The Dissertation Committee for Constance Beryl Bailey Certifies that this is the approved version of the following dissertation:

Enzymatic Features and Biocatalytic Applications of Modular Polyketide Synthase Domains

Committee:

Adrian Keatinge-Clay, Supervisor

Eric Anslyn

Hung-Wen Liu

David Hoffman

Sylvie Garneau-Tsodikova

Enzymatic Features and Biocatalytic Applications of Modular Polyketide Synthase Domains

by

Constance Beryl Bailey, B.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2015

Dedication

To my wonderful husband, John.

Acknowledgements

I would like to thank everyone who has been along with me on this journey that has been my scientific career to date. My advisor Prof. Adrian Keatinge-Clay, I appreciate for his incredible passion and enthusiasm for scientific endeavors, but for his singular focus on polyketides and PKSs. I came to his laboratory because I had a specific interest in this area, and I have been fortunate to have some great projects working toward deciphering these systems. I would like to thank those mentors that came before Adrian, specifically Prof. Arthur Glasfeld, Prof. Patrick McDougal, and Prof. Alan Shusterman at Reed College who inspired my love for chemistry and encouraged me to pursue a career in scientific research. I am also indebted to Prof. Christian Hertweck at the Hans Knoell Institute, because without my stint in his laboratory I would have not ended up here, at the University of Texas, in Adrian's laboratory pursuing polyketide research for my graduate career.

At UT several other faculty have guided my intellectual development. I thank Prof. Dionicio Siegel for countless chats in his office when he offered me both experimental advice on my syntheses as well as much appreciated career advice. Prof. Eric Anslyn undoubtedly impacted my chemical intuition through his physical organic course, as well as chats and collaborations over the years. Prof. Christopher Bielawski I thank for the opportunity to pursue the most fruitful collaboration of my career thus far. Prof. David Hoffman, thank you for letting me walk into your office with random ideas for experiments and bounce off ideas. My lab mates have undoubtedly been a great source of support, humor, frustration, and absurdity. Darren Gay, your weirdness and intelligence, and conspiracies kept the Keatinge-Clay lab from ever being boring.

Amanda Hughes, thanks for the laughs, the therapeutic vents, the ridiculous nights drinking, and the enthusiastic scientific daydreams. Shawn Piasecki, thank you for paving the way on my KR project and for cutting my hair and clothing me with your designer hand-me-downs throughout graduate school. Chris Fage, I am glad that Shawn, Amanda, Eta and I were able to give you a glimpse of what you missed out on, as you don't have any sisters, and I hope you are doing well in Germany. Eta Isiorho, you were a great desk mate--I enjoyed very much discussing anything from enzymatic mechanisms and natural products biosynthesis, to TLC reality shows. Andy Harper, we had a very nice paper together. Jessica Mienke, thank you for the moments of entertainment, YouTube clips, and general camaraderie. Jianting Zheng, I thank for all the patience and molecular biology advice and for putting up with all of our silliness. Jia Zeng, I am jealous of your zen and ability to sleep. Drew Wagner, I admire your passion and endless curiosity. Cole Stevens, I will miss our discussions of the literature, academic gossip, and I thank you for your advice and guidance. Alex Zheng, please keep surprising us with unexpected events like caviar in group meeting. Thanks to Marji Pasman for being a wonderful undergraduate coauthor. Thanks also to my other undergraduate mentees, Andrew Tran, and Seong-Meen Yoon, and my graduate rotation student mentee, Nick Brenes, for their help on various projects. Also thank you to the honorary Keatinge-Clay lab member, Jessica Momb, for always letting her office be a place for advice and commiseration. Also thanks to all the great friends outside of my group that I have made in graduate school including Jordan Dinser, Chris Chen, Christina Davis, Garrett Blake, Kristin Suhr, Dominika Lastovickova, Aaron Teator, and Alex Todd. Thank you to Penny Kile and Betsy Hamblen for all their help navigating bureaucratic aspects of UT.

I would like to also thank old friends that I made prior to graduate school, who have commiserated with my ups and downs including Ian McCulloch, Becky Phillips,

Laurel Brehm, Leigh Walton, Gaurav Vekataraman, Emma Friedman, Luke Kaiser, and Hanna Levitin. And finally, my family: my parents Michael Bailey and Debra Favre, and my brother Michael. Most importantly, I thank my husband John Brantley, who has been a deep source of emotional support and intellectual development. The collaboration on the GFP project sparked our relationship, but our collaboration is a life-long one.

Enzymatic Features and Biocatalytic Applications of Modular Polyketide Synthases

Constance Beryl Bailey, Ph.D.

The University of Texas at Austin, 2015

Supervisor: Adrian T. Keatinge-Clay

Polyketides are a class of secondary metabolites that are notable for their chemical diversity and therapeutic relevance. They are biosynthesized by polyketide synthases (PKSs) megasynthase enzymes in an assembly-line fashion. Though the molecular architectures of polyketides are complex, their biological precursors are chemically simple. Thus, understanding this powerful biosynthetic machinery is of interest for synthetic biology and biocatalytic applications. This dissertation presents three projects that decipher underlying mechanistic features and explore biocatalytic applications of PKSs.

In modular PKSs, one module corresponds to one round of keto-elongation followed by modification through the action of β-carbon processing domains. The first project employs a system wherein a single module is used *in vitro* to generate small, chiral PKS products (triketide lactones). Although triketide lactones are a common output for PKS enzymology assays, usually they are only observed in trace quantities. In this study, we performed a number of strategies to scale up the production of triketide lactones to facilitate their use as chiral building blocks for chemical synthesis. In this process, we also gained new insights regarding the interacting kinetics and selectivities of the domains in an *in vitro* environment.

The second project focused on the ketoreductase (KR) domain, which sets the majority of the stereogenic centers within a polyketide, and thus has obvious potential for biocatalytic applications. This project employs a structure-activity relationship (SAR)-type approach to dissecting stereocontrol. The SAR results, in concert with crystallographic data inspired two rational mutations that were sufficient to reverse the stereoselectivity of a representative KR. Thus, we were able to employ a rational approach to engineering stereocontrol.

The final project also focuses on the KR domain, however from a subclass of PKSs termed *trans*-acyltranferase (AT) PKSs. In contrast to the canonical *cis*-AT PKSs, the *trans*-AT PKSs have more varied modular organizations and architectures. One of these peculiar organizations one termed a "split" bimodule, wherein domains within a module are present on different polypeptides. Structural characterization of a KR from a split bimodule revealed features that may correspond to interpeptide interactions that afford communication between the two polypeptides of the split bimodule. Additionally, bioinformatic analysis of KRs from split bimodules reveals a number of diagnostic sequence motifs.

Table of Contents

List of Tables

List of Figures

List of Schemes

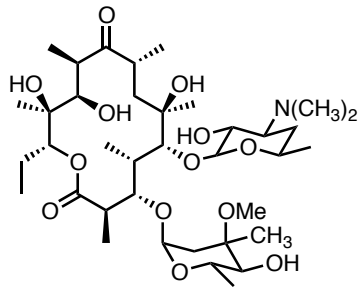
Chapter 1: Introduction

Overview

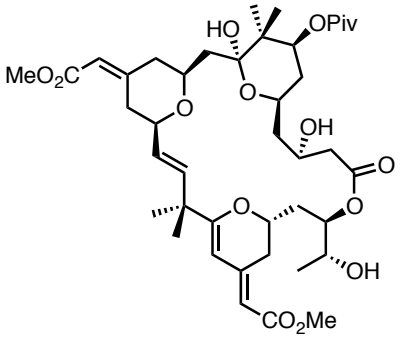
This dissertation describes three projects related to the overarching theme of dissecting mechanistic features of modular polyketide synthases (PKSs) and, in turn, harnessing PKSs for biocatalytic applications. The first project examines a miniaturized model system that generates small, chiral PKS fragments, termed triketide lactones (**Chapter 2**).[1] The second project focuses on the mechanistic details of an individual domain, the ketoreductase domain, which sets two stereogenic centers during an enzymatic reduction (**Chapter 3**). The third project also focuses on the ketoreductase domain, however in the context of an unusual (and largely uncharacterized) class of modular polyketide synthases, the *trans*-ATs (**Chapter 4**). An additional project unrelated to the main doctoral work is described in **Chapter 5**.[2]

Polyketide Secondary Metabolites and Polyketide Synthases

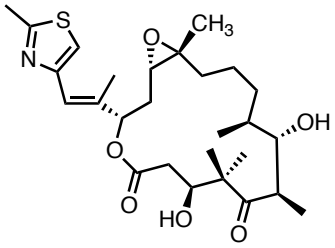
Polyketides comprise a diverse class of natural products with varied and intriguing carbon skeletons that are known for their pharmaceutical relevance.[3,4] Although (in most cases) their exact physiological roles are unknown, it is believed that they frequently function as pigments, virulence factors, signaling molecules, or as a form of defense against competing organisms.[3] Regardless of their evolutionary function, polyketides have been fruitful for drug discovery efforts: the "hit rate" for polyketide natural products has been estimated to be ~0.3%, which is far superior to the typical "hit rate" of <0.001% for a standard synthetic compound library.[4] As medicinal entities, they are primarily used as antibiotics, antifungals, immunosuppressants, antitumoral agents, and cholesterol-lowering agents (**Figure 1.1**).



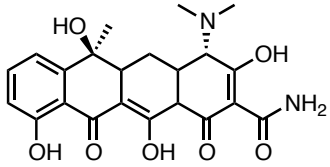
Erythromycin A, Antibiotic



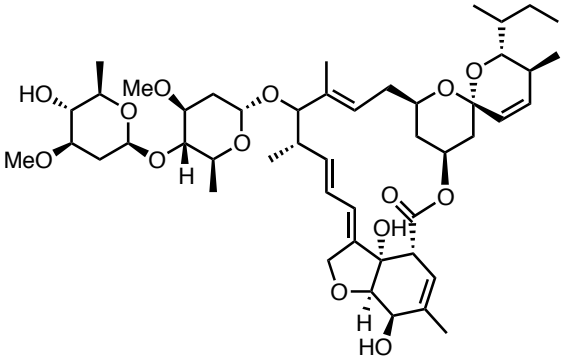
Bryostatin 16, Anticancer/anti-Alzheimers



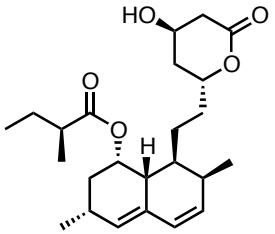
Epothilone B, Anticancer



Tetracycline, Antibiotic



Avermectin B_{1a}, Antiparasitic



Lovastatin, Cholesterol-Lowering

Figure 1.1 Selected pharmaceutically relevant polyketides.

Due to the stereochemical complexity of many polyketides, members of this class of natural products have historically been vibrant targets for developing and showcasing asymmetric methodologies.[5–9] Indeed in 1956, R.B. Woodward commented that the

polyketide, erythromycin A "looks at the present time quite hopelessly complex, particularly in view of its plethora of asymmetric centers."[10] In a pioneering effort, his group reported the first enantioselective total synthesis of the macrolide in 1981.[11] Nonetheless, Woodward's comment illustrates the inspirational role of polyketide scaffolds in the development of asymmetric C-C bond formation strategies. In stark contrast to their molecular complexity, the biological precursors of polyketides are remarkably simple: they are typically acetate or propionate-derived building blocks. Polyketides are biosynthesized by polyketide synthases (PKSs), which are multidomain megasynthases that assemble polyketides in an assembly-line fashion. PKSs select an extender unit (typically malonyl- or methylmalonyl-CoA) with an acyltranferase (AT) domain, which then undergoes a decarboxylative Claisen-like condensation through the action of a ketosynthase (KS) domain. Subsequently, the keto group may be processed by optional β-carbon processing domains which include: 1) the ketoreductase (KR), which reduces the keto group to a hydroxy moiety, 2) the dehydratase (DH), which can dehydrate the hydroxy moiety to form an olefin, and 3) the enoylreductase (ER), which results in a fully saturated carbon backbone. Finally, the polyketide is either cyclized or hydrolyzed through the action of a thioesterase (TE) domain. Throughout this process, the elongating chain is anchored on acyl carrier proteins (ACPs), which are tethered to the intermediate through an 18Å phosphopantetheinyl prosthetic group.[i] Following the action of the PKS, the carbon scaffold is usually decorated by post-PKS tailoring enzymes (*i.e.* oxidations, glycoylations, etc.)[3,12–14] A useful attribute of modular polyketide synthases is the principle of "colinearity," or the one-to-one correspondence

[i] Additionally, some polyketides contain a methyltransferase (MT) domain, which can introduce α branching when malonyl-CoA is used as the extender unit.

between gene sequence and metabolite structure.[ii] This relationship affords *a priori* predictions of the metabolite structures *via* their enzyme architectures (i.e. the domains present within the module;[3,12] consider 6-dEB biosynthesis,[13] **Figure 1.2**). Thus the colinearity principle has obvious implications for the rational reprogramming of their biosynthetic pathways for synthetic biology purposes.

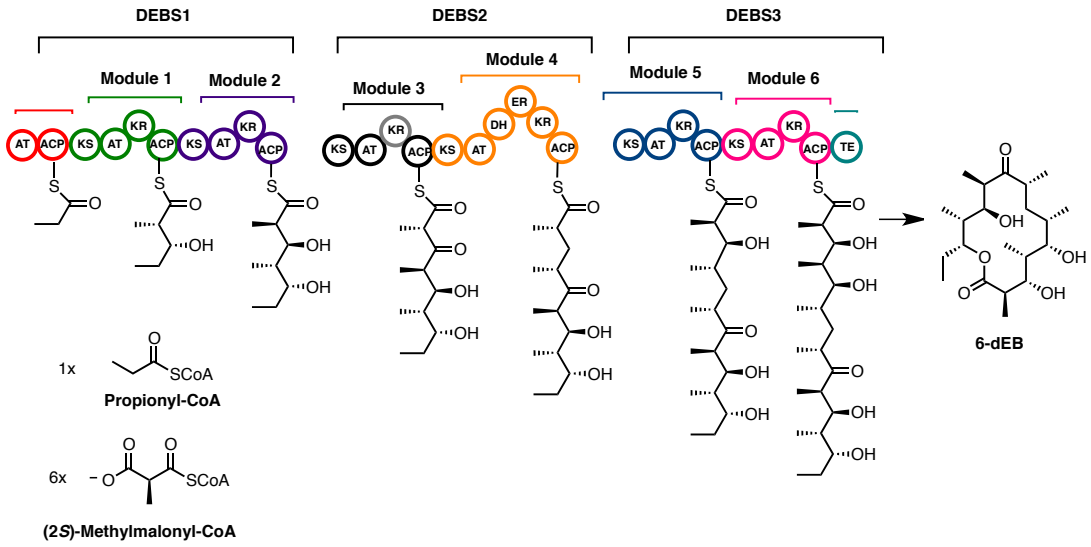


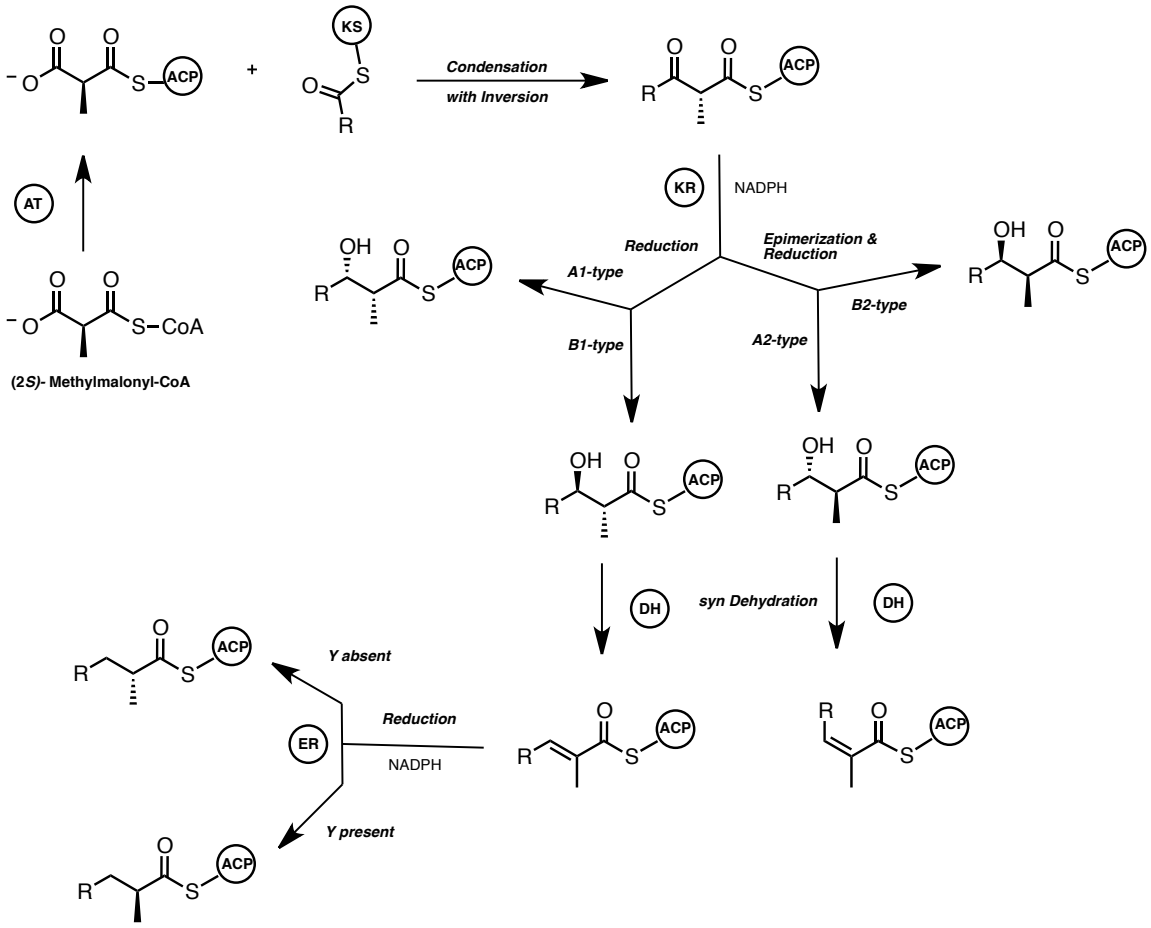
Figure 1.2 Biosynthetic pathway for 6-deoxyerythronolide B (6-dEB), the agylcone precursor to the polyketide antibiotic, erythromycin.

MECHANISTIC FEATURES OF PKS CATALYTIC DOMAINS

The colinearity principle can be taken one step further: in addition to predicting the connectivity and degree of reductive processing that the metabolite undergoes, examination of the protein sequence often provides "fingerprints," or diagnostic residues that correlate to enzymatic selectivities (including stereochemical outcomes; **Scheme 1.1**).[14] The first stereospecific step in PKS biosynthesis is often through the selection of

[ii] This review discusses modular type I polyketides. Other subtypes of polyketides exist, including type II polyketides (where the domains are present on separate polypeptides[109,110]), and type III polyketides (which lack ACPs and instead have CoA-linked intermediates).[111,112]

the extender unit by the AT domain. The AT domain possesses an ~240 residue catalytic subdomain with an α/β hydrolase fold. Its active site consists of a serine, activated by a histidine and a backbone carbonyl.[14] The mechanism for extender unit selection is through formation of an acyl-enzyme intermediate that will either be transthioesterified with the next ACP or hydrolyzed. Specificity arises from the respective rates of formation of the acyl-enzyme intermediate and the rate of transthioesterification to the ACP; the remaining "incorrect" extender units are then edited by AT domains' hydrolytic activity.[15] Whether the AT is selective for malonyl or methylmalonyl-CoA can be determined by a fingerprint found ~90-100 residues C-terminal of the active site serine (a YASH motif for methylmalonyl-CoA and a HAFH motif for malonyl-CoA;[16] other motifs have been correlated to more unusual extender units, however they are less robust[17,18]). For AT domains that accept methylmalonyl-CoA, there is an exclusive preference for the 2*S* isomer.[19] The current model for substrate selection of the 2*S* isomer is one in which the C_2 methyl forms favorable hydrophobic interactions with the tyrosine while being sterically accommodated by the serine. The 2*R* isomer is precluded by steric repulsion with the histidine of the YASH motif.[20]



Scheme 1.1 Summary of the introduction of stereochemistry by PKSs.

Following extender unit selection, the KS domain (when α-substitution is present) catalyzes the next stereospecific step. KSs have a thiolase fold,[21] and they catalyze the decarboxylative condensation of ACP-linked extender units. The KS active site consists of two histidines and a cysteine (to which the acyl group is covalently attached). Once the decarboxylation occurs, the nascent enolate performs a nucleophilic attack on the cysteine-linked thioester electrophile.[14] During the condensation process, there is an inversion of configuration of the α-methyl group, resulting in an α-methyl-β-keto

intermediate with a D-configured methyl group.[22] Often KSs serve as "gatekeeping" domains and prevent the condensation of incorrect polyketide intermediates.[23,24]

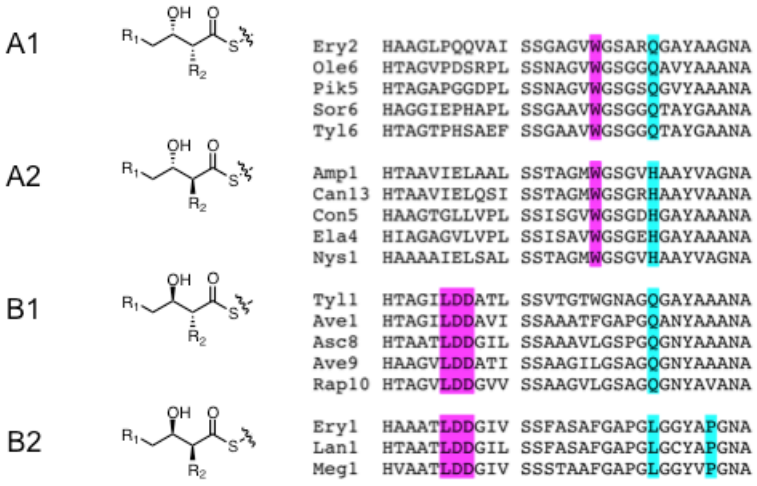


Figure 1.3 Conserved fingerprints correlating to KR stereochemical outcome. Residues correlating with ß–hydroxy stereochemistry are shown in magenta (the LDD motif correlates to D-ß-hydroxyls and the conserved W correlates to L-ß-hydroxyls). Residues are correlated with α-stereochemistry are shown in cyan.[25]

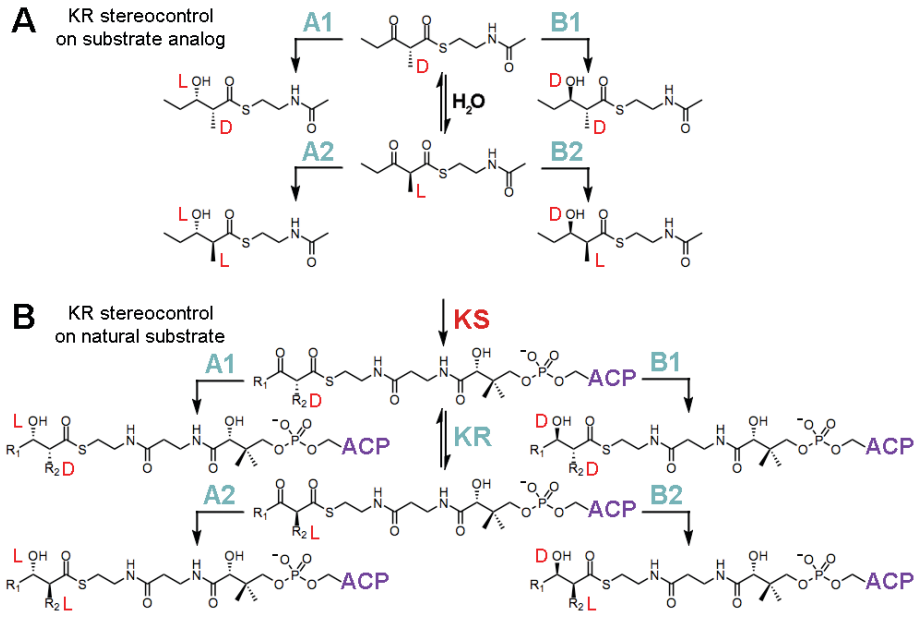
The domain that confers the most stereochemical complexity in PKS systems is the KR domain, as they can set two sp^3-stereogenic centers in a single reduction. Accordingly, KRs are some of the most thoroughly studied of the domains embedded within PKSs, and they possess some of the most robust predictive sequence fingerprints. KRs are a member of the short chain dehydrogenase/reductase (SDR) superfamily, and they are comprised of two domains, which each have a Rossman fold: an N-terminal structural subdomain and a C-terminal catalytic one. The catalytic subdomain harbors a conserved dinucleotide-binding site with consensus sequence TGGTGxLG.[26] KRs are classified as B-type KRs (which generate D-ß-hydroxyls), A-type KRs (which generate L-

