, 8736.

(35) Parenty, A.; Moreau, X.; Niel, G.; Campagne, J. M. *Chem. Rev.* **2013**, *113*, PR1.

(36) Woodward, R. B.; Au-Yeung, B. W.; Balaram, P.; Browne, L. J.; Ward, D. E.; Au-Yeung, B. W.; Balaram, P.; Browne, L. J.; Card, P. J.; Chen, C. H. *J. Am. Chem. Soc.* **1981**, *103*, 3213.

(37) Stork, G.; Rychnovsky, S. D. *J. Am. Chem. Soc.* **1987**, *109*, 1565.

(38) Schaffer, J. E.; Reck, M. R.; Prasad, N. K.; Wencewicz, T. A. *Nat. Chem. Biol.* **2017**, *13*, 737.

(39)   Horsman, M. E.; Hari, T. P. A.; Boddy, C. N. *Nat. Prod. Rep.* **2016**, *33*, 183.

(40)   Lin, T.-Y.; Borketey, L. S.; Prasad, G.; Waters, S. A.; Schnarr, N. A. *ACS Synthetic Biology* **2013**, *2*, 635.

(41)   Weissman, K. J. *Nat. Prod. Rep.* **2016**, *33*, 203.