ß-hydroxyls),[27,28] or C-type KRs (which are reductase-incompetent).[29][iii] A conserved tryptophan correlates to the A-type stereochemical outcome, whereas a leucine-aspartate-aspartate (LDD motif) correlates to the B-type stereochemical outcome.[27,28] Furthermore, KRs that generate products with a D-α-methyl are denoted with "1" and KRs that generate products with an L-α-methyl are denoted as "2"; sequence motifs are also correlated with these stereochemical outcomes (**Figure 1.3**).[29] Mechanistically, KRs undergo reduction by NADPH with a catalytic tyrosine (which acts as a general acid) and a serine that aids in the stabilization of negative character on the ß oxygen during the reduction process.[26]

KRs are also the best characterized of the PKS domains structurally, with ten PKS KR crystal structures to date (eight published).[25,30–35] Structural characterization has revealed that all KRs bind the NADPH cofactor in the same orientation (transferring the 4-*pro*-*S*-hydride[36]). Thus, stereoselectivity arises from the orientation of the ß-keto intermediate in the KR active site. To date, there are no crystal structures of KRs where a substrate mimic is bound; thus all evidence for interactions between various active-site residues and substrates come from modeling[37] and mutagenesis studies.[32,38–40] However, several features have been hypothesized to influence stereocontrol. Although the exact role of the conserved typtophan in A-type KRs is unknown, in B-type KRs it is believed that the last aspartate in the LDD motif hydrogen bonds with the first amide after the thioester in the phosphopantetheinyl arm of the ACP, guiding the diketide to expose its *si* face to the cofactor.[25,26,34] Supporting this hypothesis, in the related SDR enzyme (an acetoacetyl-CoA reductase, PhaB from *Ralstonia eutropha* H16) such an interaction is

[iii] Because during the course of biosynthesis, the Cahn-Ingold-Prelog *R* and *S* priorities of reduced intermediates can change, when discussing polyketide intermediates the D/L system is used. When discussing small molecule substrates (such as *N*-acetyl cysteamine β–keto thioesters), the more common Cahn-Ingold-Prelog system is appropriate, and is used accordingly.

observed between bound acetoacetyl-CoA and a homologous aspartate.[41] Additionally, mutagenesis experiments have implicated several residues surrounding the active site.[32,38,42,43]



Scheme 1.2 A) Model of KR stereocontrol with the natural substrate, the phosphopantethienyl arm of the ACP. B) Model of KR sterecontrol on substrate analogs. Reproduced with permission.[26]

One of the more mysterious aspects of KRs is the mechanism by which they set α-stereocenters. KRs can be classified into non-epimerizing and epimerizing KRs (denoted by "1" and "2," respectively, *vide supra*).[25] Early isotope-labeling studies with DEBS indicated that while the hydrogens geminal to D-methyl groups were propionate-derived, the hydrogens geminal to L-methyl groups were derived from water. Subsequent experiments confirming this finding were performed through the incubation of DEBS1-TE with deuterium-labeled 2*S*-methylmalonyl-CoA.[22] As spontaneous epimerization from water was determined to be too slow, and no other short-chain dehydrogenase enzymes are known to catalyze epimerization, it was initially hypothesized that this

reaction was mediated by the KS. However, *in vitro* reconstitution experiments did not link the KS to epimerization events,[44,45] and subsequent experimental results with KR domains both *in vitro* and *in vivo* have established their intrinsic epimerase activity.[46,47] To date, the molecular basis for epimerization remains somewhat elusive, as sequence and structural analysis does not reveal any residues that are clearly identifiable as a general base.[25] However, recent studies *in vivo* demonstrate that epimerization can be influenced by the modular context of the KR.[47] A hypothesis presented by Weissman and coworkers to explain this phenomenon is that, perhaps, epimerase activity is mediated by controlling access of water to the KR active site.[47] For racemic small molecule substrate mimics (such as β-keto *N*-acetyl cysteamine thioesters), KRs are believed to undergo a process equivalent to dynamic kinetic resolution, wherein the KR stereospecifically binds and stereoselectively reduces one epimer of the racemic substrate (**Scheme 1.2**).[32,38,42]

Geometric isomerism is introduced by the DH domain, which undergoes a *syn*-coplanar elimination of the KR-installed β-hydroxy group to yield an olefin.[48–50] Consequently, olefins with *trans* configurations result from the dehydration of D-ß-hydroxyls, whereas olefins with *cis* configurations result from L-ß-hydroxyls.[49,50] When α-substitution is present, the KRs associated with modules with harboring DHs are exclusively anti-selective (*i.e.* B1 or A2).[49] The DH domain possesses a double hot-dog fold, and has two catalytic active site residues: a histidine and an aspartatic acid. The aspartatic acid (found in the HPALLD motif) has been hypothesized to act as a general acid, donating a proton to the ß-hydroxy group. The catalytic histidine (found in the HxxxGxxxP motif) acts as a general base, abstracting the α-proton.[48,50] The active sites of *syn*- and *anti*-selective DH domains are remarkably similar, and indeed, some DH

domains perform two sequential dehydrations, one conferring *cis* geometry and the other conferring *trans* geometry (see **Chapter 4**).[51]

Interestingly, as the DH domain performs general acid/general base-catalyzed chemistry, the double hot dog fold has diverged in function to perform isomerizations. Enzymes that are essentially structurally identical to DHs, termed enoyl isomerases (EIs), have been shown to shift a double bond from the α, β position to the γ, δ position.[52–55] Additionally, DH-like domains termed pyran synthases (PSs) have been shown to catalyze cyclization.[56] DH-like domains have also been hypothesized to catalyze polyene geometric isomerization (see **Chapter 4**). In general, DH-like domains that have evolved to perform isomerizations typically possess a sequence motif wherein the catalytic aspartic acid is replaced with an asparagine, or, less commonly, a histidine.[53,56,57] Examples of DHs that possess both pyran synthase activity and dehydratase activity also exist.[57,58]

The ER domain belongs to the medium-chain dehydrogenase/reductase (MDR) superfamily.[14] When α substitution is present, ER domains can confer either D- or L-methyl group stereochemistry. In ERs that have α substitution, when a diagnostic tyrosine residue is present, the ER confers an L-methyl configuration, whereas if this diagnostic tyrosine residue is absent the ER confers a D-methyl configuration.[59,60] In general, ER domains are less well studied than the other β-carbon processing domains; to date, only one structure of an embedded PKS ER has been solved.[33]

The elongating polyketide chain is terminated through the action of a TE domain. TE domains consist of an α/β hydrolase catalytic core, which has a serine-hisitidine-aspartate catalytic triad, akin to the active site found in serine proteases.[14,61] Typically, thioesterases catalyze cyclization (e.g. in the case of the macrolides, such as erythromycin, **Figure 1.2**). However, some TEs catalyze hydrolysis, yielding a linear

polyketide product.[14,61] Understanding how TE domains mediate regio- and stereospecificity is an area of active investigation.[61] The best studied of the TEs is the DEBS TE, which has been shown to have hydrolytic activity toward a broad range of thioester substrates, but has a strict stereochemical preference for cyclization. In DEBS TE, the stereochemical configuration at the nucleophilic hydroxyl and corresponding α-methyl position must be the same as the "natural" hexaketide intermediate (*i.e.*, the stereochemistry conferred by the B2-type KR in module one).[62–64] This stereospecificity appears to be general among macrolactonizing TEs,[65] which has consequences for the rational reprogramming of polyketides through "domain swaps," such as switching KR types within modules of the PKSs. Additionally, the hydrolytic activity towards unnatural small molecule substrates has consequences for PKSs utility in biocatalytic endeavors (see **Chapter 2**).[1,65–68]

PKSs as a Platform for Synthetic Biology

As commented in a recent review by Keasling and coworkers,[69] there is a notable gap between the promise and reality of PKSs as a synthetic biology platform. This is in no small part due to the protein-protein interactions that mediate the collaborative action between polypeptides[14,20] that are disrupted during a "domain swap" approach. Additionally, (as alluded to previously) the interacting selectivities of downstream domains within the module can sometimes hinder the incorporation of varied structural elements, such as introducing different stereochemistries.[20] Although the notion of using a "molecular lego" type approach is intellectually attractive, these attributes render this concept complicated to implement in practice.

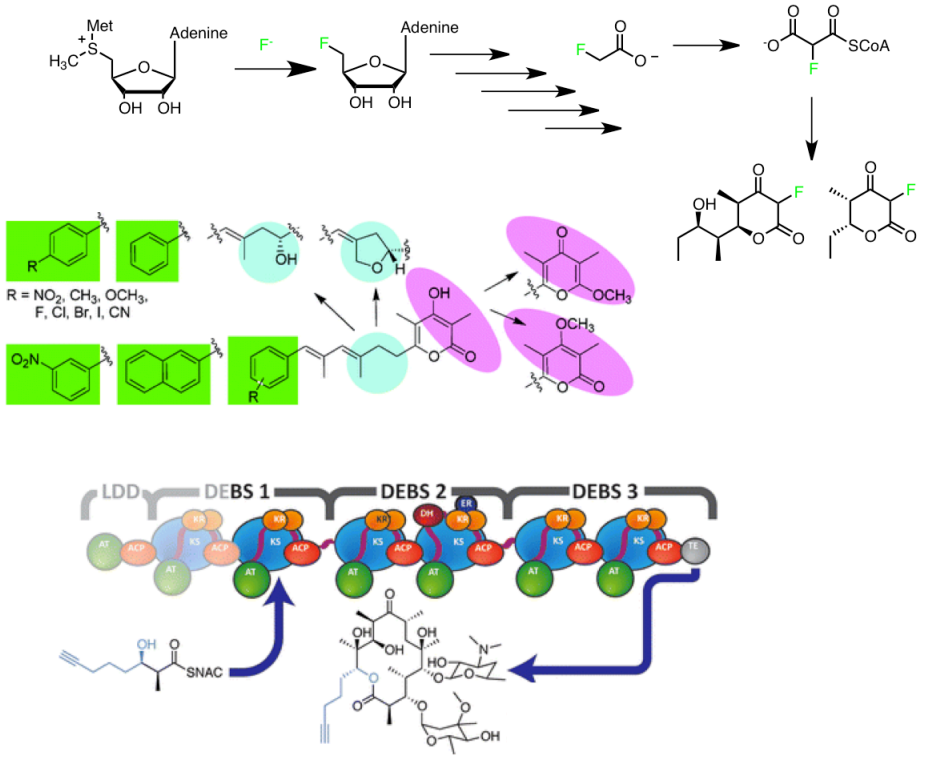


Figure 1.4 Examples of PKS systems used for synthetic biology applications. A) Incorporation of fluorine into DEBS-derived triketide lactones.[70] B) Generation of a focused library of aureothin analogs through a mutasynthetic approach (reproduced with permission). [71] C) Introduction of an alkyne for further modifications in erythromycin (reproduced with permission).[72]

Swaps of individual domains have had some limited success. Early examples of were reported by Menzella and co-workers, who established that this approach can be applied to generate small, chiral PKS fragments (termed triketide lactones) with varying degrees of reduction and stereoisomeric configurations.[73,74] Other examples include applying a "mutasynthetic" approach, wherein the domains are swapped within the synthase to generate chemical diversity through altered specificity of the entire

metabolite's biosynthetic pathway. This mutasynthetic approach has some limited success in the spinosin,[75] erythromycin,[76] and aureothin[77] pathways, particularly when the loading module's specificity is altered. Other strategies include either relying on a degree of intrinsic promiscuity[72,78] or applying a directed evolution approach to altering specificity,[79,80] which has been employed successfully to introducing new chemical functionality (*e.g.* introducing alkynes[72,79] or fluorine[70,81] into PKS scaffolds) through new extender units *via* relaxed or altered specificity of the AT domain. However, frequently these architectural disruptions to the PKS only serve to either abolish or significantly diminish enzymatic activity. These examples of successful synthetic biology incorporations to yield either libraries of compounds[71] or novel functionalities[70,72] within polyketide scaffolds, however elegant, are somewhat limited in scope. In addition to more global "synthetic biology" approaches to engineering a pathway overall, using individual domains or modules for *in vitro* biocatalytic applications also has promise for further harnessing the potential of these remarkable enzymes.

ModTESs as a Biocatalytic Plaform for Generation of Chiral Building Blocks

Despite the proven track record of polyketide scaffolds as lead compounds, their molecular complexity often hampers their development as drug targets. Many pharmaceutical companies have phased out screening complex natural products due to difficulties accessing sufficient quantities for clinical trials.[4,82] A prominent example of such limitations is the bryostatins, a family of 20 marine natural products that were originally isolated from the bryozoan *Bugula nertina*.[83] These structurally complex compounds exhibit exceptional biological activity against cancer as well as neurological activity with promise for treating Alzheimer's disease.[84–88] Unfortunately, their low abundance (approximately 1 gram per ton of bryozoan[89]) remains an insurmountable

hurdle to their development. Unlike most polyketide drugs (such as erythromycin or rapamycin) that are fermented, culturing the strain that produces the bryostations remains elusive and the pathway is intractable to heterologous expression. Indeed, with the exception of eribulin, all polyketide drugs to date are prepared by semi-synthesis or fermentation.[82] An attractive approach for the construction of polyketide scaffolds would be generating small, chiral polyketide chiral synthons that could be then used to generate more complex stereopolyad architectures.[1,90–92]

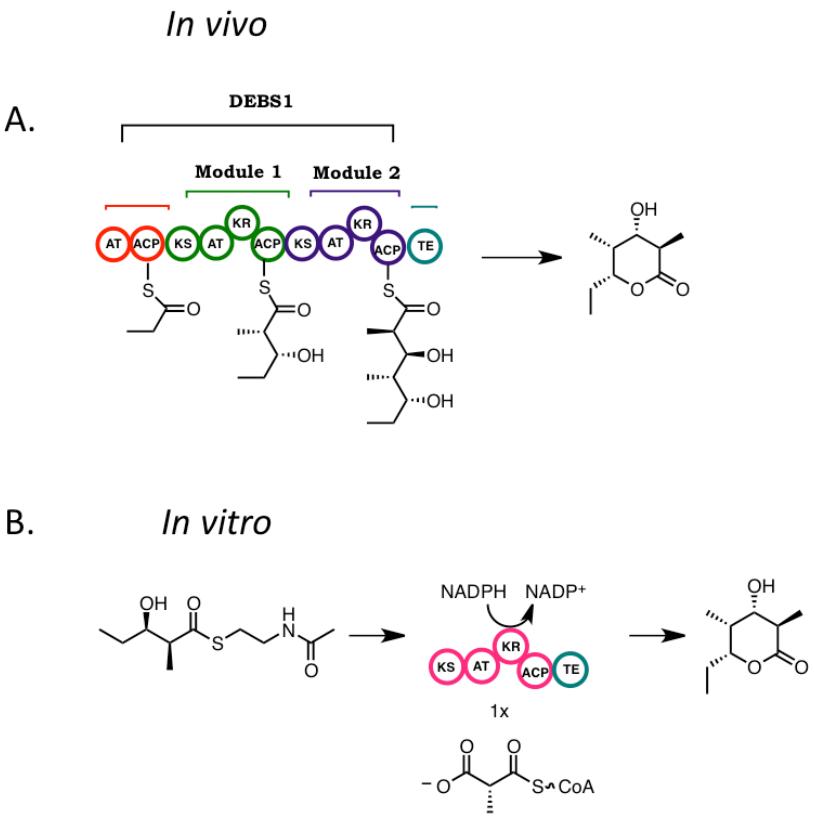


Figure 1.5 A) Schematic of *in vivo* formation of a triketide lactone by DEBS1TE.[93] B) *In vitro* formation of a triketide lactone by priming a ModTE construct with a diketide priming unit.[1,49]

Triketide lactones are the enzymatic products for a several miniaturized polyketide synthase model systems both *in vitro*[94] and *in vivo* (**Figure 1.5**).[95] Early seminal work by Khosla and coworkers demonstrated that DEBS1 could be used as a platform for the generation of one of these compounds (**Figure 1.5**).[95] Building on this result, Leadlay and coworkers demonstrated the *in vivo* efficiency of DEBS1 as a catalyst for triketide lactone formation resulted in much higher titers when it was fused the thioesterase domain.[93] Since these early reports, miniaturizing PKSs to simple bimodular

or modular-TE constructs and using triketide lactone output has been a useful *in vitro* strategy to interrogate enzymatic selectivity.[45,94] However, in addition to their utility as a model system, they could serve as intriguing chiral building blocks.[1] Indeed, one could envision generating a library of all 16 stereochemical combinations of triketide lactone as a library as synthons for the construction of polyketide architectures (**Figure 1.6**). However, typically for enzymology readouts, detection of these compounds is at the trace level (*e.g.* requiring techniques such as mass spectrometry or radio-thin layer chromatography for detection).[94] Thus, developing such molecules requires scaling up this reactivity to a preparative level. In **Chapter 2**, efforts to develop a model ModTE biocatalytic platform for the generation of triketide lactones are described.[1]

Building on this and other work, Sherman and coworkers recently expanded the use of ModTE catalyze the formation of entire macrolides (pikromycin and pikromycin derivatives). This work involved a hybrid of chemical synthesis and biocatalysis, wherein pentaketide intermediates were generated through synthetic chemistry, and extension and macrolactonization were performed by the ModTE construct.[68,96,97] The application of ModTEs for these purposes beyond small molecules such as triketides demonstrates that ModTE constructs are versatile biocatalytic agents that can be applied in various ways to yield chemical diversity.

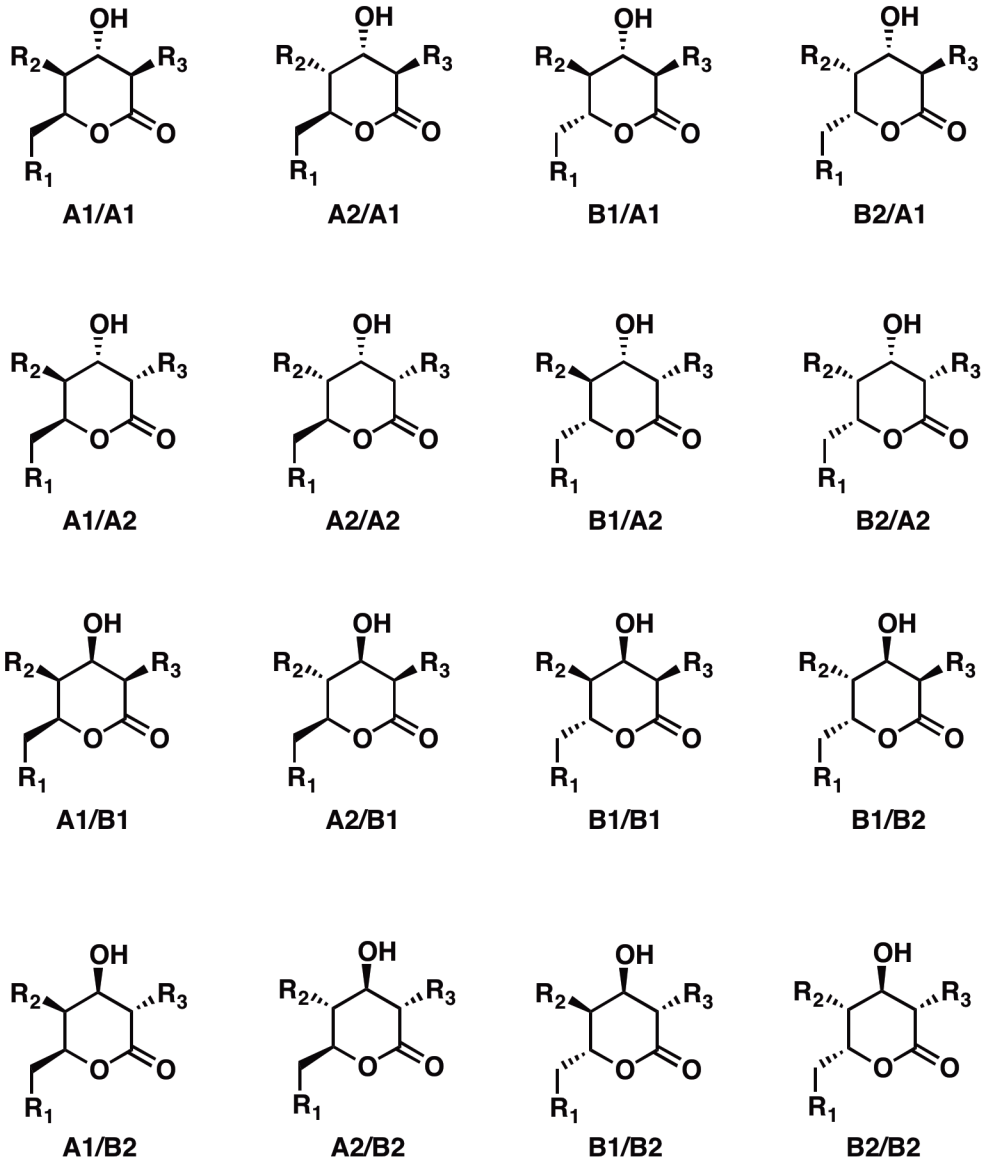
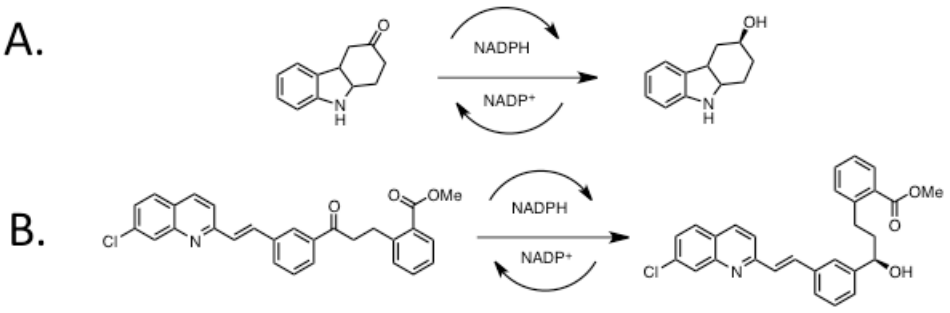


Figure 1.6 Library of all possible stereoisomers of triketide lactone chiral building blocks.

THE KETOREDUCTASE DOMAIN: BIOCATALYTIC APPLICATIONS AND IMPLICATIONS FOR ENZYME ENGINEERING

While the collaborative action of interacting domains in PKSs is crucial for generating varied chemical architectures, there is still much to decipher regarding mechanistic aspects of each individual domain. Additionally, as well as harnessing the collaborative action of PKS domains, individual domains can be developed for their biocatalytic applications. Of all the β-carbon processing domains, KR domains have particular promise for development as biocatalysts.[26] PKS KRs fall within the more general class of enzymes, oxidoreductases. Oxidoreductases, and in particular ketoreductases, are applied in a number of industrial manufacturing processes.[98–100] Indeed, the introduction of an asymmetric hydroxyl functionality is performed by a ketoreductase in a number of drug manufacturing processes (including the blockbuster drugs atorvastin, montelukast, duloxetine, phenylephrine, ezetimibe, and crizotinib) (**Scheme 1.3**).[100] Because many PKS KRs set both the stereochemistry of the β-hydroxyl as well as the α-methyl group, they are particularly attractive for such applications.

Further establishing their biocatalytic potential, KRs have been shown to have a somewhat high degree of substrate promiscuity. Early work by Leadlay and coworkers established that in addition to reducing small molecule analogs (such as β-keto *N*-acetyl cysteamine substrates), KRs possessed activity toward decalone substrates.[101] Additional studies established that NADPH turnover is observed when KRs from the mycolactone, erthryomycin, and tylosin pathways are incubated with a panel of commercially available keto compounds (including 3-heptanone, cyclooctanone, ethylvinyl ketone, dicyclohexyl ketone, and others).[102,103]



Scheme 1.3 Examples of ketoreductases applied in the synthesis of pharmaceutical intermediates. A) Intermediate in the production of the drug rambatroban. B) Intermediate in the production of motolukast.[99]

Activity toward (9*R*, *S*)-*trans*-1-decalone was later used as an output for a saturation mutagenesis study, which explored the role of altering residues surrounding the active site, and those mutations' impacts on stereochemical outcomes. Leadlay and coworkers also generated homology models from other SDR enzymes, which guided rational mutagenesis experiments.[38] Through these experiments, Leadlay and coworkers determined that EryKR1 (a B2 type enzyme) could be made A2 selective with as few as three mutations to residues surrounding the active site.[39] These experiments, then, suggest that subtle perturbations to the active site environment mediate stereocontrol with small molecule mimics. Later mutagenesis efforts were guided by the crystal structures, especially those of AmpKR2[43] and AmpKR11.[32] In the case of AmpKR2, two residues (G55T and Q364H) were sufficient to reverse the diastereoselectivity and increase the efficiency (~4 fold increase in k_{cat}/K_m), resulting in a robust A2 type enzyme.[32]

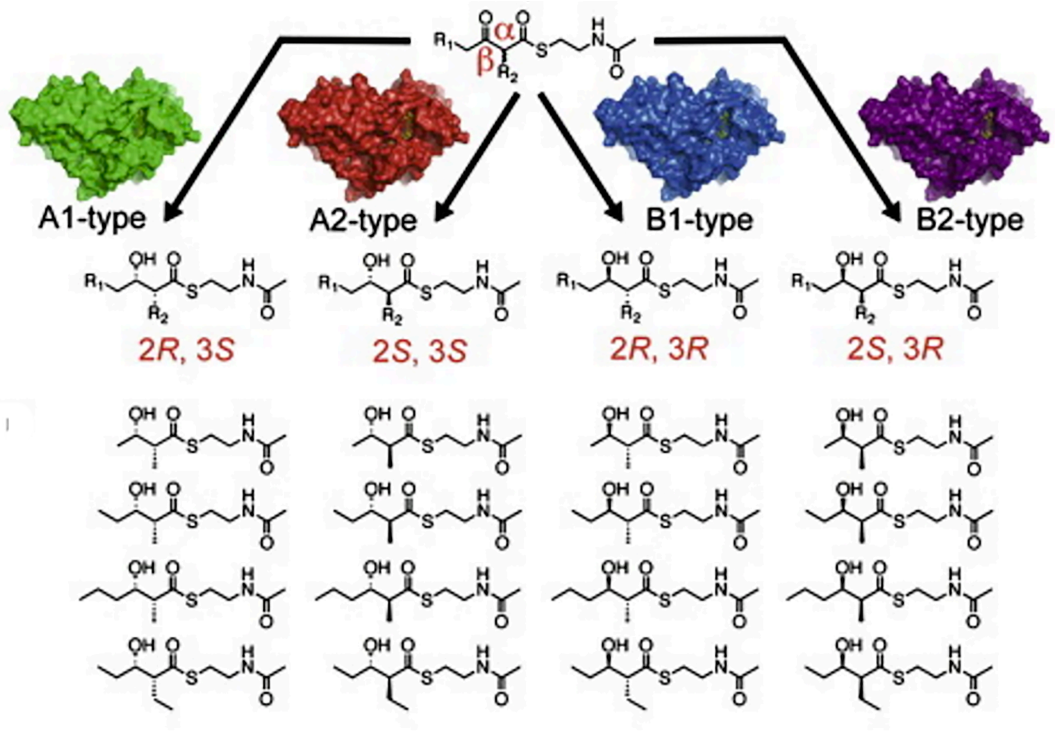


Figure 1.7 Library of diketide building blocks generated by 11 isolated KRs, indicating their biocatalytic utility. Reproduced with permission.[104]

The facility with which the A2 stereoisomer apparently forms when the enzyme is altered through mutagenesis is interesting in light of many KRs activities towards substrates that deviate significantly from the natural substrates. Leadlay and coworkers demonstrated that EryKR2, an A1 type enzyme that naturally accepts a triketide intermediate, has very poor activity towards diketide intermediates (the natural substrate for EryKR2 is a triketide; however δ-hydroxy thioester small molecule substrate mimics spontaneously cyclize in solution and are thus not feasible for *in vitro* assays). What little activity was observed as the A2 *anti* product.[42] A later, expansive study investigated the selectivity of eleven ketoreductases to interrogate their reactivity with five different diketide *S*-NAC thioesters[104] (**Figure 1.7**). This study indicated that KRs frequently

maintain their natural stereocontrol when presented with truncated substrates. In general, KRs that appear early in the synthase (and thus reduce smaller intermediates) tended to retain their stereocontrol; conversely, KRs found later in the biosynthetic pathways tended to have looser substrate selectivity and poorer activity toward diketide substrates. Presumably, this is because KRs found later in biosynthetic pathways have enlarged active sites, which are evolved to accommodate larger PKS substrates.

The A2 stereoisomer is often the major product when the KR does not encounter α-branching in its natural context. For example, MycKRA, which is an A type KR and naturally reduces an unsubstituted keto moiety, preferentially formed the A2 *anti* product over the A1 *syn* product when presented with diketides with α-stereochemistry.[104] Interestingly MycKRB, which is a B type KR from the same pathway (mycolactone), when presented with diketides bearing α-stereochemistry also preferentially forms the *anti* A2 product as the major product, with the *anti* B1 product as the minor product (the *syn* isomers are not observed).[102] Additionally, an odd phenomenon, wherein TylKR1, a naturally B1-type KR switched substrate selectivity when the diketide substrate was truncated by one methyl group, from a 2-methyl-3-oxopentyl-*S*-NAC to a 2-methyl-3-oxobutyl-*S*-NAC thioester, was also observed.[104] This apparent "substrate-dependent enatioselectivity" was further interrogated by Müller and coworkers, who explored the selectivity of TylKR1 using a panel of *S*-NAC and oxo-esters. Frequently, when TylKR1 reduced the more unnatural oxo-ester substrates, the enatioselectivity was reversed, resulting in the A2 product (the enantiomer of the natural stereoisomer) being the major product instead of the B1 product.[105] An approach that harnesses such observations of differences in stereochemical outcome with various substrates and/or point mutations are introduced is further explored in **Chapter 3**.

TRANS-AT PKSS: DEVIATIONS FROM CANONICAL CO-LINEARITY

The majority of the work reviewed above corresponds to "canonical" PKSs, such as erythromycin and other PKSs from filamentous actinomyces bacteria, which are commonly found in the soil.[13] However, in recent years, there has been interest in pursuing a wider variety of taxa and bacterial habitats for the discovery of secondary metabolites. These investigations lead to the discovery of a new class of PKSs, which possess modular megasynthases that have distinctive architectural features that deviate from textbook colinearity.[51] The most notable feature of these PKSs is that each module receives its extender unit from the action of an AT domain that is present on a separate polypeptide. As such, these ATs are termed "*trans*-AT PKSs," whereas textbook PKSs (such as DEBS, which harbor embedded ATs) are termed "*cis*-AT PKSs." The same ATs, then, are used iteratively to acylate each module, and thus typically use the same building block throughout the synthase (malonyl-CoA).[51,106] While *cis*-AT PKSs have only eight different modular organizations (KS-AT-ACP, KS-AT-KR-ACP, KS-AT-DH-KR-ACP, KS-AT-ER-DH-KR-ACP, and their counterparts with methyltransferase domains), over 50 modular organizations have been identified in *trans*-AT pathways.[107] These variant modular organizations frequently possess repeated or missing domains, unusual ordering of domains, or modules split in various ways between two proteins, which altogether serves to obfuscate the metabolite-structure colinearity. However, more recent biosynthetic and phylogenetic analysis has afforded some insights for determining new colinearity rules that are unique to *trans*-AT PKSs. [51]

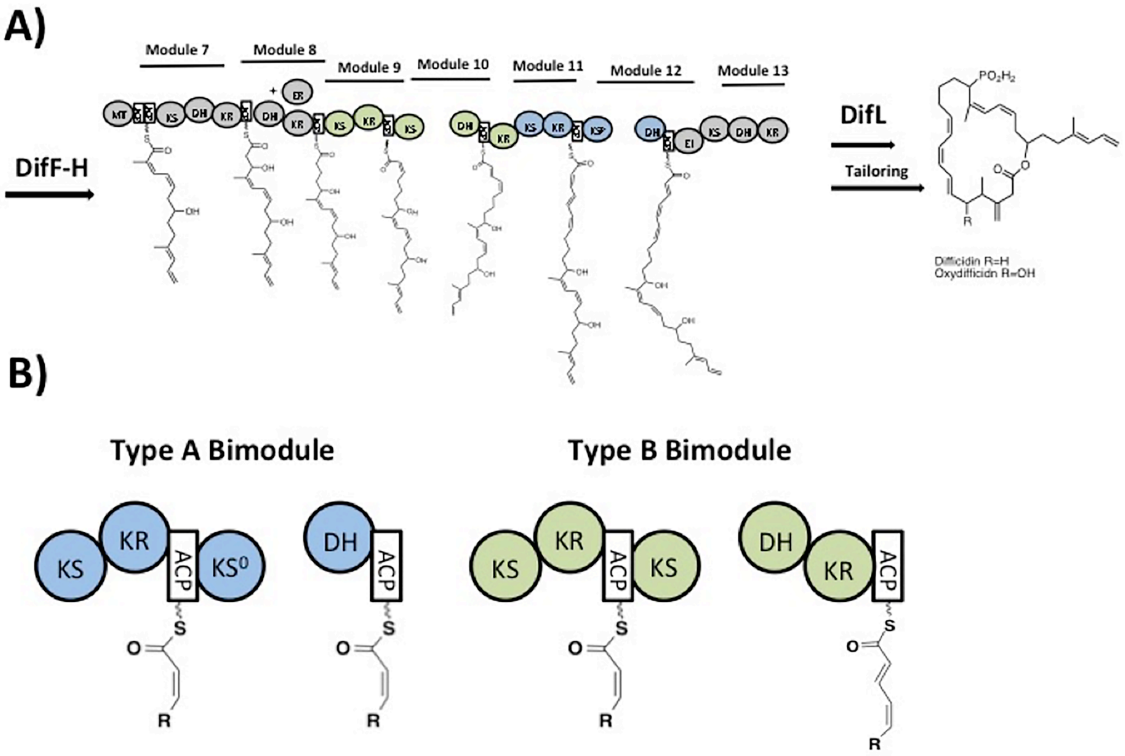


Figure 1.8 An example of a *trans* AT PKS (the difficidin PKS) that contains two tandem split bimodules (a type B bimodule followed by a type A bimodule). B) Schematic of type A and type B bimodules.

One of the interesting deviations from canonical colinearity arises from "split bimodules." Because domains, which must collaboratively process the elongating polyketide chain, are located on separate polypeptides, split modular organizations necessitate protein-protein interactions that mediate chain transfer. Two "split bimodule" types that exemplify the deviant nature of *trans*-AT PKSs in comparison to *cis* AT PKSs are termed by Piel "type A" and "type B" bimodules (**Figure1.8**).[51] Type A bimodules harbor the following domain sequence: KS-KR-ACP-KS^0 (where the KS^0 denotes a KS that is catalytically inactive, identifiable by a lack of one of the active site histidines) on

one gene product, and a subsequent DH on the next polypeptide. In contrast, type B bimodules have a similar organization, except the C-terminal KS is catalytically active, and there is a KR following the N-terminal DH (KS-KR-ACP-KS then DH-KR). One of the clearest illustrations of "new" colinearity rules arising from *trans*-AT pathways is the action of the DH of type B bimodules, which performs two sequential dehydrations. First the hydroxyl arising from the KR from the downstream module is dehydrated, and subsequently the hydroxyl arising from the KR on the upstream module is dehydrated.[51] This likely suggests that the DH must visit two ACPs (one from the upstream module and one from the downstream module). In **Chapter 4**, the structure of a KR from a type A bimodule reveals several structural and sequence features that may explain some of these bridging interactions between the two polypeptides.

OUTLOOK

The colinear, assembly-line nature of PKSs renders them exciting platforms for synthetic biology and biocatalysis. As a more detailed understanding of individual domains interact in unnatural contexts has been uncovered, their applicability has improved in recent years. In addition to canonical systems (such as DEBS), a separate subclass of PKSs, the *trans*-AT PKSs have been recently identified and, even more recently, structurally and functionally dissected. The utility of the *trans*-ATs as a synthetic biology platform will undoubtedly increase, as we gain a better understanding of their unique colinearity rules and protein-protein bridging interactions. Encouragingly, the gap between applications we can envision for PKSs and the reality of their use as a synthetic biology platform has begun to narrow over recent years[69] due to advances in our knowledge, both in terms of their enzymology[108] and of structural features that mediate these protein-protein interactions.[14] The results presented within this dissertation

highlight a few incremental advances in terms of enzymatic mechanism, collaborative action of the domains in *in vitro* environments, and structural dissection that contribute to the ultimate goal of realizing the biocatalytic potential of these enzymes.

REFERENCES

[1] A. D. Harper, C. B. Bailey, A. D. Edwards, J. F. Detelich, A. T. Keatinge-Clay, *ChemBioChem* **2012**, *13*, 2200.

[2] J. N. Brantley, C. B. Bailey, J. R. Cannon, K. a. Clark, D. a. Vandenbout, J. S. Brodbelt, A. T. Keatinge-Clay, C. W. Bielawski, *Angew. Chem.* **2014**, *53*, 5088.

[3] C. Hertweck, *Angew. Chem.* **2009**, *48*, 4688.

[4] J. W.-H. Li, J. C. Vederas, *Science* **2009**, *325*, 161.

[5] A.-M. R. Dechert-Schmitt, D. C. Schmitt, X. Gao, T. Itoh, M. J. Krische, *Nat. Prod. Rep.* **2014**, *31*, 504.

[6] D. Leonori, V. K. Aggarwal, *Acc. Chem. Res.* **2014**, *47*, 3174.

[7] Y. Zhu, K. Burgess, *J. Am. Chem. Soc.* **2008**, *130*, 8894.

[8] J. Li, D. Menche, *Synthesis (Stuttg).* **2009**, *2009*, 2293.

[9] B. Schetter, R. Mahrwald, *Angew. Chem.* **2006**, *45*, 7506.

[10] R. B. Woodward, *Perspectives in Organic Chemistry*, Wile-Interscience, New York, **1956**.

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

[12] C. Khosla, *J. Org. Chem.* **2009**, *74*, 6416.

[13] C. Khosla, Y. Tang, A. Y. Chen, N. a Schnarr, D. E. Cane, *Annu. Rev. Biochem.* **2007**, *76*, 195.

[14] A. T. Keatinge-Clay, *Nat. Prod. Rep.* **2012**, *29*, 1050.

[15] B. J. Dunn, D. E. Cane, C. Khosla, *Biochemistry* **2013**, *52*, 1839.

[16] F. Del Vecchio, H. Petkovic, S. G. Kendrew, L. Low, B. Wilkinson, R. Lill, J. Cortés, B. A. M. Rudd, J. Staunton, P. F. Leadlay, *J. Ind. Microbiol. Biotechnol.* **2003**, *30*, 489.

[17] A. S. Eustáquio, R. P. McGlinchey, Y. Liu, C. Hazzard, L. L. Beer, G. Florova, M. M. Alhamadsheh, A. Lechner, A. J. Kale, Y. Kobayashi, et al., *Proc. Natl. Acad. Sci., U. S. A.* **2009**, *106*, 12295.

[18] A. S. Eustáquio, S.-J. Nam, K. Penn, A. Lechner, M. C. Wilson, W. Fenical, P. R. Jensen, B. S. Moore, *ChemBioChem* **2011**, *12*, 61.

[19] A. F. Marsden, P. Caffrey, J. F. Aparicio, M. S. Loughran, J. Staunton, P. F. Leadlay, *Science* **1994**, *263*, 378.

[20] K. J. Weissman, *Nat. Chem. Biol.* **2014**, *11*, 660.

[21] A. M. Haapalainen, G. Meriläinen, R. K. Wierenga, *Trends Biochem. Sci.* **2006**, *31*, 64.

[22] K. J. Weissman, M. Timoney, M. Bycroft, P. Grice, U. Hanefeld, J. Staunton, P. F. Leadlay, *Biochemistry* **1997**, *36*, 13849.

[23] J. Wu, K. Kinoshita, C. Khosla, D. E. Cane, *Biochemistry* **2004**, *43*, 16301.

[24] D. C. Gay, G. Gay, A. J. Axelrod, M. Jenner, C. Kohlhaas, A. Kampa, N. J. Oldham, J. Piel, A. T. Keatinge-Clay, *Structure* **2014**, *22*, 444.

[25] A. T. Keatinge-Clay, *Chem. Biol.* **2007**, *14*, 898.

[26] J. Zheng, A. T. Keatinge-Clay, *MedChemComm* **2012**. *4*, 34.

[27] R. Reid, M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D. V Santi, C. R. Hutchinson, R. McDaniel, *Biochemistry* **2003**, *42*, 72.

[28] P. Caffrey, *ChemBioChem* **2003**, *4*, 654.

[29] A. T. Keatinge-Clay, *Chem. Biol.* **2007**, *14*, 898.

[30] J. Zheng, A. T. Keatinge-Clay, *J. Mol. Biol.* **2011**, *410*, 105.

[31] S. K. Piasecki, J. Zheng, A. J. Axelrod, M. E Detelich, A. T. Keatinge-Clay, *Proteins* **2014**, *82*, 2067.

[32] J. Zheng, S. K. Piasecki, A. T. Keatinge-Clay, *ACS Chem. Biol.* **2013**, *8*, 1964.

[33] J. Zheng, D. C. Gay, B. Demeler, M. A. White, A. T. Keatinge-Clay, *Nat. Chem. Biol.* **2012**, *8*, 615.

[34] S. A. Bonnett, J. R. Whicher, K. Papireddy, G. Florova, J. L. Smith, K. A. Reynolds, *Chem. Biol.* **2013**, *20*, 772.

[35] A. T. Keatinge-Clay, R. M. Stroud, *Structure* **2006**, *14*, 737.

[36] Y. Yin, R. Gokhale, C. Khosla, D. E. Cane, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1477.

[37] M. L. Mugnai, Y. Shi, A. T. Keatinge-Clay, R. Elber, *Biochemistry* **2015**, *54*, 2346.

[38] A. Baerga-Ortiz, B. Popovic, A. P. Siskos, H. M. O'Hare, D. Spiteller, M. G. Williams, N. Campillo, J. B. Spencer, P. F. Leadlay, *Chem. Biol.* **2006**, *13*, 277.

[39] H. M. O'Hare, A. Baerga-Ortiz, B. Popovic, J. B. Spencer, P. F. Leadlay, *Chem. Biol.* **2006**, *13*, 287.

[40] J. Zheng, C. A. Taylor, S. K. Piasecki, A. T. Keatinge-Clay, *Structure* **2010**, *18*, 913.

[41] J. Kim, J. H. Chang, E.-J. Kim, K.-J. Kim, *Biochem. Biophys. Res. Commun.* **2014**, *443*, 783.

[42] A. P. Siskos, A. Baerga-Ortiz, S. Bali, V. Stein, H. Mamdani, D. Spiteller, B. Popovic, J. B. Spencer, J. Staunton, K. J. Weissman, et al., *Chem. Biol.* **2005**, *12*, 1145.

[43] J. Zheng, C. A. Taylor, S. K. Piasecki, A. T. Keatinge-Clay, *Structure* **2010**, *18*, 913.

[44] J. Wu, T. J. Zaleski, C. Valenzano, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.* **2005**, *127*, 17393.

[45] R. Castonguay, W. He, A. Y. Chen, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.* **2007**, *129*, 13758.

[46] A. Garg, X. Xie, A. Keatinge-Clay, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.* **2014**, *136*, 10190.

[47] T. Annaval, C. Paris, P. F. Leadlay, C. Jacob, K. J. Weissman, *ChemBioChem* **2015**, *16*, 1357.

[48] A. Keatinge-Clay, *J. Mol. Biol.* **2008**, *384*, 941.

[49] C. R. Valenzano, Y.-O. You, A. Garg, A. Keatinge-Clay, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.* **2010**, *132*, 14697.

[50] D. L. Akey, J. R. Razelun, J. Tehranisa, D. H. Sherman, W. H. Gerwick, J. L. Smith, *Structure* **2010**, *18*, 94.

[51] J. Piel, *Nat. Prod. Rep.* **2010**, *27*, 996.

[52] B. Kusebauch, B. Busch, K. Scherlach, M. Roth, C. Hertweck, *Angew. Chem.* **2010**, *49*, 1460.

[53] D. C. Gay, P. J. Spear, A. T. Keatinge-Clay, *ACS Chem. Biol.* **2014**, *9*, 2374.

[54] J. Moldenhauer, D. C. G. Götz, C. R. Albert, S. K. Bischof, K. Schneider, R. D. Süssmuth, M. Engeser, H. Gross, G. Bringmann, J. Piel, *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 1465.

[55] F. Lohr, I. Jenniches, M. Frizler, M. J. Meehan, M. Sylvester, A. Schmitz, M. Gütschow, P. C. Dorrestein, G. M. König, T. F. Schäberle, *Chem. Sci.* **2013**, *4*, 4175.

[56] P. Pöplau, S. Frank, B. I. Morinaka, J. Piel, *Angew. Chem.* **2013**, *52*, 13215.

[57] G. Berkhan, F. Hahn, *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 14240.

[58] R. Ueoka, A. R. Uria, S. Reiter, T. Mori, P. Karbaum, E. E. Peters, E. J. N. Helfrich, B. I. Morinaka, M. Gugger, H. Takeyama, et al., *Nat. Chem. Biol.* **2015**, *11*, 705.

[59] D. H. Kwan, Y. Sun, F. Schulz, H. Hong, B. Popovic, J. C. C. Sim-Stark, S. F. Haydock, P. F. Leadlay, *Chem. Biol.* **2008**, *15*, 1231.

[60] D. H. Kwan, P. F. Leadlay, *ACS Chem. Biol.* **2010**, *5*, 829.

[61] M. E. Horsman, T. P. A. Hari, C. N. Boddy, *Nat. Prod. Rep.* **2015**, DOI 10.1039/c4np00148f.

[62] A. Pinto, M. Wang, M. Horsman, C. N. Boddy, *Org. Lett.* **2012**, *14*, 2278.

[63] K. K. Sharma, C. N. Boddy, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 3034.

[64] S. C. Tsai, L. J. Miercke, J. Krucinski, R. Gokhale, J. C. Chen, P. G. Foster, D. E. Cane, C. Khosla, R. M. Stroud, *Proc. Natl. Acad. Sci., U. S. A.* **2001**, *98*, 14808.

[65] T. P. A. Hari, P. Labana, M. Boileau, C. N. Boddy, *ChemBioChem* **2014**, *15*, 2656.

[66] A. J. Hughes, J. F. Detelich, A. T. Keatinge-Clay, *MedChemComm* **2012**, *3*, 956.

[67] S. Piasecki, A. Keatinge-Clay, *Synlett* **2012**, *23*, 1840.

[68] D. A. Hansen, C. M. Rath, E. B. Eisman, A. R. H. Narayan, J. D. Kittendorf, J. D. Mortison, Y. J. Yoon, D. H. Sherman, *J. Am. Chem. Soc.* **2013**, *135*, 11232.

[69] S. Poust, A. Hagen, L. Katz, J. D. Keasling, *Curr. Opin. Biotechnol.* **2014**, *30C*, 32.

[70] M. C. Walker, B. W. Thuronyi, L. K. Charkoudian, B. Lowry, C. Khosla, M. C. Y. Chang, *Science*. **2013**, *341*, 1089.

[71] M. Werneburg, B. Busch, J. He, M. E. A. Richter, L. Xiang, B. S. Moore, M. Roth, H.-M. Dahse, C. Hertweck, *J. Am. Chem. Soc.* **210**, *132*, 10407.

[72] C. J. B. Harvey, J. D. Puglisi, V. S. Pande, D. E. Cane, C. Khosla, *J. Am. Chem. Soc.* **2012**, *134*, 12259.

[73] H. G. Menzella, R. Reid, J. R. Carney, S. S. Chandran, S. J. Reisinger, K. G. Patel, D. a Hopwood, D. V Santi, *Nat. Biotechnol.* **2005**, *23*, 1171.

[74] H. G. Menzella, J. R. Carney, D. V Santi, *Chem. Biol.* **2007**, *14*, 143.

[75] L. S. Sheehan, R. E. Lill, B. Wilkinson, R. M. Sheridan, W. A. Vousden, A. L. Kaja, G. D. Crouse, J. Gifford, P. R. Graupner, L. Karr, et al., *J. Nat. Prod.* **2006**, *69*, 1702.

[76] J. R. Jacobsen, C. R. Hutchinson, D. E. Cane, C. Khosla, *Science* **2009**, *277*, 367.

[77] B. Busch, N. Ueberschaar, S. Behnken, Y. Sugimoto, M. Werneburg, N. Traitcheva, J. He, C. Hertweck, *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 5285.

[78] M. Werneburg, B. Busch, J. He, M. E. A. Richter, L. Xiang, B. S. Moore, M. Roth, H.-M. Dahse, C. Hertweck, *J. Am. Chem. Soc.* **2010**, *132*, 10407.

[79] U. Sundermann, K. Bravo-Rodriguez, S. Klopries, S. Kushnir, H. Gomez, E. Sanchez-Garcia, F. Schulz, *ACS Chem. Biol.* **2013**, *8*, 443.

[80] I. Koryakina, J. B. McArthur, M. M. Draelos, G. J. Williams, *Org. Biomol. Chem.* **2013**, *11*, 4449.

[81] A. S. Eustáquio, D. O'Hagan, B. S. Moore, *J. Nat. Prod.* **2010**, *73*, 378.

[82] M. J. Yu, W. Zheng, B. M. Seletsky, *Nat. Prod. Rep.* **2013**, *30*, 1158.

[83] G. R. Pettit, C. L. Herald, D. L. Doubek, D. L. Herald, E. Arnold, J. Clardy, *J. Am. Chem. Soc.* **1982**, *104*, 6846.

[84] B. M. Trost, G. Dong, *Nature* **2008**, *456*, 485.

[85] D. L. Alkon, M.-K. Sun, T. J. Nelson, *Trends Pharmacol. Sci.* **2007**, *28*, 51.

[86] Y. Lu, S. K. Woo, M. J. Krische, *J. Am. Chem. Soc.* **2011**, *133*, 13876.

[87] G. E. Keck, Y. B. Poudel, T. J. Cummins, A. Rudra, J. A. Covel, *J. Am. Chem. Soc.* **2011**, *133*, 744.

[88] D. A. Evans, P. H. Carter, E. M. Carreira, A. B. Charette, J. A. Prunet, M. Lautens, *J. Am. Chem. Soc.* **1999**, *121*, 7540.

[89] D. E. Schaufelberger, M. P. Koleck, J. A. Beutler, A. M. Vatakis, A. B. Alvarado, P. Andrews, L. V. Marzo, G. M. Muschik, J. Roach, J. T. Ross, et al., *J. Nat. Prod.* **1991**, *54*, 1265.

[90] Stephen Hanessian, *Pure Appl. Chem.* **1993**, *65*, 1189.

[91] S. Hanessian, *Total Synthesis of Natural Products: The "Chiron" Approach*, Pergamon Press, Oxford, **1883**.

[92] Z. A. Kasun, X. Gao, R. M. Lipinski, M. J. Krische, *J. Am. Chem. Soc.* **2015**, *137*, 8900.

[93] J. Cortes, K. E. Wiesmann, G. A. Roberts, M. J. Brown, J. Staunton, P. F. Leadlay, *Science* **1995**, *268*, 1487.

[94] A. Y. Chen, N. A. Schnarr, C.-Y. Kim, D. E. Cane, C. Khosla, *J. Am. Chem. Soc.* **2006**, *128*, 3067.

[95] C. M. Kao, G. Luo, L. Katz, D. E. Cane, C. Khosla, *J. Am. Chem. Soc.* **1994**, *116*, 11612.

[96] D. A. Hansen, A. A. Koch, D. H. Sherman, *J. Am. Chem. Soc.* **2015**, *137*, 3735.

[97] J. A. Chemler, A. Tripathi, D. A. Hansen, M. O'Neil-Johnson, R. B. Williams, C. Starks, S. R. Park, D. H. Sherman, *J. Am. Chem. Soc.* **2015**, DOI 10.1021/jacs.5b04842.

[98] E. Busto, V. Gotor-Fernández, V. Gotor, *J. Org. Chem.* **2012**, *77*, 4842.

[99] G. W. Huisman, J. Liang, A. Krebber, *Curr. Opin. Chem. Biol.* **2010**, *14*, 122.

[100] G. W. Huisman, S. J. Collier, *Curr. Opin. Chem. Biol.* **2013**, *17*, 284.

[101] L. H. Østergaard, L. Kellenberger, J. Cortés, M. P. Roddis, M. Deacon, J. Staunton, P. F. Leadlay, *Biochemistry* **2002**, *41*, 2719.

[102] S. Bali, K. J. Weissman, *ChemBioChem* **2006**, *7*, 1935.

[103] S. Bali, H. M. O'Hare, K. J. Weissman, *ChemBioChem* **2006**, *7*, 478.

[104] S. K. Piasecki, C. A. Taylor, J. F. Detelich, J. Liu, J. Zheng, A. Komsoukaniants, D. R. Siegel, A. T. Keatinge-Clay, *Chem. Biol.* **2011**, *18*, 1331.

[105] M. Häckh, M. Müller, S. Lüdeke, *Chemistry* **2013**, *19*, 8922.

[106] T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck, J. Piel, *Nat. Biotechnol.* **2008**, *26*, 225.

[107] J. Piel, *Nat. Prod. Rep.* **2010**, *27*, 996.

[108] C. Khosla, D. Herschlag, D. E. Cane, C. T. Walsh, *Biochemistry* **2014**, *53*, 2875.

[109] J. Zhan, *Curr. Top. Med. Chem.* **2009**, *9*, 1958.

[110] C. Hertweck, A. Luzhetskyy, Y. Rebets, A. Bechthold, *Nat. Prod. Rep.* **2007**, *24*, 162.

[111] S. Sheng, S. Zhao, *Sheng Wu Gong Cheng Xue Bao* **2009**, *25*, 1601.

[112] Y.-Q. Cheng, J. M. Coughlin, S.-K. Lim, B. Shen, *Methods Enzymol.* **2009**, *459*, 165.

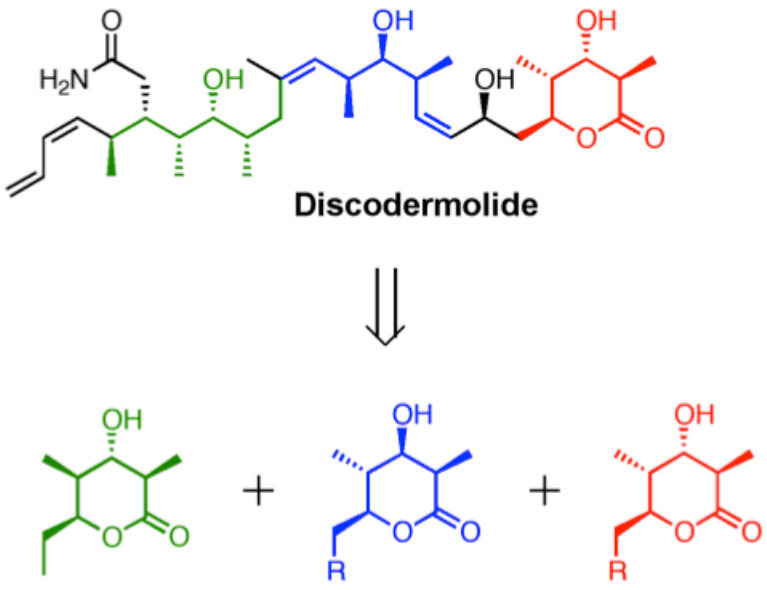
Chapter 2: Preparative Biocatalysis of Triketide Lactone Chiral Building Blocks[iv]

INTRODUCTION

Complex polyketides are structurally-diverse secondary metabolites biosynthesized through successive rounds of decarboxylative Claisen condensations and reductive processing by type I polyketide synthases (PKSs).[1,2] The broad medicinal utility of these compounds has generated intense interest in their production. The antibacterial erythromycin and the immunosuppressant rapamycin are accessed through fermentation; however, this strategy is not always a viable route to a complex polyketide as the producing organism may not be culturable and the biosynthetic pathway may not be amenable to heterologous production.[3] Identifying efficient syntheses of such targets then becomes the bottleneck to their development as therapeutics.[4] Medicinally-relevant marine natural products[5] such as bryostatin[6,7] and discodermolide,[8] are particularly notorious for these difficulties. Chiral precursors can simplify such syntheses, although the current chiral pool is largely limited to sugars, amino acids, and Roche esters.[8,9] PKSs themselves are accessible catalysts that synthesize complex chiral products at room temperature in aqueous conditions. As such, we envisioned harnessing this machinery to generate triketides that could be used as synthetic precursors to accelerate the total syntheses of natural products and their analogs. For example, one can envision constructing the anticancer agent discodermolide from triketide lactones (**Scheme 2.1**).

[iv] Portions of this chapter were reproduced from: Harper, A.D., **Bailey, C.B.,** Edwards, A.D., and Detelich, J.F., and Keatinge-Clay, A.T. *ChemBioChem*. 2012, *13*, 2200. ADH and CBB performed the majority of the experimental work. Specifically, ADH performed the HPLC-based assays and initial enzymatic screens, whereas CBB performed the biocatalytic reactions, isolation, and characterization of triketide lactones. ADE prepared proteins and assisted with initial enzymatic screens and JFD assisted with substrate synthesis. ADH, CBB, and ATK evaluated the experimental data and wrote the original text.

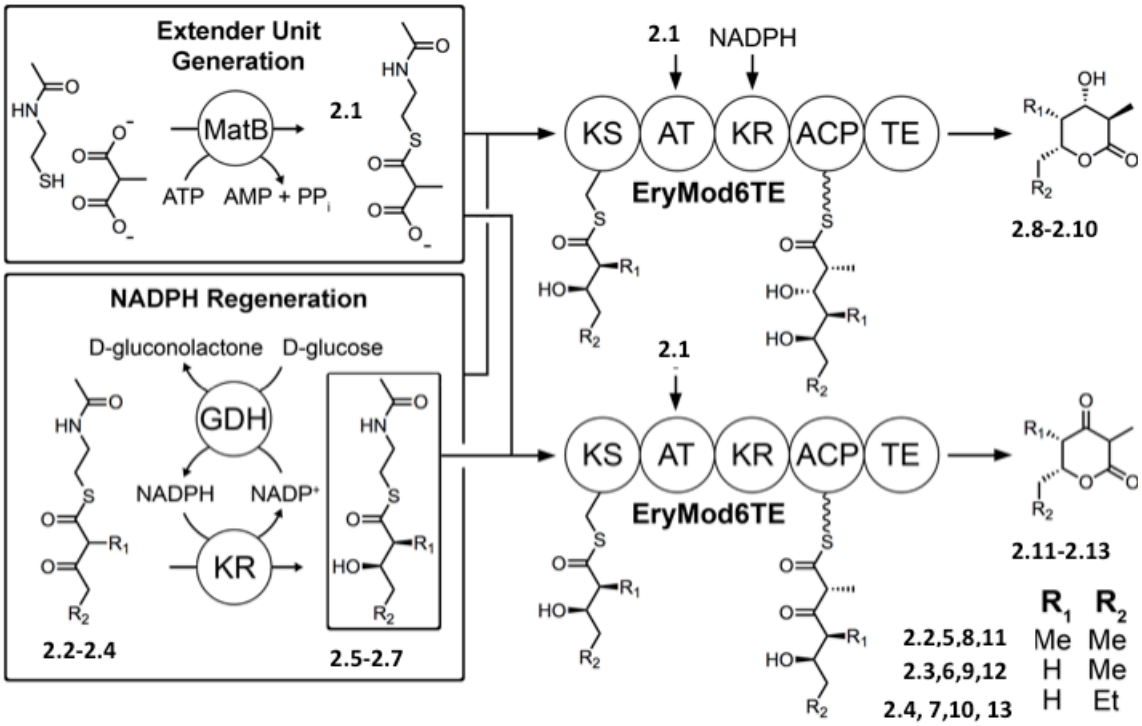
Isolated PKS modules C-terminally fused to a thioesterase (TE) domain (ModTEs) have been shown to generate triketide lactones harboring up to four contiguous stereogenic centers both *in vivo*[10-14] and *in vitro*.[15-18] Such compounds could be applied as synthetic precursors; however, *in vivo* titers rarely exceed a few milligrams per liter[10, 12-15] and, until recently, *in vitro* strategies have only yielded such compounds in trace quantities.[15-18] Herein, we report a versatile in vitro platform that combines enzymatic promiscuity and cofactor regeneration to afford a representative library of triketide lactones produced at unprecedented scales for in vitro PKS synthesis (up to 77 mg). Six triketide lactones were generated in quantities amenable to rigorous characterization. Through the generation of the described library it became apparent that the biocatalytic platform also serves as a powerful strategy to study the activities and selectivities of PKS modules.



Scheme 2.1. Retrosynthetic analysis of the antitumorial agent, discodermolide using triketide lactone chiral synthons.

A ModTE Platform for Triketide Lactone Biocatalysis

The core of this biocatalytic strategy is the terminal module and TE of the erythromycin PKS (EryMod6TE), a module previously demonstrated to accept non-natural substrates[19,20] (**Scheme 2.2**). We employed this miniature enzymatic assembly line to accept and extend a diketide, stereoselectively reduce the resulting triketide intermediate, and catalyze its cyclization.



Scheme 2.2 Chemoenzymatic syntheses of triketide lactone chiral building blocks. In this biocatalytic platform extender units and chiral diketides are generated in separate reactions and then combined with EryMod6TE.

The extension reaction is promoted by excess methylmalonyl-*S*-N-acetylcysteamine (NAC) (**2.1**). This economical substitute for the natural extender unit, 2*S*-methylmalonyl-*S*-CoA, was generated by incubating methylmalonate, NAC, and ATP

with the *Streptomyces coelicolor* malonyl-CoA ligase MatB.[21,22] β-Ketoacyl-*S*-NACs **2.2-2.4** were obtained through facile syntheses, enabling the production of chiral diketide-*S*-NACs **2.5-2.7** through separate biocatalytic reactions.[23] (2*S*,3*R*)-3-Hydroxy-2-methylpentanoyl-*S*-NAC (**2.5**) was generated through the reduction of α-substituted diketide **2** by the first ketoreductase (KR) of the erythromycin PKS (EryKR1), while both (3*R*)-hydroxypentanoyl-*S*-NAC (**2.6**) and (3*R*)-hydroxyhexanoyl-*S*-NAC (**2.7**) were generated through the reduction of α-unsubstituted diketides **2.3** and **2.4** by the first KR of the tylosin PKS (TylKR1). KRs were chosen according to prior studies that identified KRs with high activity and stereoselectivity toward NAC-based substrates.[23] Since purification of these KRs was demonstrated to be unnecessary for *in vitro* biocatalytic reactions, stereoselective reductions of achiral or racemic β-ketoacyl-*S*-NACs were performed in KR-containing dialyzed cell lysate that was supplemented with *Bacillus subtilis* glucose dehydrogenase (GDH), $NADP^+$, and D-glucose (the NADPH-regeneration system).[23] The extender unit generation and diketide reduction reactions were then combined with EryMod6TE-containing lysate from *E. coli* K207-3 cells, expressing the *B. subtilis* phosphopantetheinyl transferase Sfp that phosphopantetheinylates acyl carrier protein (ACP) domains in PKS modules.[20,24] After one day, EryMod6TE reactions were extracted and purified by flash chromatography to yield quantities of triketide lactone suitable for full characterization (4-77 mg).

PREPARATIVE BIOCATALYTIC SYNTHESES

Biocatalytic reactions initiated with diketides **2.2-2.4** (0.5 mmol scale) generated triketide lactones **2.8-2.10** (7%, 9%, and 10% isolated yields, respectively) (**Table 2.1**). While no isolable quantity of ketolactones **2.12** or **2.13** were produced in the reactions generating **2.9** and **2.10**, ketolactone **2.11** was the major product in reactions generating

2.8 (13% vs. 7% isolated yield), resulting from TE operating before EryKR6 on the triketide-*S*-ACP intermediate. The significant amount of ketolactone product in reactions initiated with α-substituted diketide **2.2** suggests that α,γ-dimethyl, β-ketoacyl-*S*-ACP intermediates are better substrates for EryTE than EryKR6, while the small amounts of ketolactone product in reactions initiated with **2.3** and **2.4** suggest the opposite for α-methyl, β-ketoacyl-*S*-ACP intermediates. The previously unmeasured differences in the reactivity of the EryTE and EryKR6 toward various polyketide intermediates demonstrate the utility of this biocatalytic platform to reveal subtle features of PKS enzymology even for a system as well-studied as the erythromycin PKS.

We next examined the scalability of the biocatalytic platform. From a reaction initiated with 7 mmol of **2.4**, 77 mg of **2.10** was readily isolated (6% isolated yield, a decrease from the 10% yield of the 0.5 mmol scale reaction). Triketide lactone **2.10** was easily separated from a small amount of the ketolactone side-product **2.13** (1% isolated yield) by flash chromatography. The scale of this *in vitro* synthesis of a complex polyketide is unprecedented.

Table 2.1. Summary of the triketide lactone isolations. Each row represents a reaction.

Diketide Precursor	Scale (mmol)	NADPH-Regen. Sys.[a]	Product	Product Mass (mg)	Isolated Yield (%)[b]
2.2	0.5	+	**2.8** **2.11**	6.3 11.0	7 13
2.3	0.5	+	**2.9**	7.3	9
2.3	0.5	-	**2.12**	5.6	7
2.4	0.5	+	**2.10**	8.4	10
2.4	0.5	-	**2.13**	3.4	4
2.4	7	+	**2.10** **2.13**	77 14.1	7 1

[a] In the EryMod6TE reaction. [b] Yields reflect quantities of triketide lactones (isolated by flash chromatography) generated from precursors **2.2-2.4.**

MONITORING KETOLACTONE FORMATION

We next sought to monitor the generation of triketide products by the biocatalytic platform. Chiral diketides, produced by KRs from β-ketoacyl-*S*-NACs, were ethyl acetate-extracted from reactions containing the NADPH-regeneration system. The incubation of chiral diketides **2.6** and **2.7** with EryMod6TE but without the NADPH-regeneration system yielded ketolactones **2.12** and **2.13** (7% and 4% isolated yields on a 0.5 mmol scale, respectively). Although the β-ketoesters of these ketolactones are weak chromophores (λ_{max} = 248 nm), products were generated in sufficient concentrations for their synthesis to be monitored with an HPLC coupled to a photodiode array detector. Since β-hydroxyacyl-*S*-NACs and methylmalonyl-*S*-NAC are also UV-active (λ_{max} = 233 nm and λ_{max} = 235 nm, respectively), all reactants and desired products were easily tracked throughout the biocatalytic reactions (**Figures 2.1a** and **2.1b**; experimental

section). The quantities of ketolactone **2.11** generated from diketide **2.5** in this manner facilitated its structural determination by X-ray crystallography (**Figure 2.1c**).

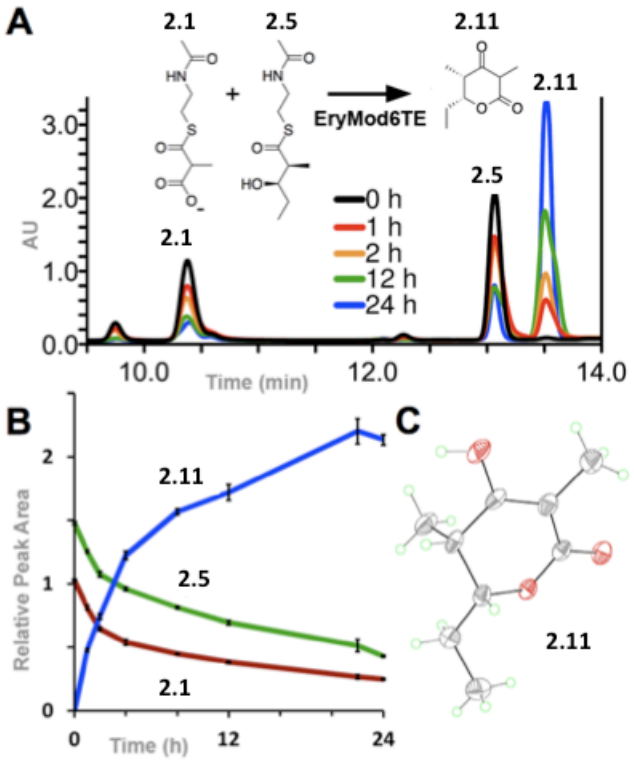


Figure 2.1. A-B) Stacked HPLC chromatograms and a chart reveal the progress of the biocatalytic production of ketolactone **2.11**. C) Oak Ridge Thermal Ellipsoid Plot (ORTEP) representation of the crystal structure of **2.11** (ellipsoids scaled at the 50% probability level).

STEREOCHEMICAL CONSIDERATIONS IN MODTE PLATFORMS

EryMod6TE was then examined for its ability to accept each of the α-methyl, β-hydroxyacyl-S-NAC stereoisomers generated through the reduction of **2.2** by KRs that set different stereochemical combinations at the α- and β-carbons (AmpKR2, RifKR7, TylKR1, and EryKR1).[23,25,26] A strict stereochemical requirement for diketide

incorporation by EryMod6TE was observed in that ketolactone was only produced when EryMod6TE was supplied with **2.5**, which possesses 2*S* and 3*R* stereochemistries. This result was unexpected as this stereoconfiguration is opposite to that of the natural substrate of EryKS6 during erythromycin biosynthesis (its hexaketide-*S*-ACP intermediate possesses 2*R* and 3*S* stereochemistries). Narrow specificity has been previously observed from ModTEs – **2.5** and each of its stereoisomers transthioesterified onto the KS of EryMod2TE but each of the stereoisomers rendered this KS condensation-incompetent.[27] The exclusive incorporation by EryMod6TE of the non-natural (2*S*)-methyl, (3*R*)-hydroxyl diketide into triketide lactones may either indicate differences in KS selectivity towards NAC-bound substrates versus ACP-bound substrates or an inability of EryTE to cyclize the other triketide stereoisomers.[27-30] The described biocatalytic platform will help in determining which domains control ModTE specificities.

Enabling the general extension of diketide stereoisomers may represent the largest challenge to accessing stereochemically diverse libraries of triketide lactone building blocks. We are currently investigating other ModTE constructs such as the second module of the amphotericin PKS fused to the erythromycin TE (AmpMod2TE) since AmpKS2 naturally accepts a (2*S*)-methyl, (3*S*)-hydroxyl diketide. We are also exploring other TEs to address the possibility that EryTE may not be active toward all triketide stereoisomers[27-30] To obtain a ModTE with desired activities it may be necessary to construct chimeras containing KSs, ATs, KRs, and TEs from different modules and synthases.[31]

CONCLUSIONS AND OUTLOOK

While the described biocatalytic system has already yielded a promising first step in accessing triketide lactone chiral building blocks from abundant achiral and racemic precursors, optimization is still necessary. We will attempt to improve the yields compromised by TE-mediated hydrolysis of diketide thioester substrates by employing a promiscuous acyl-CoA synthetase to regenerate hydrolyzed diketide-*S*-NACs *in situ*. We also seek to engineer ModTEs in which KR kinetically outcompetes TE to increase the efficiency of reduced triketide lactone production. Finally, we aim to incorporate chemical handles (*e.g.* terminal olefin, alkyne, chloride) into the building blocks to broaden their synthetic utility.[12,32]

Polyketide synthesis from in vitro PKS reactions has been typically limited to the microgram scale by the cost of substrates and cofactors, particularly CoA-bound extender units and NADPH, and analysis of such in vitro reactions has typically been restricted to sensitive methods such as radio-TLC or mass spectrometry [15-18,26,33,34] The presented platform overcomes such restraints and renders more rigorous characterization techniques (*e.g.* HPLC, NMR, and crystallography) practical through the use of glucose-fueled NADPH regeneration[23] and truncated extender units.[20-22, 35] The increased scale of these in vitro reactions demonstrates that PKS enzymes can indeed be employed as biosynthetic biocatalysts.[36] In addition to helping generate chiral building blocks, the described biocatalytic platform will be valuable in studying the activities and selectivities of enzymes within PKS modules.

ABBREVIATIONS

NAC, *N*-acetylcysteamine; PKS, modular polyketide synthase; KS, ketosynthase; AT, acyltransferase; KR, ketoreductase; ACP, acyl carrier protein; TE, thioesterase; GDH, *Bacillus subtilis* glucose dehydrogenase; NADPH, nicotinamide adenine

dinucleotide phosphate hydrogen; EryTE, erythromycin PKS thioesterase; ModTE, module+thioesterase; EryMod6TE, the 6th module of the erythromycin PKS fused to EryTE; EryMod2TE, the 2nd module of the erythromycin PKS fused to EryTE; AmpMod2TE, the 2nd module of the amphotericin PKS fused to EryTE; AmpKR2, KR from 2nd module of the amphotericin PKS; RifKR7, KR from 7th module of the rifamycin PKS; TylKR1, KR from 1st module of the tylosin PKS; EryKR1, KR from 1st module of the erythromycin PKS; EryKR6, KR from 6th module of the erythromycin PKS; EryKS2, KS from the 2nd module of the erythromycin PKS; AmpKS2, KS from the 2nd module of the amphotericin PKS.

ACKNOWLEDGEMENTS

We thank Seong-Meen Yoon and Andrew Tran for their help with scaled-up biocatalytic reactions. We also thank Vince Lynch for solving the crystal structure of **2.11.** Funding was provided by the Robert A. Welch Foundation (Grant F-1712).

REFERENCES AND NOTES

[1] C. Khosla, Y. Tang, A. Y. Chen, N. A. Schnarr, D. E. Cane, *Annu. Rev. Biochem.* **2007**, *76*, 195.

[2] C. Hertweck, *Angew. Chem.* **2009**, *48*, 4688.

[3] S. F. Brady, L. Simmons, J. H. Kim, E. W. Schmidt, *Nat. Prod. Rep.* **2009**, *11*, 1488.

[4] A. Rivkin, T-. C. Chou, S. J. Danishefsky. *Angew. Chem.* **2005**, *44*, 2838.

[5] R. Montaser, H. Luesch. *Future Med. Chem.* **2011**, *12*, 1475.

[6] B. M. Trost, G. Dong, *Nature*, **2008**, *456*, 485.

[7] B. A. DeChristopher, B. A. Loy, M. D. Marsden, A. J. Schrier, J. A. Zack, P. A. Wender. *Nat. Chem.* **2012**, *4,* 705.

[8] S. J. Mickel, G. H. Sedelmeier, D. Niederer, R. Daeffler, A. Osmani, K. Schreiner, M. Seeger-Weibel, B. Bérod, K. Schaer, R. Gamboni, *Org. Process Res. Dev.* **2004**, *8*, 92.

[9] L. Hilterhaus, A. Liese, *Adv. Biochem. Eng. Biotechnol.* **2007**, *105*, 133.

[10] C. M. Kao, G. Luo, L. Katz, D. E. Cane, C. Khosla, *J. Am. Chem. Soc.* **1994**, *116*, 11612.

[11] J. Cortes, K. E. Weismann, G. A. Roberts, M. J. Brown, J. Staunton, P. F. Leadlay, *Science* **1995**, *268*, 1487.

[12] R. Regentin, J. Kennedy, N. Wu, J. R. Carney, P. Licari, J. Galazzo, R. Desai, *Biotechnol. Prog.* **2004**, *20*, 122.

[13] H. G. Menzella, R. Reid, J. R. Carney, S. S. Chandran, S. J. Reisinger, K. G. Patel, D. A. Hopwood, D. V. Santi, *Nat. Biotechnol.* **2005**, *23*, 1171.

[14] H. G. Menzella, J. R. Carney, Y. Li, D. V. Santi, *Appl. Environ. Microbiol.* **2010**, *76*, 5221.

[15] A. Y. Chen, D. E. Cane, C. Khosla, *Chem. Biol.* **2007**, *14*, 784.

[16] R. Castonguay, W. He, A. Y. Chen, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.* **2007**, *129*, 13758.

[17] M. Oliynyk, M. J. B. Brown, J. Cortés, J. Staunton, P. F. Leadlay, *Chem. Biol.* **1996**, *10*, 833.

[18] K. E. H. Weismann, J. Cortes, M. J. B. Brown, A. L. Cutter, J. Staunton, P. F. Leadlay, *Chem. Biol.* **1995**, *2*, 583.

[19] J. D. Mortison, J.D. Kittendorf, D.H. Sherman, *J. Am. Chem. Soc.* **2009**, *131*, 15784.

[20] A. J. Hughes, J. F. Detelich, A. T. Keatinge-Clay, *MedChemComm.* **2012**, *3,* 956.

[21] N. L. Pohl, M. Hans, H. Y. Lee, Y. S. Kim, D. E. Cane, C. Khosla, *J. Am. Chem. Soc.* **2001**, *123*, 5822.

[22] A. J. Hughes, A. T. Keatinge-Clay, *Chem. Biol.* **2011**, *18*, 165.

[23] S. K. Piasecki, C. A. Taylor, J. F. Detelich, J. Liu, J. Zheng, A. Komsoukaniants, D. R. Siegel, A. T. Keatinge-Clay, *Chem. Biol.* **2011**, *18,* 1331.

[24] S. Murli, J. Kennedy, L. C. Dayem, J. R. Carney, J. T. Kealey, *J. Ind. Microbiol. Biotechnol.* **2003**, *30*, 500.

[25] A. T. Keatinge-Clay, *Chem. Biol.* **2007**, *14*, 898.

[26] C. R. Valenzano, Y. You, A. Keatinge-Clay, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.* **2010**, *132*, 14697.

[27] J. Wu, K. Kinoshita, C. Khosla, D. E. Cane, *Biochemistry* **2004**, *43*, 16301.

[28] R. S. Gokhale, D. Hunziker, D. E. Cane, C. Khosla, *Chem. Biol.* **1999**, *6*, 117.

[29] S. Tsai, L. J. W. Miercke, J. Krucinski, R. Gokhale, J. C. H. Chen, P. G. Foster, D. E. Cane, C. Khosla, R. M. Stroud, *Proc. Natl. Acad. Sci., U. S. A.* **2001**, *98*, 14808.

[30] A. Pinto, M. Wang, M. Horsman, C. N. Boddy, *Org. Lett.* **2012**, *14*, 2278-2281.

[31] L. Kellenberger, I. S. Galloway, G. Sauter, G. Böhm, U. Hanefeld, J. Cortés, J. Staunton, P. F. Leadlay, *ChemBioChem* **2008**, *9*, 2740.

[32] J.B. Harvey, J.D. Puglisi, V.S. Pande, D.E. Cane, C. Khosla. *J. Am. Chem. Soc.* **2012,** *134*, 12259-.

[33] S. Kapur, B. Lowry, S. Yuzawa, S. Kenthirapalan, A. Y. Chen, D. E. Cane, C. Khosla, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 4110.

[34] ^{1}H NMR of reduced triketide lactones has been performed in only one other reported in vitro study: K. J. Weissman, M. Timoney, M. Bycroft, P. Grice, U. Hanefeld, J. Staunton, P. F. Leadlay, *Biochemistry* **1997**, *36*, 13849.

[35] H. Zhao, W. A. van der Donk, *Curr. Opin. Biotechnol.* **2003**, *14*, 583.

[36] A. Kirschning, F. Hahn, *Angew. Chem.* **2012**, *51*, 4012.

Experimental Section for Chapter 2

General Considerations

N-acetylcysteamine (NAC),[1,2] and all β-ketoacyl-*S*-NAC substrates (**2.2-2.4**)[2] were synthesized according to literature procedures. Methylmalonic acid was purchased from TCI America, ATP was purchased from Meiya Pharmaceuticals, and $NADP^+$ was purchased from CalBioChem. IPTG was purchased from either CarboSynth or Anatrace, and Ni-NTA agarose was purchased from Amintra. For purified proteins, final concentrations were determined using a Thermo Scientific Nanodrop 1000. Thin layer chromatography (TLC) was conducted with EMD gel 60 F_{254} pre-coated plates (0.25 mM). Fisher scientific silica gel 60 (particle size 230-400 μm) was used for flash column chromatography. All HPLC monitoring was performed on a Waters 1525 binary HPLC pump connected to a Waters 2998 photodiode array detector using a Varian Microsorb-MV C_{18} column (250 x 4.6 mm, 5 μm particle size, 100 Å pore size) with a matching Metaguard column and mobile phases consisting of water with 0.1% TFA (solvent A) and methanol with 0.1% TFA (solvent B) at a flow rate of 1 mL/min. ^{1}H NMR data were acquired on a Varian Mercury 400 MHz instrument at ambient temperature and are reported in terms of chemical shift (δ ppm), multiplicity, coupling constant, and integration and are referenced downfield from $(CH_3)_4Si$ to the residual solvent peak at 7.26 ppm for $CDCl_3$ as an internal standard. ^{13}C NMR data were acquired on either a Varian Mercury 400 MHz instrument or a Varian Oxford 600 MHz instrument at ambient temperature and are reported in terms of chemical shift and referenced to the residual solvent peak at 77.16 ppm for $CDCl_3$ as an internal standard. High-resolution mass spectrometry measurements were obtained by chemical ionization (CI) with a VG analytical ZAB2-E instrument. LC-MS analysis was performed on an Agilent Technologies 1200 Series HPLC with a Gemini C_{18} column (5 μm, 2 x 50 mm,

Phenomenex) coupled to an Agilent Technologies 6130 quadrupole mass spectrometer system equipped with an electrospray-ionization source. A 5- 95% B gradient over 12 minutes at a flow rate of 0.7 mL/min was run in which the mobile phases were water with 0.1% formic acid (solvent A) and acetonitrile with 0.1% formic acid (solvent B). Specific rotation measurements were made with an Atage AP-300 Automatic Polarimeter.

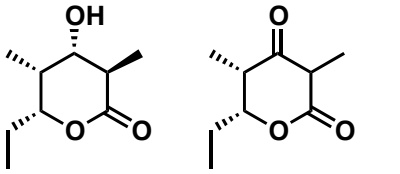
PROTEIN EXPRESSION AND PURIFICATION

Streptomyces coelicolor MatB,[1] TylKR1,[2] EryKR1,[2] AmpKR,[2] MycKR5,[2] RifKR7,[3] and *Bacillus subtilis* glucose dehydrogenase (GDH),[2] were expressed in *E. coli* BL21(DE3), and EryMod6TE[1] was expressed in *E.coli* K207-3[4] (the expression plasmid for all proteins was pET28b, except RifKR7, which was pET28a). Starter cultures (50 mL) were grown to inoculate pre-warmed Luria broth supplemented with 25 μg/mL kanamycin. When OD_{600}=0.4, the media was cooled to 15 °C and then induced with 0.5 mM IPTG. After 16 hours, the protein was harvested by centrifugation (3,000 xg for 20 minutes), and the pellets of 6 L of cell growth were re-suspended in lysis buffer (100 mM HEPES, 500 mM NaCl, 10% v/v glycerol, pH 7.5). The cells were then lysed by sonication on ice and centrifuged (30,000 xg for 45 minutes) to remove cellular debris. For MatB and GDH, the proteins were purified by passing the crude lysate over a nickel-NTA column equilibrated with lysis buffer. The column was washed with lysis buffer containing 15 mM imidazole, and the protein was eluted with lysis buffer containing 150 mM imidazole. For isolated ketoreductases (EryKR1, TylKR1, AmpKR2, and RifKR7) and EryMod6TE, lysate was used. To generate lysate, cells from 6 L of cell growth were pelleted after expression (3,000 xg for 20 minutes), and the pellets of were resuspended in 50 mL lysis buffer before sonication and centrifugation (30,000 xg for 45 minutes). The crude lysate was then twice dialyzed at 4 °C in 10 kDa MWCO cellophane dialysis

tubing against dialysis buffer (30 mM HEPES, 200 mM NaCl, 10% v/v glycerol, pH 7.5), each time for 8 hours.

BIOCATALYTIC SYNTHESES OF TRIKETIDE LACTONES 2.8-2.13

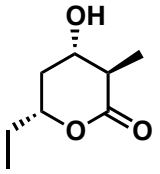
General Considerations for Biocatalytic Syntheses: All biocatalytic reactions were performed at ambient temperature (~23 °C). Reaction progress for the ketoreductase reactions yielding **2.5-2.7** was monitored by TLC (15% MeOH:$CHCl_3$) and HPLC absorbance at 235 nm (linear gradient of 15-35% B over 20 minutes).[2] The progress of reactions yielding methylmalonyl-*S*-NAC (**2.1**) was monitored via HPLC at 235 nm (linear gradient of 0-50% B over 15 minutes). The progress of reactions yielding **2.8-2.13** was monitored via TLC (70% EtOAc:hexanes). TLCs were stained with vanillin.



2.8, 7.3% yield **2.11**, 7.3% yield

(2*R*, 3*S*, 4*S*, 5*R*)-3,5-Dihydroxy-2,4-dimethylheptanoic acid d-lactone (2.8) and (4*S*, 5*R*)-2,4-dimethyl-3-oxoheptanoic acid d-lactone (2.11). Generation of methylmalonyl-S-NAC (2.1) was performed in a 5 mL total volume containing 500 mM HEPES (pH 7.5), 100 mM NaCl, 10% v/v glycerol, 120 mM methylmalonate, 120 mM ATP, 240 mM MgCl2, 100 mM NAC, and 0.5 mg/mL MatB and was reacted for 24 hours. Generation of reduced diketide 5 was performed in a 5 mL total volume containing 500 mM HEPES (pH 7.5), 100 mM NaCl, 15% v/v glycerol, 180 mM D-glucose, 2 mM NADP+, 100 mM **2.2**, 0.05 mg/mL GDH, and 1.3 mL EryKR1 lysate and went to completion after 16 hours. The reactions yielding **2.1** and **2.5** were then combined with 20 mL of

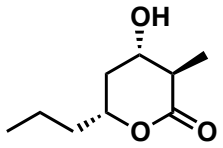
EryMod6TE lysate and incubated in a 30 mL total reaction volume. After 24 hours the reaction was then extracted with EtOAc (3 x 150 mL), dried with MgSO4, filtered, and concentrated under vacuum. The crude extract was then extracted with chloroform (1 x 100 mL) to remove remaining glycerol and concentrated under vacuum. The crude product was next purified *via* flash column chromatography (silica, 50% EtOAc:hexanes) to afford both triketide lactone 2.8 and ketolactone 2.11. Triketide lactone **2.8** was isolated as a yellow oil (6.3 mg, 7.3% yield from 2). Ketolactone **2.11** was isolated as a white solid (11.0 mg, 12.9% yield from 2.2).



2.9, 9.2% yield

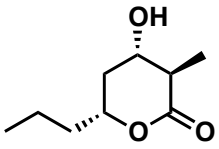
(2*R*, 3*S*, 5*R*)-3,5-Dihydroxy-2-methylheptanoic acid δ-lactone (2.9). Generation of methylmalonyl-*S*-NAC (**2.1**) was performed in a 5 mL total volume containing 500 mM HEPES (pH 7.5), 100 mM NaCl, 10% v/v glycerol, 120 mM methylmalonate, 120 mM ATP, 240 mM $MgCl_2$, 100 mM NAC, and 0.5 mg/mL MatB and was reacted for 24 hours. Generation of reduced diketide **2.6** was performed in a 5 mL total volume containing 500 mM HEPES (pH 7.5), 100 mM NaCl, 15% v/v glycerol, 180 mM D-glucose, 2 mM $NADP^+$, 100 mM **2.3,** 0.05 mg/mL GDH, and 1.3 mL TylKR1 lysate and went to completion after 16 hours. The reactions yielding **2.1** and **2.6** were then combined with 20 mL of EryMod6TE lysate and incubated in a 30 mL total reaction volume. After 24 hours the reaction was extracted with ethyl acetate (3 x 150 mL), dried with $MgSO_4$, filtered, and concentrated under vacuum. The crude extract was extracted with chloroform (1 x 100 mL) to remove remaining glycerol and concentrated under

vacuum. The crude product was then purified via flash column chromatography (silica, 60% EtOAc:hexanes) to afford triketide lactone **2.9** as a yellow oil (7.3 mg, 9.2% isolated yield from **2.3**). No ketolactone side-product (**2.12**) was isolated.



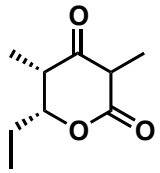
2.10, 9.6% yield

(2*R*, 3*S*, 5*R*)-3,5-Dihydroxy-2-methyloctanoic acid δ-lactone (2.10). Generation of methylmalonyl-*S*-NAC (**2.1**) was performed in a 5 mL total volume containing 500 mM HEPES (pH 7.5), 100 mM NaCl, 10% v/v glycerol, 120 mM methylmalonate, 120 mM ATP, 240 mM $MgCl_2$, 100 mM NAC, and 0.5 mg/mL MatB and was reacted for 24 hours. Generation of reduced diketide **2.7** was performed in a 5 mL total volume containing 500 mM HEPES (pH 7.5), 100 mM NaCl, 15% v/v glycerol, 5% v/v DMSO, 180 mM D-glucose, 2 mM $NADP^+$, 100 mM **2.4**, 0.05 mg/mL GDH, and 1.3 mL TylKR1 lysate. The reaction went to completion after 16 hours. The reactions yielding **2.1** and **2.7** were then combined with 20 mL of EryMod6TE lysate to a 30 mL total reaction volume. After 24 hours the reaction was extracted with ethyl acetate (3 x 150 mL), dried with $MgSO_4$, filtered, and concentrated under vacuum. The crude extract was then extracted with chloroform (1 x 100 mL) to remove remaining glycerol and concentrated under vacuum. The crude product was then purified via dry flash column chromatography (silica, 70% EtOAc:hexanes) to afford compound **2.10** as a yellow oil (8.4 mg, 9.6% yield from **4**). No ketolactone side-product (**2.13**) was isolated.



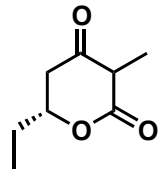
2.10, 6.4% yield, 77 mg

Preparative Biocatalysis of (2*R*, 3*S*, 5*R*)-3,5-Dihydroxy-2-methyloctanoic acid δ-lactone (2.10). Generation of methylmalonyl-S-NAC (**2.1**) was performed in a 70 mL total volume containing 500 mM HEPES (pH 7.5), 100 mM NaCl, 10% v/v glycerol, 120 mM methylmalonate, 120 mM ATP, 240 mM $MgCl_2$, 113 mM NAC, and 0.42 mg/mL MatB and went to completion in 36 hours. Generation of reduced diketide **2.7** was performed in a 70 mL total volume containing 500 mM HEPES (pH 7.5), 100 mM NaCl, 15% v/v glycerol, 5% v/v DMSO, 180 mM D-glucose, 2 mM $NADP^+$, 100 mM **2.4**, 0.05 mg/mL GDH, and 18.2 mL TylKR1 lysate and went to completion after 16 hours. The reactions yielding **2.1** and **2.7** were then combined with 190 mL EryMod6TE lysate and reacted in a total reaction volume of 330 mL. After two days the reactions were extracted with EtOAc (6 x 250 mL). The remaining emulsion was centrifuged (1000 xg for 1 minute) and the organic phase was removed. The remaining aqueous phase was then extracted with EtOAc (2 x 250 mL), and the remaining emulsion was centrifuged again (1000 xg for 1 minute) and the organic phase was again removed. The combined organic phases were dried with $MgSO_4$, filtered, and concentrated under vacuum. The crude extract was then extracted with chloroform (2 x 200 mL) to remove remaining glycerol and concentrated. The crude product was next purified via flash column chromatography (silica, 50% EtOAc:hexanes) to afford **2.10**, a yellow oil (77 mg, 6.4% yield from **2.4**). The ketolactone lactone side-product, **2.13,** was also isolated as a grey solid (14.1 mg, 1.18% yield from **2.4**).



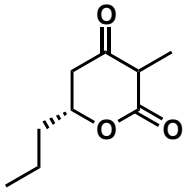
2.12, 7.2% yield

(5*R*)-2-methyl-3-oxoheptanoic acid δ-lactone (2.12). Generation of methylmalonyl-S-NAC (**2.1**) was performed in a 5 mL total volume containing 500 mM HEPES (pH 7.5), 100 mM NaCl, 10% v/v glycerol, 120 mM methylmalonate, 120 mM ATP, 240 mM $MgCl_2$, 100 mM NAC, and 0.5 mg/mL MatB and was reacted for 24 hours. Generation of reduced diketide **2.6** was performed in a 5 mL total volume containing 500 mM HEPES (pH 7.5), 100 mM NaCl, 15% v/v glycerol, 5% v/v DMSO,180 mM D-glucose, 2 mM $NADP^+$, 100 mM **2.3,** 0.05 mg/mL GDH, and 1.3 mL TylKR1 lysate and went to completion after 16 hours. The reaction yielding reduced diketide **2.6** was then extracted with EtOAc (3 x 50 mL), dried with $MgSO_4$, filtered, and concentrated under vacuum to isolate reduced diketide **2.6** from the NADPH regeneration system. Reduced diketide **2.6** was diluted in 5 mL 500 mM HEPES (pH 7.5) and combined with the reaction yielding **2.1** and 20 mL of EryMod6TE lysate. The reaction was incubated in a total volume of 30 mL. After 24 hours, the reaction was extracted with EtOAc (3 x 150 mL), dried with $MgSO_4$, filtered, and concentrated under vacuum. The crude extract was then extracted with chloroform (1 x 100 mL) to remove remaining glycerol and concentrated to afford the crude product. The crude product was next purified via dry flash column chromatography (silica, 50% EtOAc:hexanes) to afford ketolactone **2.12**, a white solid (5.6 mg, 7.2% yield from **2.3**).



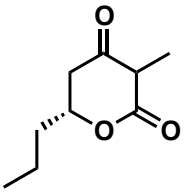
2.12, 7.2% yield

(5*R*)-2-methyl-3-oxoheptanoic acid δ-lactone (2.12). Generation of methylmalonyl-S-NAC (**2.1**) was performed in a 5 mL total volume containing 500 mM HEPES (pH 7.5), 100 mM NaCl, 10% v/v glycerol, 120 mM methylmalonate, 120 mM ATP, 240 mM $MgCl_2$, 100 mM NAC, and 0.5 mg/mL MatB and was reacted for 24 hours. Generation of reduced diketide **2.6** was performed in a 5 mL total volume containing 500 mM HEPES (pH 7.5), 100 mM NaCl, 15% v/v glycerol, 5% v/v DMSO,180 mM D-glucose, 2 mM $NADP^+$, 100 mM **2.3,** 0.05 mg/mL GDH, and 1.3 mL TylKR1 lysate and went to completion after 16 hours. The reaction yielding reduced diketide **2.6** was then extracted with EtOAc (3 x 50 mL), dried with $MgSO_4$, filtered, and concentrated under vacuum to isolate reduced diketide **2.6** from the NADPH regeneration system. Reduced diketide **2.6** was diluted in 5 mL 500 mM HEPES (pH 7.5) and combined with the reaction yielding **2.1** and 20 mL of EryMod6TE lysate. The reaction was incubated in a total volume of 30 mL. After 24 hours, the reaction was extracted with EtOAc (3 x 150 mL), dried with $MgSO_4$, filtered, and concentrated under vacuum. The crude extract was then extracted with chloroform (1 x 100 mL) to remove remaining glycerol and concentrated to afford the crude product. The crude product was next purified via dry flash column chromatography (silica, 50% EtOAc:hexanes) to afford ketolactone **2.12**, a white solid (5.6 mg, 7.2% yield from **2.3**).



2.13, 4% yield

(5*R*)-2-methyl-3-oxooctanoic acid δ-lactone (2.13). Generation of methylmalonyl-S-NAC (**2.1**) was performed in a 5 mL total volume containing 500 mM HEPES (pH 7.5), 100 mM NaCl, 10% v/v glycerol, 120 mM methylmalonate, 120 mM ATP, 240 mM $MgCl_2$, 100 mM NAC, and 0.5 mg/mL MatB and was reacted for 24 hours. Generation of reduced diketide **7** was performed in a 5 mL total volume containing 500 mM HEPES (pH 7.5), 100 mM NaCl, 15% v/v glycerol, 180 mM D-glucose, 5% v/v DMSO, 2 mM $NADP^+$, 100 mM **2.4,** 0.05 mg/mL GDH, and 1.3 mL TylKR1 lysate and went to completion after 16 hours. The reaction yielding reduced diketide **2.7** was then extracted with EtOAc (3 x 50 mL), dried with $MgSO_4$, and concentrated under vacuum to isolate reduced diketide **7** from the NADPH regeneration system. Reduced diketide **2.7** was then diluted in 5 mL 500 mM HEPES (pH 7.5) and combined with the reaction yielding **2.1** and 20 mL of EryMod6TE lysate. The reaction was incubated in a total volume of 30 mL. After 24 hours the reaction was extracted with EtOAc (3 x 150 mL), dried with $MgSO_4$, filtered, and concentrated under vacuum. The crude extract was then extracted with chloroform (1 x 100 mL) to remove glycerol and concentrated to afford the crude product. The crude product was next purified via dry flash column chromatography (silica, 50% EtOAc:hexanes) to afford ketolactone **2.13**, a grey solid (3.4 mg, 4.0% yield from **2.4**).



2.13, 4% yield

(5*R*)-2-methyl-3-oxooctanoic acid δ-lactone (2.13). Generation of methylmalonyl-S-NAC (**2.1**) was performed in a 5 mL total volume containing 500 mM HEPES (pH 7.5), 100 mM NaCl, 10% v/v glycerol, 120 mM methylmalonate, 120 mM ATP, 240 mM $MgCl_2$, 100 mM NAC, and 0.5 mg/mL MatB and was reacted for 24 hours. Generation of reduced diketide **2.7** was performed in a 5 mL total volume containing 500 mM HEPES (pH 7.5), 100 mM NaCl, 15% v/v glycerol, 180 mM D-glucose, 5% v/v DMSO, 2 mM $NADP^+$, 100 mM **2.4,** 0.05 mg/mL GDH, and 1.3 mL TylKR1 lysate and went to completion after 16 hours. The reaction yielding reduced diketide **2.7** was then extracted with EtOAc (3 x 50 mL), dried with $MgSO_4$, and concentrated under vacuum to isolate reduced diketide **7** from the NADPH regeneration system. Reduced diketide **2.7** was then diluted in 5 mL 500 mM HEPES (pH 7.5) and combined with the reaction yielding **2.1** and 20 mL of EryMod6TE lysate. The reaction was incubated in a total volume of 30 mL. After 24 hours the reaction was extracted with EtOAc (3 x 150 mL), dried with $MgSO_4$, filtered, and concentrated under vacuum. The crude extract was then extracted with chloroform (1 x 100 mL) to remove glycerol and concentrated to afford the crude product. The crude product was next purified via dry flash column chromatography (silica, 50% EtOAc:hexanes) to afford ketolactone **2.13**, a grey solid (3.4 mg, 4.0% yield from **2. 4**).

CHARACTERIZATION

***(2R, 3S, 4S, 5R)-3,5-Dihydroxy-2,4-dimethylheptanoic acid δ-lactone* (8).** ^1H NMR (400 MHz, $CDCl_3$) δ= 4.13 (m, 1H), 3.82 (dd, J=4.5 Hz, J=10 Hz, 1H), 2.47 (dq, J=4.5 Hz, 7 Hz, 1H), 2.17 (m, 1H), 1.83 (m, 1H), 1.55 (m, 1H), 1.41 (d, J=7 Hz, 3H), 1.00 (t, J=7 Hz, 3H), 0.97 (d, J=7 Hz, 3H). ^{13}C NMR (150 MHz, $CDCl_3$) δ=173.51, 81.31, 73.96,

39.84, 36.75, 25.26, 14.26, 9.86, 4.35. $[\alpha]_D^{23}$ = +33 (c=0.21, $CHCl_3$). HRMS (CI) (*m/z*) $[M+H]^+$: calcd. for $C_9H_{16}O_3$: 173.1178, found: 173.1176. This characterization is in agreement with literature reported data.[6]

***(2R, 3S, 5R)-3,5-Dihydroxy-2-methylheptanoic acid δ-lactone* (2.9)**. ^1H NMR (400 MHz, $CDCl_3$) δ= 4.16 (m, 1H), 3.76 (td, 10.5, 4.5 1H), 2.35 (m, 1H), 2.20 (m, 1H), 1.82-1.62 (m, 3H), 1.41 (d, J=7 Hz, 3H), 1.00 (t, J=7 Hz, 3H). ^{13}C NMR (100 MHz, $CDCl_3$) 173.53, 77.89, 70.48, 45.24, 37.84, 28.94, 13.63, 9.29. $[\alpha]_D^{23}$= +31 (c=0.30, $CHCl_3$). HRMS (CI) (*m/z*) $[M+H]^+$: calcd. for C_8H_{14}: 159.1021, found: 159.1020.

***(2R, 3S, 5R)-3,5-Dihydroxy-2-methyloctanoic acid δ-lactone* (2.10)**. ^1H NMR (400 MHz, $CDCl_3$) δ= 4.24 (m, 1H), 3.76 (m, 1H), 2.35 (m, 1H), 2.20 (m, 1H), 1.78-1.46 (m, 5H), 1.40 (d, J=7 Hz, 3H), 0.92 (t, J=7 Hz, 3H). ^{13}C NMR (100 MHz, $CDCl_3$). δ= 174.06, 76.70, 70.15, 45.15, 38.26, 38.01, 18.12, 13.87, 13.62. $[\alpha]_D^{21}$= +43 (c=1.01, $CHCl_3$). HRMS (CI) (*m/z*) $[M+H]^+$: calcd. for $C_9H_{16}O_3$: 173.1178, found 173.1178.

***(4S, 5R)-2,4-Dimethyl-3-oxoheptanoic acid δ-lactone* (2.11)**. ^1H NMR (400 MHz, $CDCl_3$) δ= 4.65 (m, 1H), 3.61 (q, J=7 Hz, 1H), 2.62 (dq, J=7 Hz, J=3.2 Hz, 1H), 1.85 (m, 1H), 1.65 (m, 1H), 1.36 (d, J=7 Hz, 3H), 1.12 (d, J=7 Hz, 3H), 1.07 (t, J=7 Hz, 3H).[v] ^{13}C NMR (100 MHz, $CDCl_3$) δ= 205.59, 170.23, 78.68, 50.55, 44.52, 24.18, 10.07, 9.88, 8.39. HRMS (CI) (*m/z*): calcd. for $C_9H_{14}O_3$ $[M+H]^+$: 171.1021, found: 171.1020.

***(5R)-2-Methyl-3-oxoheptanoic acid δ-lactone* (2.12)**. ^1H NMR (400 MHz, $CDCl_3$) δ= 4.64 (m, 1H), 3.57 (q, J=6.5 Hz, 1H), 2.74 (dd, J=2.8 Hz, J=19 Hz, 1H), 2.44 (dd, J=12 Hz, J=19 Hz, 1H), 1.90-1.75 (m, 2H), 1.37 (d, J=6.5 Hz, 3H), 1.08 (t, J=7 Hz, 3H). ^{13}C NMR (100 mHz, $CDCl_3$) δ= 201.71, 170.07, 75.46, 51.80, 42.83, 27.65, 9.33, 7.93.

[v] Note that the ^1H splitting suggests that **2.11** is present as the keto tautomer in solution, which is in contrast to the solid-state (Figure C 6) where it is present as the enol tautomer.

HRMS (CI) (*m/z*) $[M+H]^+$: calcd. for $C_8H_{12}O_3$: 157.0865, found 157.0864.

***(5R)-2-Methyl-3-oxooctanoic acid δ-lactone* (2.13)**. ^1H NMR (400 MHz, $CDCl_3$) δ= 4.70 (m, 1H), 3.56 (q, J=6.5 Hz, 1H), 2.73 (dd, J=2.7 Hz, J=19 Hz, 1H), 2.47 (dd, J=12 Hz, J=18 Hz, 1H), 1.6-1.4 (m, 3H), 1.36 (d, J=6.5 Hz, 3H), 0.99 (t, J=7 Hz, 3H). ^{13}C NMR (100 MHz, $CDCl_3$) δ=201.76, 170.13, 74.04, 51.82, 43.22, 36.47, 18.16, 13.79, 7.83. HRMS (CI) (*m/z*) $[M+H]^+$: cacld. for $C_9H_{14}O_3$: 171.1021, found: 171.1021.

MONITORING KETOLACTONE FORMATION VIA HPLC

Generation of methylmalonyl-S-NAC (**2.1**) was performed in 500 mM HEPES (pH 8.0), 100 mM NaCl, 10% v/v glycerol, 100 mM methylmalonate, 100 mM ATP, 200 mM $MgCl_2$, 100 mM NAC, and 0.5 mg/mL MatB and was reacted for 48 hours. Generation of reduced diketides **2.5-2.7** was performed in 500 mM HEPES (pH 7.5), 100 mM NaCl, 15% v/v glycerol, 5% v/v DMSO, 300 mM D-glucose, 1 mM $NADP^+$, 100 mM β-ketoacyl-S-NAC **2.2**-**2.4,** 0.05 mg/mL GDH, and 20% v/v KR-containing lysate (EryKR1 for **2.5**, TylKR1 for **2.6**-**2.7**) and went to completion after 16 hours. β-hydroxyacyl-S-NACs **2.5**-**2.7** were extracted with ethyl acetate (3 x 100 mL), and concentrated under vacuum. Triplicate ketolactone reactions consisted of 250μL MatB reaction, 1 mL EryMod6TE lysate, an amount of extracted β-hydroxyacyl-S-NAC equivalent to that of methylmalonyl-S-NAC estimated by HPLC peak area in a total volume of 2.1 mL. Timepoints were taken at 0, 1, 2, 4, 8, 12, 22, and 24 hr by quenching 200 μL of the reaction with an equal volume of 2 M acetic acid in methanol and storing at -20 °C until analyzed. Before analysis, timepoint samples were centrifuged at 21,000 xg for 1 min (to pellet precipitate) and the supernatant was filtered through a 0.45 μm nylon membrane. For reactions yielding **2.11-2.13**, 20-minute linear gradients (5-20% solvent B) were performed. For reactions yielding **2.12**, an additional isocratic run (30% B) was necessary

to separate **2.6** from **2.12**. 25 μL of the quenched reaction was injected for each timepoint and absorbance was monitored at 242 nm.[vi] Ketolactone formation was further confirmed via LC-MS (**Table E2.1**). Completion of KR reactions was judged by HPLC (15-minute gradient of 15-35% B, monitored at 234 nm).

[vi] Note that when monitoring the disappearance of compound **1**, the compound eluting directly before **1** with a retention time of 9.3 minutes is residual acetyl-SNAC thioester, a minor impurity from the synthesis of N-acetylcysteamine.

Table E2.1 LC/ESI-MS (*m/z*) $[M+H]^+$ confirming ketolactone formation in time course assays.

Compound	Formula	Expected	Found
2.11	$C_9H_{14}O_3$	171.09	171.4
2.12	$C_8H_{12}O_3$	157.08	157.2
2.13	$C_9H_{14}O_3$	171.09	171.2

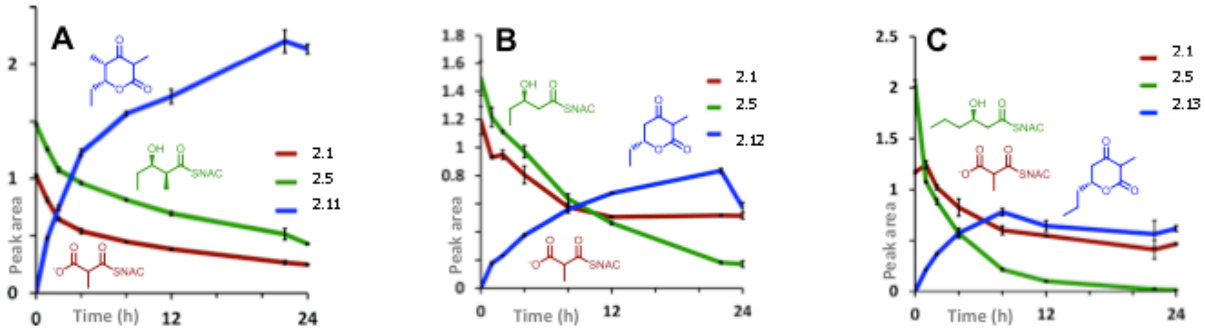


Figure E2.1 Time course assays quantifying the formation of ketolactones **2.11-2.13** and the disappearance of both the reduced diketide priming unit (**2.5-2.7**) and extender unit **2.1** as monitored by HPLC peak area (in arbitrary units). A) Formation of ketolactone **2.11** primed with diketide **2.5**. B) Formation of ketolactone **2.12** primed with diketide **2.6**. C) Formation of ketolactone **2.13** primed with diketide **2.7.**

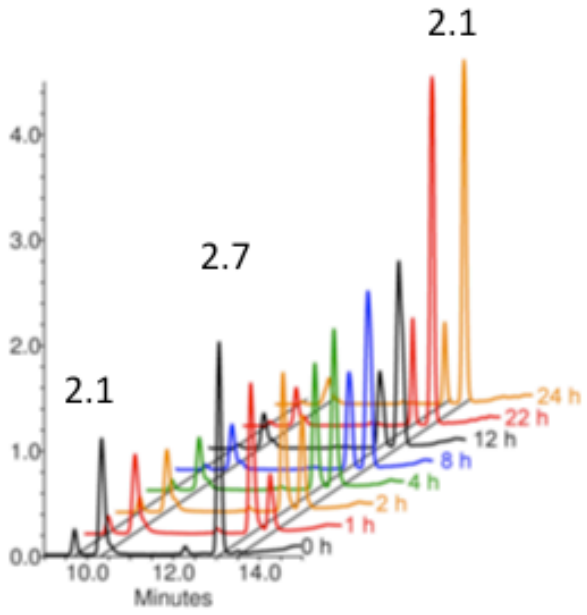


Figure E2.2 Stacked HPLC plot of the disappearance of diketide **2.5** and extender unit **2.1** and the formation of ketolactone **2.11** (monitored at 242 nm).

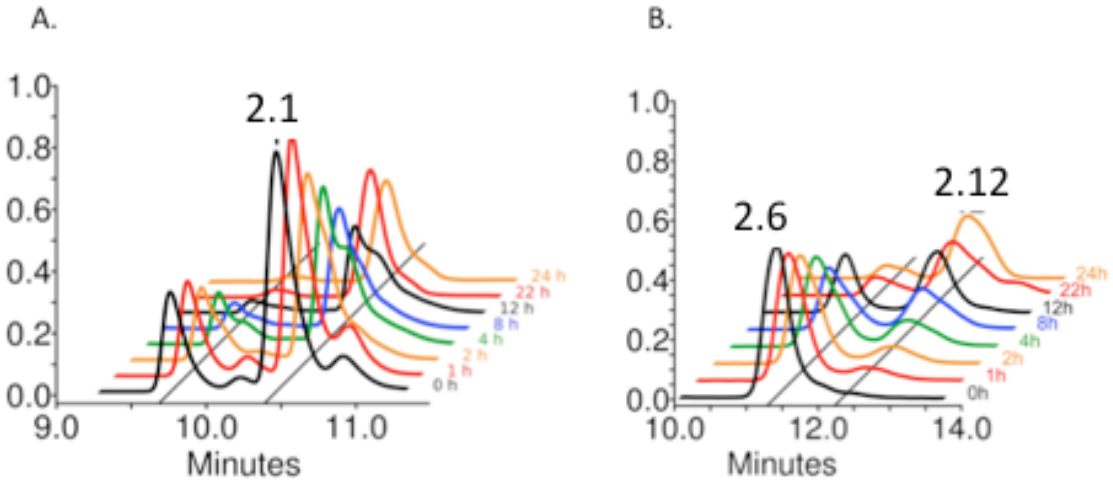


Figure E2.3 Stacked HPLC plots of A) the disappearance of extender unit **2.1** and B) the disappearance of diketide **2.6** and the formation of ketolactone **2.12** (monitored at 242 nm).

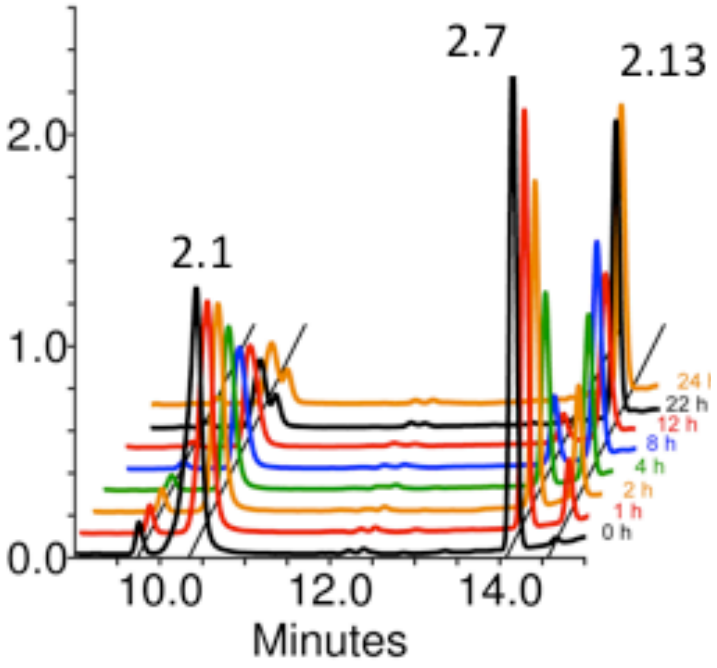
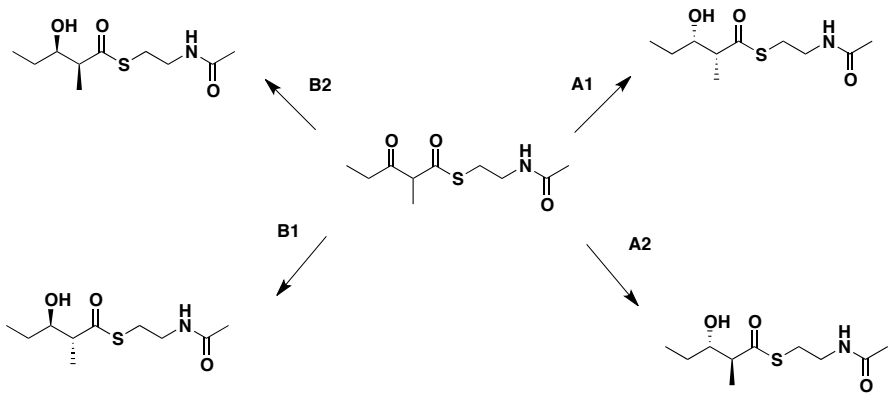


Figure E2.4 Stacked HPLC plot of the disappearance of diketide **2.7** and extender unit **2.1** and the formation of ketolactone **2.13** (monitored at 242 nm).

THE 2*S*, 3*R* REQUIREMENT FOR DIKETIDE INCORPORATION BY MOD6TE



Scheme E2.1 Schematic for the generation of all stereoisomers of diketide **5** by reducing with different types of isolated ketoreductases (A1=AmpKR2,[2] A2=RifKR7,[3] B1=TylKR1,[2] B2=EryKR2[2]).

Each of the stereoisomers of reduced diketide **2.5** were incubated with EryMod6TE; however, only reduced diketide **2.5** generated triketide lactone product. KR-containing lysate was employed to generate the diketide stereoisomers from diketide **2.2** [AmpKR2(A1-type)[2] for the (2*R*, 3*S*)-diketide, RifKR7(A2-type)[2,3,5] for the (2*S*, 3*S*)-diketide, TylKR1(B1-type)[2] for the (2*R*, 3*R*)-diketide, and EryKR1(B2-type)[2,5] for the (2*S*, 3*R*)-diketide] **(Scheme E2.1**). The reactions to form **2.5** and its stereoisomers were performed, extracted, and analyzed using the procedure detailed for the HPLC assay of ketolactone formation. Generation of **2.1** was also performed using the conditions described above. Triplicate ketolactone reactions consisted of 250 μL MatB reaction, 1 mL EryMod6TE lysate, an amount of extracted β-hydroxyacyl-S-NAC equivalent to the methylmalonyl-S-NAC (as estimated by HPLC peak area), in a total volume of 2.1 mL. Timepoints taken at 0, 4, and 22 hr were quenched and assayed by HPLC as described above (**Figure E2.5**).

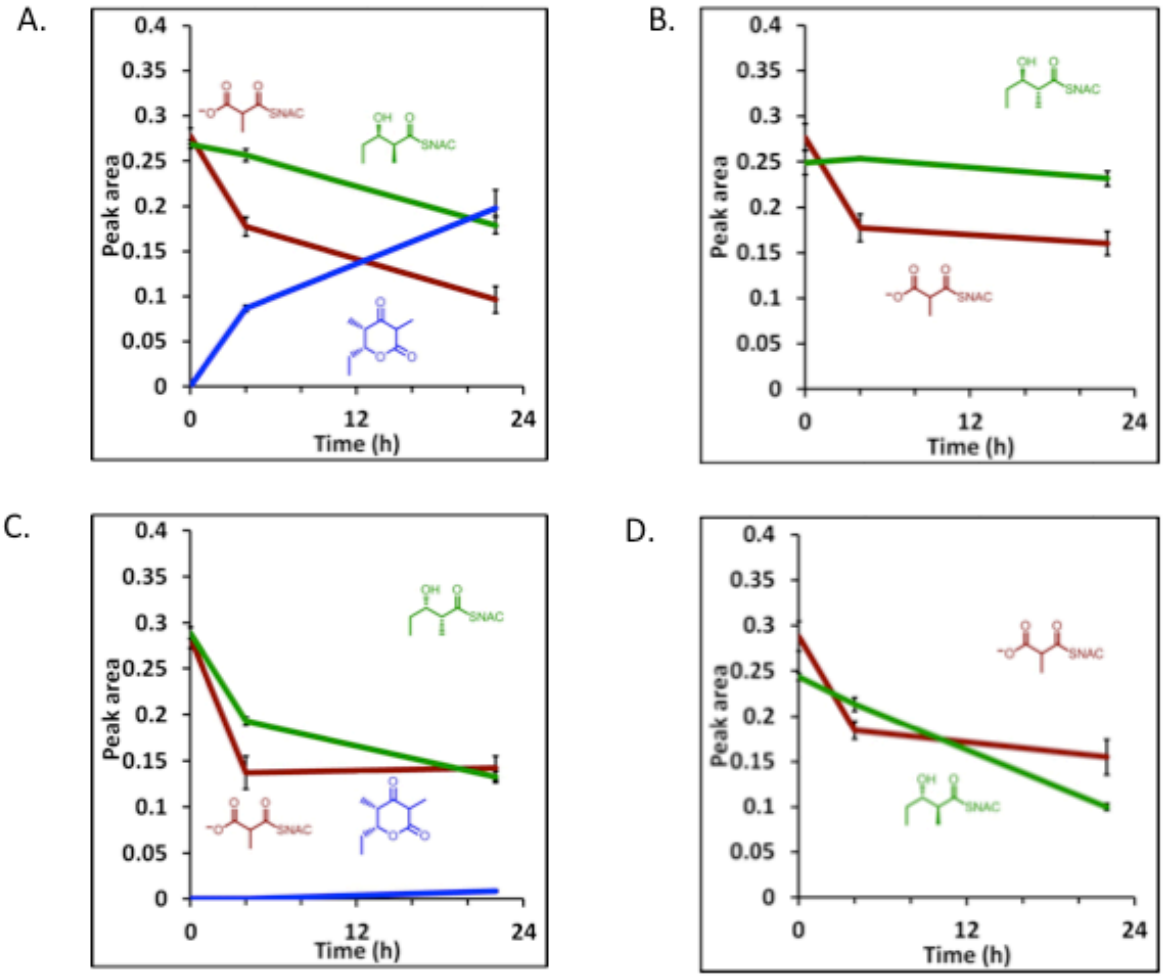


Figure E2.5 Time course plots of the incubation of methylmalonyl-S-NAC (**2.1**) and the stereoisomers of reduced diketide **5** (see Scheme E5.1) with EryMod6TE. Timepoints taken were at 0, 4, and 22 hours. All species were monitored by HPLC peak area (in arbitrary units). A) (2*S*,3*R*)-3-hydroxy-2-methylpentanoyl-S-NAC (**2.5**), generated by incubating **2.5** with EryKR1 (B2-type), yielded ketolactone **8**. B) (2*R*,3*R*)-3-hydroxy-2-methylpentanoyl-S-NAC, generated by incubating **2.5** with TylKR1 (B1-type), did not yield ketolactone. C) (2*S*,3*R*)-3-hydroxy-2-methylpentanoyl-S-NAC, generated by incubating **2.5** with AmpKR2 (A1-type), yielded trace quantities of ketolactone (consistent with the formation of the trace quantities of **2.5** as a side-product of the reduction of **2.5** by AmpKR2[2]). D) (2*R*,3*R*)-3-hydroxy-2-methylpentanoyl-S-NAC, generated by incubating **2.5** with RifKR7 (A2-type), did not yield ketolactone.

X-RAY EXPERIMENTAL DATA FOR $C_9H_{14}O_3$ (2.11)

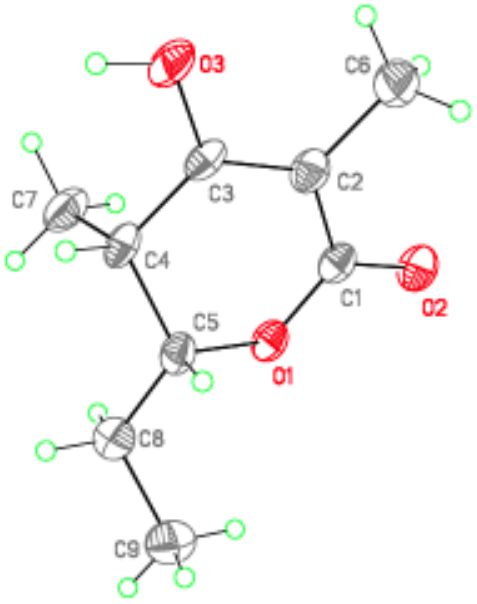


Figure E2.6 ORTEP diagram of **2.11**

Crystals grew as clusters of long, colorless needles by slow evaporation from 50% EtOAc:hexanes. The data crystal was cut from a larger crystal and had approximate dimensions; 0.48 x 0.12 x 0.10 mm. The data were collected on a Rigaku SCX-Mini diffractometer with a Mercury CCD using a graphite monochromator with Mo K_α radiation (λ = 0.71075 Å). A total of 540 frames of data were collected using ω-scans with a scan range of 1° and a counting time of 50 seconds per frame. The data were collected at 153 K using a Rigaku XStream temperature device. Details of crystal data, data collection and structure refinement are listed in Table 1. Data reduction were performed using the Rigaku Americas Corporation's Crystal Clear version 1.40.[7] The structure was solved by direct methods using SIR97[8] and refined by full-matrix least-squares on F^2 with anisotropic displacement parameters for the non-H atoms using

SHELXL-97.9[8] Structure analysis was aided by use of the programs PLATON98[10] and WinGX.[11] The hydrogen atoms on carbon were calculated in ideal positions with isotropic displacement parameters set to 1.2xUeq of the attached atom (1.5xUeq for methyl hydrogen atoms). The function, $\Sigma w(|F_o|^2 - |F_c|^2)^2$, was minimized, where $w = 1/[(\sigma(F_o))^2 + (0.0425*P)^2]$ and $P = (|F_o|^2 + 2|F_c|^2)/3$. $R_w(F^2)$ refined to 0.114, with R(F) equal to 0.0599 and a goodness of fit, S, = 1.17. Definitions used for calculating R(F), $R_w(F^2)$ and the goodness of fit, S, are given below.[vii] The data were checked for secondary extinction effects but no correction was necessary. Neutral atom scattering factors and values used to calculate the linear absorption coefficient are from the International Tables for X-ray Crystallography (1992).[12] All figures were generated using SHELXTL/PC.[13] Tables of positional and thermal parameters, bond lengths and angles, torsion angles and figures are found elsewhere. CCDC number: 878292.

REFERENCES

[1] A.J. Hughes, A. and Keatinge-Clay, A. *Chem. Biol.* **2011**, 18, 165.

[2] S.K. Piasecki, C.A. Taylor, J.F. Detelich, J. Liu, J. Zheng, A. Komsoukaniants, D.R. Siegel, A.T. Keatinge-Clay. *Chem Biol.* **2011**, 18, 1331.

[3] C.R. Valenzeno, Y.You, A. Garg, A. Keatinge-Clay, C. Khosla, D.E. Cane. *J. Am. Chem Soc*. **2010**, *132*, 14697-9.

[4] S. Murli, J. Kennedy, L.C. Dayem, J.R. Carney, J.T. Kealey, *J. Ind. Microbiol. Biotechnol*. **2003**, *30*, 500.

[5] Keatinge-Clay, A.T. *Chem Biol*. **2007**, 14, 898.

[6] Kao, C.M., Luo, G., Katz, L., Cane, D.E., and Khosla, C. *J. Am. Chem Soc*. **1994,** 116, 111612.

[vii] $R_W(F^2) = \{\Sigma w(|F_o|^2 - |F_c|^2)^2/\Sigma w(|F_o|)^4\}^{1/2}$ where w is the weight given each reflection.

$R(F) = \Sigma(|F_o| - |F_c|)/\Sigma|F_o|\}$ for reflections with $F_o > 4(\sigma(F_o))$.

$S = [\Sigma w(|F_o|^2 - |F_c|^2)^2/(n - p)]^{1/2}$, where n is the number of reflections and p is the number of refined parameters.

[7] DENZO-SMN. (1997). Otwinowski, Z. and Minor, W., Methods in Enzymology, **276**: Macromolecular Crystallography, part A, 307 – 326, C. W. Carter, Jr. and R. M. Sweets, Editors, Academic Press.

[8] SIR97. (1999). A program for crystal structure solution. A. Altomare, M.C. Burla, M. Camalli, G.L. Cascarano, C. Giacovazzo, A. Guagliardi, A.G.G. Moliterni, G. Polidori, and R.J. *Spagna Appl. Cryst.* 32, 115-119.

[9] G.M. Sheldric. (2008). SHELXL97. Program for the Refinement of Crystal Structures. *Acta Cryst,* A64, 112.

[10] A.L. Spek, (1998). PLATON, A Multipurpose Crystallographic Tool. Utrecht University, The Netherlands.

[11] WinGX 1.64. (1999). An Integrated System of Windows Programs for the Solution, Refinement and Analysis of Single Crystal X-ray Diffraction Data. Farrugia, L. J. *J. Appl. Cryst.* 32. 837.

[12] International Tables for X-ray Crystallography (1992). Vol. C, Tables 4.2.6.8 and 6.1.1.4, A. J. C. Wilson, editor, Boston: Kluwer Academic Press.

[13] G.M. Sheldrick. (1994). SHELXTL/PC (Version 5.03). Siemens Analytical X-ray Instruments, Inc., Madison, Wisconsin, USA.

Chapter 3: Substrate Structure-Activity Relationships Guide Rational Stereochemical Engineering of Modular Polyketide Synthase Ketoreductases[viii]

INTRODUCTION

Polyketides comprise a diverse class of natural products with intriguing carbon skeletons. Due to the stereochemical complexity of many polyketides, members of this class of natural products have been vibrant targets for developing and applying asymmetric methodologies.[1–4] In contrast to polyketides' molecular complexity, polyketides are biosynthesized by multidomain megasynthase enzymes (termed polyketide synthases, PKSs) using remarkably simple biogenic precursors. PKSs perform extensions of malonyl-CoA or methylmalonyl-CoA through decarboxylative condensations followed by reductive processing by optional ß-carbon processing domains.[5] Of the ß-carbon processing domains, ketoreductase (KR) domains are responsible for the majority of the stereogenic centers in polyketides: KRs both stereoselectively reduce the ß-keto intermediates to a ß-hydroxy moiety and confer the stereochemical configuration of the α-substituent.[6] Additionally, KR domains possess sequence fingerprints, which enables prediction of the product stereochemistry. KRs that yield products with D-ß-hydroxyls are referred to as "A-type" and KRs that yield products with L-ß-hydroxyls are referred to as "B-type."[5] KRs that reduce D-α-substituted products are denoted as "1" and KRs that reduce L-α-substituted products are denoted as "2."[6–9]

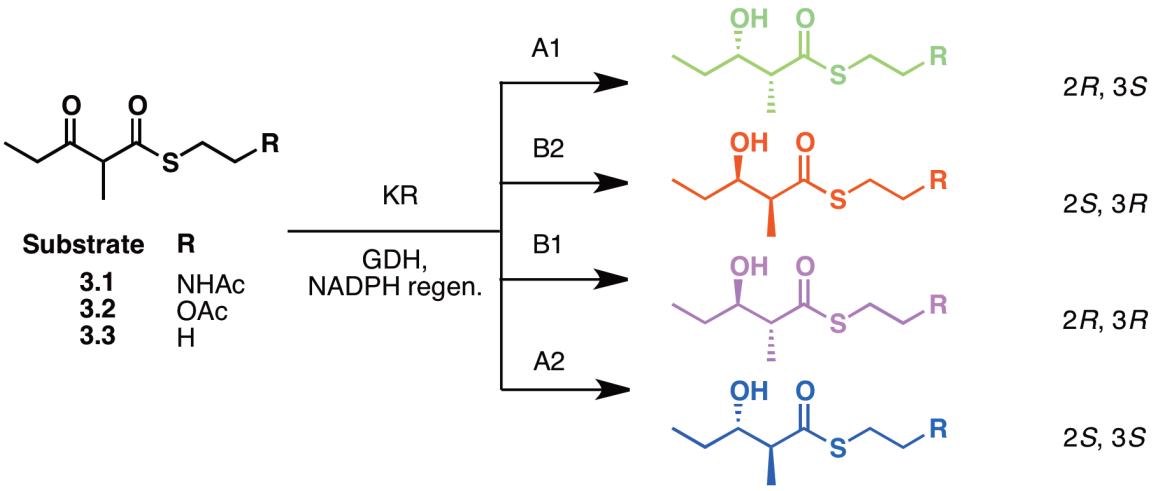
[viii] Portions of this chapter are reproduced from the following publication: C. B. Bailey, M. E. Pasman and A. T. Keatinge-Clay, *ChemComm*, 2015, DOI: 10.1039/C5CC07315D. CBB performed the majority of the experimental work, and MEP helped with synthesis of standards. CBB and ATK evaluated the data and wrote the original text.

KRs' natural substrates are ß-keto polyketide intermediates tethered to an 18Å phosphopantetheinyl prosthetic group (the phosphopantetheinyl arm") of the acyl carrier protein. Frequently, KRs retain their stereochemical fidelity when reducing diketide *N*-acetyl cysteamine (*S*-NAC) thioesters (e.g. **3.1**; Scheme **3.1**), a common truncated mimic of the phosphopantetheinyl arm.[10] The relationship between the molecular similarity of the "handle" region of the substrate to the phosophopantheinyl arm and the stereochemical outcome of KR reductions is largely unexplored, as the majority of small molecule investigations of stereocontrol have been performed with *S*-NAC substrates.[10–15] To date only one example is present in the literature wherein stereocontrol was examined with a small molecule substrate with a pantetheinyl handle. This report demonstrated that when the 1st KR of the erythromycin PKS (EryKR1) reduced 2-methyl-3-oxopentanoyl-pantetheine, the kinetic parameters and stereochemical product distributions were virtually identical to EryKR1 reducing *S*-NAC substrate **3.1**.[13] However, studies wherein the 1st KR from the tylosin PKS (TylKR1) was incubated with various oxo-ester substrates suggests that reversals in enatioselectivity may be correlated to unnatural handle structure.[16] To further dissect the interactions that mediate KR stereocontrol, we decided to systematically vary the stereoelectronic features of the KR handle. To this end, we synthesized congeneric thioester handles and evaluated the effects of structural alterations on KR stereoselectivity in a structure-activity relationship (SAR) fashion. Here, we report the effects of altering thioester substrates in two key ways from canonical *S*-NAC substrates: atom substitution and truncation.

USE OF A SUBSTRATE STRUCTURE-ACTIVITY RELATIONSHIP TO INTERROGATE STEREOCONTROL

Four KRs were selected: EryKR1, TylKR1, the 2nd KR from the amphotericin PKS (AmpKR2), and the 7th KR from the rifamycin PKS (RifKR7) as representative B2,

B1, A1, and A2 type KRs, respectively. These KRs were chosen as they retain their stereochemical fidelity with **3.1**, and are active in the presence of DMSO,[10,17] affording the ability to investigate hydrophobic substrate **3.3** (which required 20% v/v DMSO to solubilize enough substrate to observe activity). Biocatalytic screens combining KR, **3.1-3.3**, and an NADPH regeneration scheme (**Scheme 3.1**)[10] were incubated overnight, followed by analysis via chiral chromatography. To establish the elution order of the four stereoisomers for **3.3** and **3.2**, we generated synthetic standards using a combination of stereospecific aldol reactions previously detailed by our laboratory[10,18] (experimental section). The standards confirmed that each KR generated the predicted stereoisomer as the major product (experimental section).



Scheme 3.1 KR-mediated reduction of 2-methyl-3-oxopentanoyl substrates with linkages to NAC (**3.1**), thioethylacetate (**3.2**), and ethanethiol (**3.3**) handles. GDH, glucose dehydrogenase.

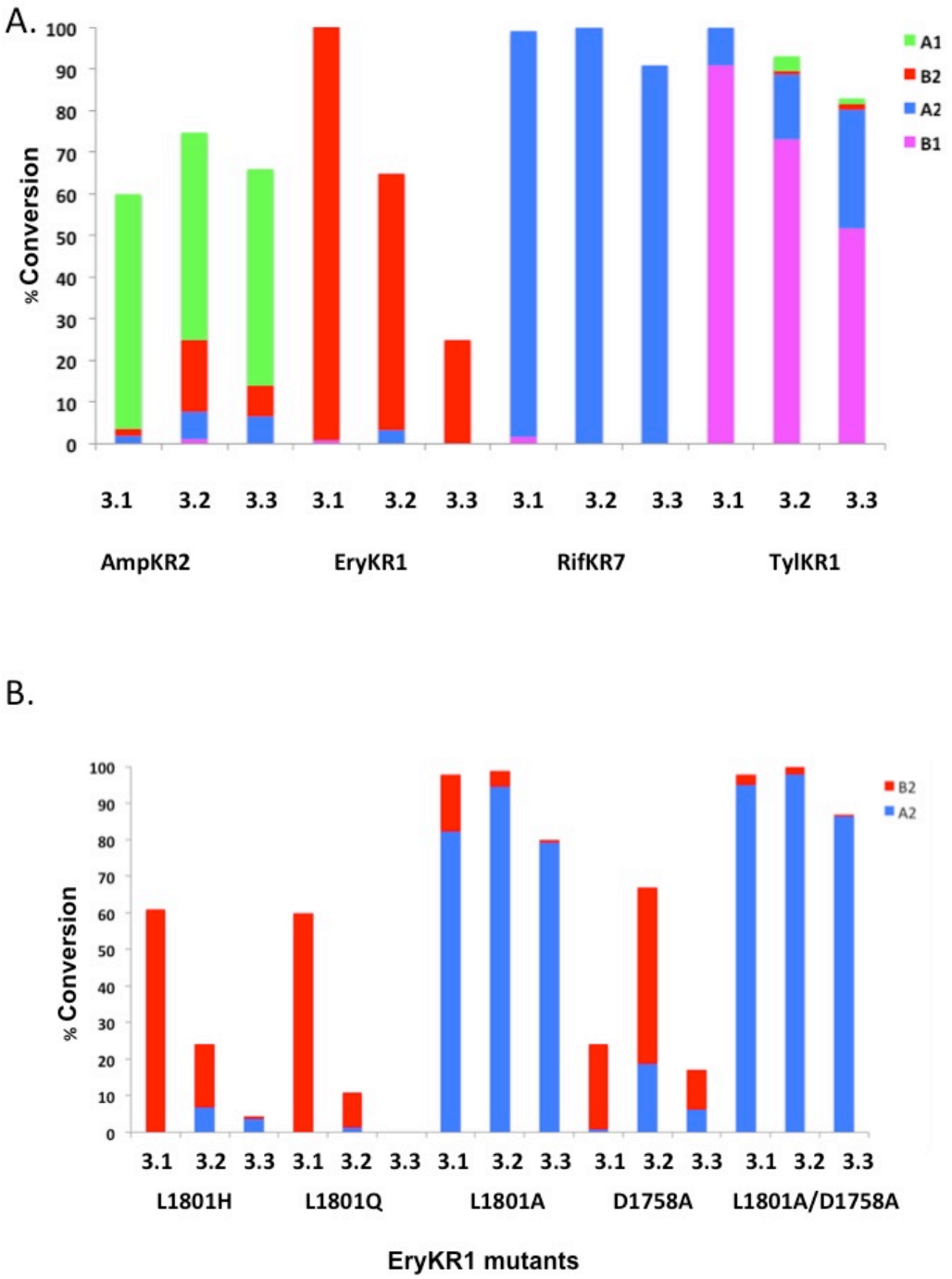

A.
% Conversion
100
90
80
70
60
50
40
30
20
10
0
A1
B2
A2
B1
3.1 3.2 3.3 3.1 3.2 3.3 3.1 3.2 3.3 3.1 3.2 3.3
AmpKR2
EryKR1
RifKR7
TylKR1
B.
% Conversion
B2
A2
3.1 3.2 3.3 3.1 3.2 3.3 3.1 3.2 3.3 3.1 3.2 3.3 3.1 3.2 3.3
L1801H
L1801Q
L1801A
D1758A
L1801A/D1758A
EryKR1 mutants

Figure 3.1

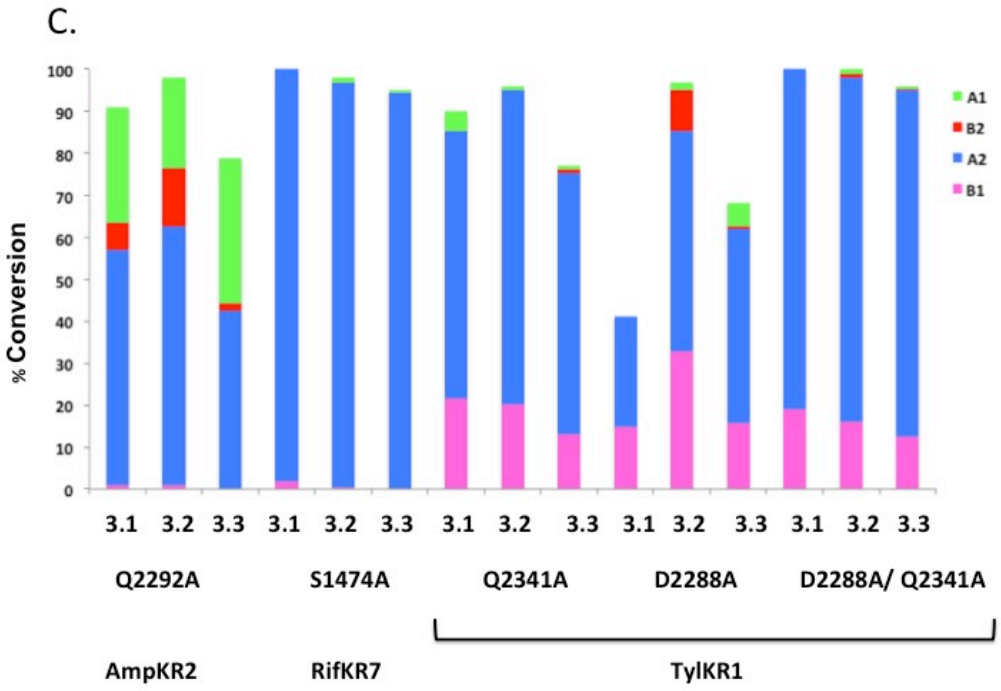


Figure 3.1. Percent conversion and stereoisomeric product ratios of (A) unmutated KRs and (B) EryKR1 point mutants and (C) AmpKR2, RifKR7, and TylK1 point mutants. See experimental section for chromatograms and tables listing product ratios and percent conversions. The colors used correspond to those in **Scheme 3.1.**

Given that stereocontrol is often observed with *S*-NAC substrates such as **3.1,**[10] we sought to probe the role of the amide moiety. As it has been shown that with EryKR1, the enzyme behaves identically toward **3.1** and the pantetheinyl analog (*vide supra*),[13] this suggests that longer mimics of the phosphopantetheinyl arm may not provide additional critical binding interactions. However, we hypothesized that the minimal polar feature retained between the pantetheine handle and the *S*-NAC handle (the amide moiety) might be a significant contributing element to stereocontrol, perhaps through hydrogen bonding interactions that guide substrate orientation. To test this hypothesis, we substituted the nitrogen atom of the amide to oxygen (**3.2**). Although this

atomic substitution is conservative, the ester linkage is distinct in a few ways: 1) it lacks a hydrogen bond donor, 2) it is a more flexible substrate, as it lacks hindered rotation arising from the amide C-N bond's increased pi character and 3) it is far less polar. Surprisingly, the overall effects of this substitution on stereocontrol were relatively subtle. AmpKR2 generated the same minor products as it had with **3.1** (the B2 and A2 products), however the stereoselectivity was significantly eroded: with **3.2**, 31% of the overall products were non-A1 products, whereas for **3.1**, only 6% of the products were unnatural stereoisomers (**Figure 3.1A**). TylKR1 produced the A2 stereoisomer (the enantiomer of its natural stereochemistry) as a minor product, which was also observed with **3.1**.[10] EryKR1 was highly selective for the natural B2 stereochemistry with substrate **3.2,** and RifKR7 generated its natural A2 product with perfect enatioselectivity (**Figure 3.1A**). We next decided to probe the role of this carbonyl *via* complete truncation to **3.3**. Even more surprisingly, the results were similar to those with **3.2**, except that stereoselectivity was further eroded for AmpKR2 and TylKR1 (**Figure 3.1A**). Additionally, EryKR1 appeared highly affected by the high DMSO concentration required to solubilize **3.3**, whereas the other KRs were comparatively robust.

A SAR-GUIDED POINT MUTANT

Stereoselectivity with **3.3** demonstrates that no polar binding interactions from the handle are required to drive the formation of energetically favored transition states. Although this result was unexpected, it corroborates with a hypothesis presented by Leadlay and coworkers that KR stereocontrol is predominantly enforced by subtle stereoelectronic effects from active site residues.[11,14] To further probe this hypothesis, we generated point mutants of a representative KR, EryKR1. We chose a residue that is conserved in B2 type enzymes, leucine 1810, which is in close vicinity to both the

catalytic tyrosine and the cofactor[6,19] (**Figure 3.2**). L1810 is a fingerprint for predicting the stereochemistry of the α-methyl group: in B2 type KRs it is a usually leucine, in A2 type KRs it is a histidine, and in B1 and A1 KRs, it is a glutamine[6] (experimental section). Furthermore, there is precedence for this residue being an important feature: the analogous residue in AmpKR2 was one of two residues mutated to completely reverse the stereocontrol of AmpKR2.[15] Thus, we generated the following point mutations to EryKR1: L1810H, and L1810Q, and L1810A.

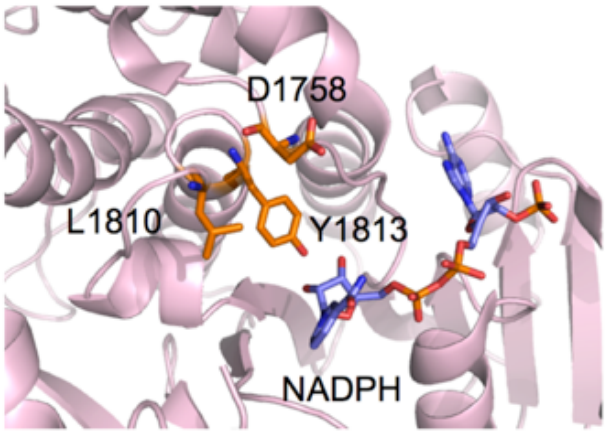


Figure 3.2 The active site of EryKR1 reveals the proximity of D1758 and L1810 to the catalytic tyrosine, Y1813, and NADPH (PDB code: 2FR1).

The L1810H and L1810Q mutants were much less robust enzymes than wild type EryKR1: for all three substrates both mutants resulted in decreased activity and selectivity (**Figure 3.1B**). With **3.1**, both mutants generated the natural B2 product whereas with **3.2**, stereocontrol was significantly eroded (with 28% and 12% of and the A2 isomer for the L1810H and L1810Q, respectively). For **3.3,** trace and no activity was observed with L1810H and L1810Q, respectively (**Figure 3.1B**). In contrast, the EryKR1 L1810A mutant was both a highly active and highly stereoselective enzyme. Remarkably, however, the stereoselectivity, was entirely reversed from that of wild type, with selective

formation of the A2 product. By percent conversion, the EryKR1 L1810A mutant was more active than wild type. There was a subtle difference in product stereochemical distribution: as the handle became an increasingly poor mimic of the phosphopantetheinyl arm: for substrates **3.3**, **3.2**, and **3.1**, the *de* values were 98%, 91%, and 69%, respectively (experimental section). Intriguingly, the enzyme generated the A2 product with *greater* selectivity as the handle became increasingly unnatural.

A SECOND RATIONAL MUTATION IS SYNERGISTIC

A hypothesis for the function of the diagnostic fingerprint of B-type KRs, the LDD motif, is that these residues interact with the phosphopantetheinyl arm of the polyketide intermediate.[6,7,15,19–23] Structural characterizations have revealed that all KRs bind the NADPH cofactor in the same orientation (transferring the 4-*pro*-*S*-hydride). Thus, stereocontrol must arise from the substrate orientation. As such, in B-type KRs, the substrate must expose the *si* face, and in A-type KR the substrate must expose the *re* face to the cofactor.[7,20] The LDD motif is on a flexible loop that could interact with the substrate amide, thus positioning the diketide for reduction at the *si* face.[6,7,19,22] Indeed, in the crystal structure of a related enzyme, PhaB from *Ralstonia eutropha* H16 (a (*R*)-3-hydroxybutyryl-CoA dehydrogenase), such an interaction is observed. A charged hydrogen bond is formed between an aspartate homologous to the second D of the LDD motif and the amide nearest the thioester in the bound acetoacetyl-CoA[24] (experimental section). Such an interaction with the LDD loop is consistent with the subtle differences in stereocontrol enforced by the EryKR1 L1810A mutant. For **3.1**, hydrogen bonding with the aspartate residue can occur, which may account for the minor production of the B2 isomer. The enatioselectivity drastically increases as the atom is substituted to an ester in **3.2**, thus unable to hydrogen bond with the aspartate in this fashion. With **3.3**, near

perfect A2-enatioselectivity is observed, which is consistent with the handle structure completely precluding interaction with the aspartate. To further probe this hypothesis, we mutated the third D of the LDD motif (D1758) to generate both a D1758A single mutant and a L1810A/D1758A double mutant.

The single D1758A mutation eroded B2 selectivity, however the B2 stereoisomer remained the major product for all three handles (Figure **3.1B**). This indicates that L1810 has a greater role in enforcing stereocontrol than hydrogen bonding with D1758. This result is consistent with the retention of stereocontrol in wild type EryKR1 with substrates **3.2** and **3.3,** which are unable to hydrogen bond with D1758 (Figure **3.1B**). Gratifyingly, for EryKR1 L1810A/D1758A, the two point mutations were synergistic: there was an enhancement of A2 selectivity over the L1810A single mutant for all substrates (99, 97, and 95% *de* for substrates **3.3**, **3.2**, and **3.1**, respectively; **Figure 3.1B**, experimental section). Although the D1758A single mutant had diminished activity, the L1810A/D1758A had greater activity than wild type, just as had been observed with the L1810A single mutant (**Figure 3.1B**). Thus, through two point mutants an enzyme was designed that was both more active and more stereocontrolled than wild type.

REVERSING STEREOCONTROL TO FORM THE A2 STEREOISOMER IS GENERALLY ACCESSIBLE *VIA* ALANINE POINT MUTATIONS

After determining that generating targeted point mutations was an effective strategy to reverse stereocontrol in EryKR1, we decided to probe how general this approach was. To this end, we generated analogous point mutations in RifKR7, TylKR1, and AmpKR2. In RifKR7, the residue appearing three residues N-terminal to the catalytic tyrosine is a serine (RifKR7 is unusual; as noted previously, typically this position is a histidine in A2 KRs), so thus we generated a RifKR7 S1474A mutant. The activity of RifKR7 S1474A was essentially identical to that of wild type, with only extremely subtle

differences in activity (**Figure 3.1C**). For substrates **3.2** and **3.3**, the A1 stereosiomer was observed as a trace side product (1 and 0.4% of the overall product, respectively), which had not been observed in wild type RifKR7 (in wild type RifKR7, only the B1 product had been observed as a trace side product with substrate **3.1**). Additionally, the activity of RifKR7 toward substrate **3.3** was improved (80% conversion in wild type RifKR7 versus 95% conversion with RifKR7 S1474A; **Figure 3.1C**, experimental section), perhaps suggesting that the alanine mutation confers enhanced DMSO tolerance relative to wild type enzyme. For AmpKR2 2292A, an increase in the amount of A2 product formed for all three substrates, however the reversals in stereochemistry were not as complete as they were for the EryKR1 L1810A mutant (**Figure 3.1C**). Although the major product in all cases was the A2 stereoisomer, for AmpKR2 2292A, significant formation of the natural A1 stereoisomer was observed (30, 22, and 40% of the overall products, for **3.1**, **3.2**, and **3.3**, respectively). Interestingly, just has been observed with wild type, substrate **3.2** had significant formation of the B2 product compared to the other two thioesters (8, 14, and 2% of the overall products for **3.1**, **3.2**, and **3.3**, respectively; **Figure 3.1C,** experimental section).

For TylKR1, the residue appearing three residues N-terminal to the catalytic tyrosine is Q2341. Additionally, as TylKR1 is a B-type enzyme, it harbors an LDD motif. Thus, we generated three point mutations, just as we had with EryKR1: TylKR1 Q2341A, TylKR1 D2288A, and TylKR1 D2288A/Q2341A (**Figure 3.1C**). For TylKR1 Q2341A, in a similar fashion to EryKR1 L1810, increased formation of the A2 product occurred as the substrate became increasingly unnatural (42, 56, and 63% B1:A2 *ee* for substrates **3.1**, **3.2**, and **3.3**, respectively; **Figure 3.1C,** experimental section). For TylKR1 D2288A, the mutation also increased the amount of A2 product formation, however not as effectively as TylKR1 Q2341A. Additionally, the TylKR1 D2288A

resulted in a less active enzyme when reducing substrate substrate **3.1** (41% conversion versus 97 and 69% conversion when reducing substrates **3.2** and **3.3**, respectively; for wild type TylKR1, the conversions were 100, 93, and 82% for substrates **3.1**, **3.2**, and **3.3**, respectively; **Figure 3.1C,** experimental section). Although TylKR1 D2288A was completely *anti* diasteroselective when reducing substrate **3.1**, there was an erosion of *anti:syn* diastereoselectivity for substrates **3.2** and **3.3** (76 and 80% *anti:syn de*, respectively; **Figure 3.1C**, experimental section). This suggests that in TylKR1 (like in EryKR1) the steric environment in the active site (i.e. Q2341) is a greater contributor to stereocontrol than hydrogen bonding interactions from the LDD motif (D2288). However, like the EryKR1 L1810/D1758 mutant, the two point mutations were synergistic: the TylKR1 D2288A/Q2341A double mutant had enhanced activity for all three substrates (100% conversion for substrates **3.1** and **3.2** and 96% conversion for substrate **3.3, Figure 3.1C,** experimental section). The reversal to the A2 product with TylKR1 D2288A/Q3241A was similar, although not as complete as the analogous reversal of stereocontrol in EryKR1 (66, 63, and 71% *ee* for substrates **3.1**, **3.2**, and **3.3**, respectively, **Figure 3.1C,** experimental section). Overall, the TylKR1 point mutants all behaved similarly to the analogous ones made in EryKR1.

THE A2 STEREOISOMER IS THE "DEFAULT" STEREOCHEMICAL OUTCOME: BIOINFORMATIC AND PHYSICAL EXPLANATIONS

Due to the chiral environment of the enzyme as well as the invariant orientation of the prochiral cofactor, there must be intrinsic facial selectivity. Empirically, it appears that A-type facial selectivity is favored. A few bioinformatic observations are consistent with this hypothesis. Although the first two L and D residues of the LDD motif found in *cis* AT KRs are not conserved in *trans* AT KRs, the second D of the LDD is invariant in all B-type KRs.[25–29] For A-type KRs, the analogous fingerprint is a conserved W.[8,9]

However, the W is entirely absent in *trans* AT A-type KRs (indeed, there are no identified fingerprints that correlate to *trans*-AT A type KRs).[25–29] Examples of *cis* AT A type KRs that lack the diagnostic W also exist.[30] The sequence variance in A-type KRs suggests that there may be many evolutionary solutions to A-type facial selectivity. In contrast, the D appears to be a significant contributing element to driving B-type reduction. Perhaps this sequence conservation is due to evolutionary pressure to retain features that guide the substrate to undergo the less intrinsically favorable facial selectivity (an energy diagram illustrating this putative facial bias is shown below; **Figure 3.3**).

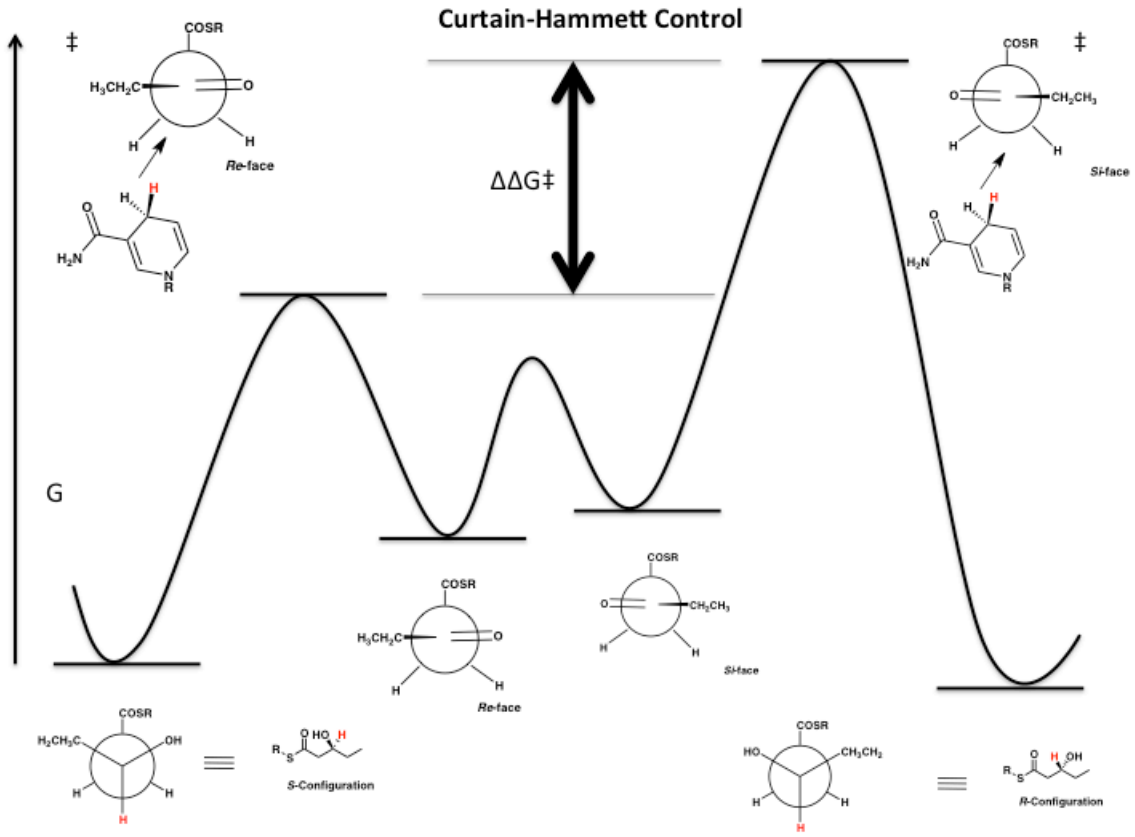
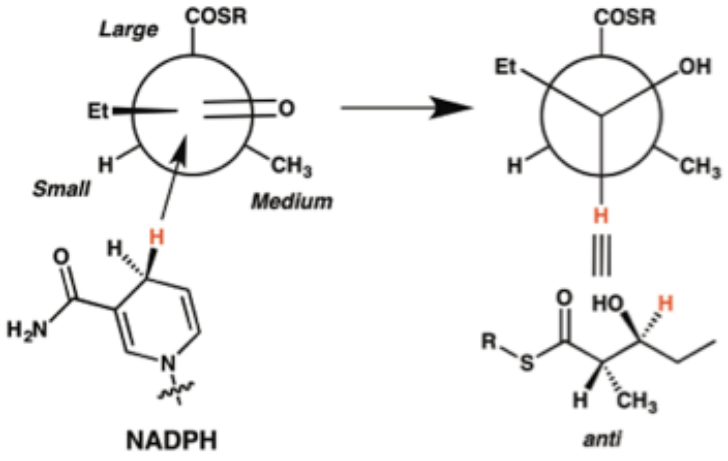


Figure 3.3 Curtain-Hammet analysis of KR reductions (α-unsubtituted shown), demonstrating an energetic a facial bias for A-type ketoreduction. An energetic difference of ~3 kcal/mol would result in ~99% *ee*.

The intrinsic preference for *anti* diasteroselectivity can be rationalized by the Felkin-Ahn model.[31] Under Felkin-Ahn selectivity, the nucleophile will approach in the least hindered fashion with the electrophile in the conformation that results to the least torsional strain (**Scheme 3.2**). Thus, interactions in the active site (*e.g.* steric clashes or hydrogen bonding) would be required to drive *syn* selectivity. Indeed, the effects of mutations on TylKR1 support this hypothesis: although they exhibited a similar reversal to the A2 product, the complete reversal of stereochemical outcome was not achieved to the same extent as with mutants of EryKR1, a naturally *syn* selective enzyme. A possible

explanation is that formation of the natural stereoisomer for TylKR1 (the B1 *anti* product) is a more intrinsically favored energetic pathway than formation of the natural stereoisomer for EryKR1 (the B2 *syn* product). Notably, *anti* selective (B1 and A2) KRs are found in modules harboring a DH.[6,9,22,30] Perhaps it makes evolutionary sense that in modules wherein stereochemical information is lost (*i.e.* the hydroxyl is dehydrated to an olefin), KRs have evolved to undergo more intrinsically energetically favorable reductions. Additionally, the hypothesis of an intrinsic preference for A-type facial selectivity and *anti* diastereoselectivity (resulting in an energetic "default" of the A2 stereoisomer) is further supported by: 1) the activity of the RifKR7 S1474A mutant was essentially identical to that of wild type RifKR7 and 2) EryKR1, AmpKR2, and TylKR1 all generated enhanced amounts of the A2 product with analogous active site mutations.

With this rationalization in mind, the next challenge will be using a similar approach to engineer KR selective for a less intrinsically favored stereochemical outcome (*e.g.* a *syn* selective KR). This should be achievable, as kinetic data with **3.1** indicates that the energetic differences that lead to different stereoisomeric products in KRs are extremely small.[32] Indeed, with unnatural substrates such as **3.1-3.3**, KRs are not acting as evolutionarily optimized catalysts; they are essentially relatively loosely bound chiral catalysts that aid asymmetric induction, and essentially follow the Curtain-Hammet principle (**Figure 3.3**). Energetic differences arising from asymmetric induction are notoriously small (95% *ee* typically results from an energetic difference of ~3 kcal/mol[32]), which explains why minor perturbations to the active site yields drastic differences in stereochemical outcome.



Scheme 3.2 Felkin-Anh analysis the reduction of a diketide by NADPH, resulting in *anti* selectivity.

FUTURE DIRECTIONS

An exciting future direction is a collaboration that has been instigated with the group of Prof. Matthew Sigman at the University of Utah. Sigman and coworkers have used a "big data" Q_{sar} approach to rationally predict and design asymmetric catalysts in the realm of synthetic organic chemistry.[33,34] Because KRs enforce stereocontrol with small molecule mimics through weak, non-colvalent interactions (*e.g.* hydrophobic interactions or hydrogen bonding), they may be an ideal model system for translating this approach to engineering enzymes.

CONCLUSIONS AND OUTLOOK

In summary, here we employ a substrate SAR approach to uncover the origins of KR selectivity, which, in turn can be used to guide mutagenesis. Through this method, we entirely reversed the stereoselectivity of EryKR1 through a mere two judicious point mutations. This approach was found to be general when similar mutations were introduced in RifKR7, TylKR1, and AmpKR2. These results demonstrate the utility of

applying simple physical organic models to both rationalize and engineer enzymatic stereocontrol.

ACKNOWLEDGEMENTS

We would like to thank the NIH (GM106112) and the Welch Foundation (F-1712) for supporting this research (A.T.K.). We also thank Dr. Shawn Piasecki and Dr. Johnathan Brantley for useful discussions.

REFERENCES

[1] A.-M. R. Dechert-Schmitt, D. C. Schmitt, X. Gao, T. Itoh, M. J. Krische, *Nat. Prod. Rep.* **2014**, *31*, 504.

[2] J. Li, D. Menche, *Synthesis*. **2009**, *2009*, 2293.

[3] D. Leonori, V. K. Aggarwal, *Acc. Chem. Res.* **2014**, *47*, 3174.

[4] B. Schetter, R. Mahrwald, *Angew. Chem.* **2006**, *45*, 7506.

[5] C. Hertweck, *Angew. Chem.* **2009**, *48*, 4688.

[6] A. T. Keatinge-Clay, *Chem. Biol.* **2007**, *14*, 898.

[7] J. Zheng, A. T. Keatinge-Clay, *MedChemComm*. **2012**. *4*, 34.

[8] P. Caffrey, *ChemBioChem*. **2003**, *4*, 654.

[9] R. Reid, M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D. V Santi, C. R. Hutchinson, R. McDaniel, *Biochemistry*. **2003**, *42*, 72.

[10] S. K. Piasecki, C. A. Taylor, J. F. Detelich, J. Liu, J. Zheng, A. Komsoukaniants, D. R. Siegel, A. T. Keatinge-Clay, *Chem. Biol.* **2011**, *18*, 1331.

[11] H. M. O'Hare, A. Baerga-Ortiz, B. Popovic, J. B. Spencer, P. F. Leadlay, *Chem. Biol.* **2006**, *13*, 287.

[12] A. P. Siskos, A. Baerga-Ortiz, S. Bali, V. Stein, H. Mamdani, D. Spiteller, B. Popovic, J. B. Spencer, J. Staunton, K. J. Weissman, et al., *Chem. Biol.* **2005**, *12*, 1145.

[13] D. H. Kwan, M. Tosin, N. Schläger, F. Schulz, P. F. Leadlay, *Org. Biomol. Chem.* **2011**, *9*, 2053.

[14] A. Baerga-Ortiz, B. Popovic, A. P. Siskos, H. M. O'Hare, D. Spiteller, M. G. Williams, N. Campillo, J. B. Spencer, P. F. Leadlay, *Chem. Biol.* **2006**, *13*, 277.

[15] J. Zheng, S. K. Piasecki, A. T. Keatinge-Clay, *ACS Chem. Biol.* **2013**, *8*, 1964.

[16] M. Häckh, M. Müller, S. Lüdeke, *Chemistry* **2013**, *19*, 8922.

[17] S. Piasecki, A. Keatinge-Clay, *Synlett* **2012**, *23*, 1840.

[18] K. K. Sharma, C. N. Boddy, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 3034.

[19] A. T. Keatinge-Clay, R. M. Stroud, *Structure* **2006**, *14*, 737.

[20] S. A. Bonnett, J. R. Whicher, K. Papireddy, G. Florova, J. L. Smith, K. A. Reynolds, *Chem. Biol.* **2013**, *20*, 772.

[21] J. Zheng, C. A. Taylor, S. K. Piasecki, A. T. Keatinge-Clay, *Structure* **2010**, *18*, 913.

[22] A. T. Keatinge-Clay, *Nat. Prod. Rep.* **2012**, *29*, 1050.

[23] S. K. Piasecki, J. Zheng, A. J. Axelrod, M. E Detelich, A. T. Keatinge-Clay, *Proteins* **2014**, *82*, 2067.

[24] J. Kim, J. H. Chang, E.-J. Kim, K.-J. Kim, *Biochem. Biophys. Res. Commun.* **2014**, *443*, 783.

[25] S. K. Piasecki, J. Zheng, A. J. Axelrod, M. E Detelich, A. T. Keatinge-Clay, *Proteins* **2014**, *82*, 2067.

[26] N. Moebius, C. Ross, K. Scherlach, B. Rohm, M. Roth, C. Hertweck, *Chem. Biol.* **2012**, *19*, 1164.

[27] K. Ishida, T. Lincke, C. Hertweck, *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 5470.

[28] D. Janssen, D. Albert, R. Jansen, R. Müller, M. Kalesse, *Angew. Chem.* **2007**, *46*, 4898.

[29] D. Menche, F. Arikan, O. Perlova, N. Horstmann, W. Ahlbrecht, S. C. Wenzel, R. Jansen, H. Irschik, R. Müller, *J. Am. Chem. Soc.* **2008**, *130*, 14234.

[30] C. R. Valenzano, Y.-O. You, A. Garg, A. Keatinge-Clay, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.* **2010**, *132*, 14697.

[31] A. Mengel, O. Reiser, *Chem. Rev.* **1999**, *99*, 1191.

[32] W. S. Knowles, *Acc. Chem. Res.* **1983**, *16*, 106.

[33] K. C. Harper, E. N. Bess, M. S. Sigman, *Nat. Chem.* **2012**, *4*, 366.

[34] A. Milo, A. J. Neel, F. D. Toste, M. S. Sigman, *Science* **2015**, *347*, 737.

Experimental Section for Chapter 3

General Considerations

Anhydrous tetrahydrofuran (THF) was distilled from sodium and benzophenone, and dichromethane (DCM) was distilled from calcium hydride. Thin layer chromatography (TLC) was conducted with EMD gel 60 F_{254} pre-coated plates (0.25 mM). Fisher scientific silica gel 60 (particle size 230-400 μm) was used for flash column chromatography. ^{1}H NMR data were acquired on a Varian Mercury 400 MHz instrument at ambient temperature and are reported in terms of chemical shift (δ ppm), multiplicity, coupling constant, and integration and are referenced downfield from $(CH_3)_4Si$ to the residual solvent peak at 7.26 ppm for $CDCl_3$ as an internal standard. ^{13}C NMR data were acquired on either a Varian Mercury 400 MHz, 500 MHz, or 600 MHz instrument and are reported in terms of chemical shift and referenced to the residual solvent peak at 77.16 ppm for $CDCl_3$ as an internal standard. High-resolution mass spectrometry measurements were obtained by chemical ionization (CI) with a VG analytical ZAB2-E instrument. Characterization of **3.2**,[1] (E)-4-ethylidene-3-methyloxetan-2-one,[2] 2-mercaptoethyl acetate,[3] and acyl oxozolidone intermediates (*S*)-4-benzyl-3-propionyloxazolidin-2-one,[1, 4-7] (*S*)-4-benzyl-3-((2*S*,3*R*)-3-hydroxy-2-methylpentanoyl)oxazolidin-2-one,[6] (*S*)-4-benzyl-3-((2*R*,3*R*)-3-hydroxy-2-methylpentanoyl)oxazolidin-2-one,[8] and (*S*)-4-benzyl-3-((2*R*,3*S*)-3-hydroxy-2-methylpentanoyl)oxazolidin-2-one[6]) were in accordance with literature reported data.

Protein Expression and Purification

Bacillus subtillus glucose dehydrogenase (GDH),[1,9] TylKR1,[1,9] EryKR1,[1,9] AmpKR2,[1,9] RifKR7,[2,10] and all their respective point mutants were expressed in *E. Coli* BL21 (the expression plasmid for all proteins was pET28b, except RifKR7, which was in

pET28a). Starter cultures (50 mL) were grown to inoculate pre-warmed Luria broth supplemented with 25 mg/L kanamycin. 5 mL of starter culture was added to 1L of Luria broth supplemented with 25 mg/L kanamycin, and cultures were grown at 37°C until OD_{600}=0.4. When OD_{600}=0.4, the media was cooled to 15 °C and then induced with 0.5 mM IPTG. After 16 hours, the proteins were harvested by centrifugation (3,000 xg for 20 minutes), and the pellets were re-suspended in lysis buffer (100 mM HEPES, 500 mM NaCl, 10% v/v glycerol, pH 7.5). The cells were then lysed by sonication on ice and centrifuged (30,000 xg for 45 minutes) to remove cellular debris. The protein was purified by passing the crude lysate over a nickel-NTA column equilibrated with lysis buffer. The column was then washed with lysis buffer containing 15 mM imidazole, and the protein was eluted with lysis buffer containing 150 mM imidazole. Final protein concentrations were determined using a Thermo Scientific Nanodrop 1000.

BIOCATALYTIC ASSAYS

Reactions were modified from a method described previously[1] using the following conditions: 200 mL HEPES, 100 mM NaCl, 10% (v/v) glycerol, 200 mM D-glucose, 5 mM $NADP^+$, 5 mM diketide substrate (**3.1-3.3**), 0.1 μM *Bacillus subtillus* GDH, and 10 μM KR in a total volume of 1 mL. For reactions with substrate **3.3** 20% (v/v) DMSO was added to solubilize the substrate.[ix] After overnight incubation (18 hours) at 23° C, the reactions were extracted with 2 volumes of ethyl acetate and evaporated to dryness. Subsequently, the reactions were diluted in ethanol (for substrate **1**) or running buffer (for substrates **3.2-3.3**) and analyzed via chiral chromatography. All

[ix] Screens were performed with less DMSO for substrate **2.3**, however activity was not observed without a minimum of 20% (v/v) DMSO because an insufficient amount of the substrate was solubilized. Screens with varying amounts of DMSO were also performed with substrates **1-2**, however because the DMSO interacted with the more polar atoms of the amide and ester and was difficult to remove, it interfered with the resolution on the chiral HPLC.

reactions were performed in duplicate. Product ratios and percent conversions were determined via HPLC integration.

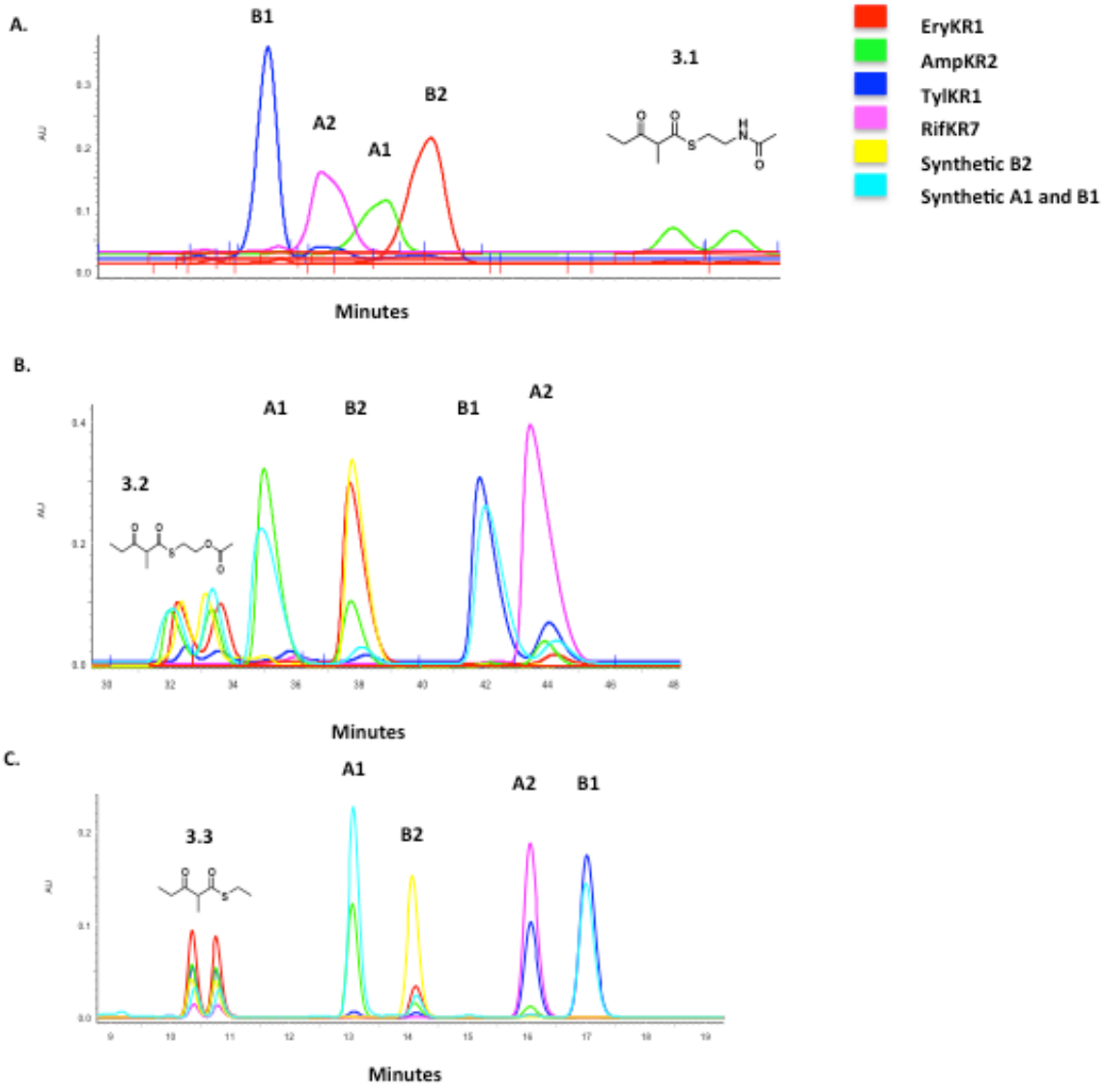


Figure E3.1 Chiral chromatograms of reactions with **3.1, 3.2** and **3.3** and synthetic standards (see synthetic methods section). Standards were spiked with substrates **3.2** and **3.3** confirm the peak alignment. The elution order of the reduction products of **3.1** was reported previously.[1]

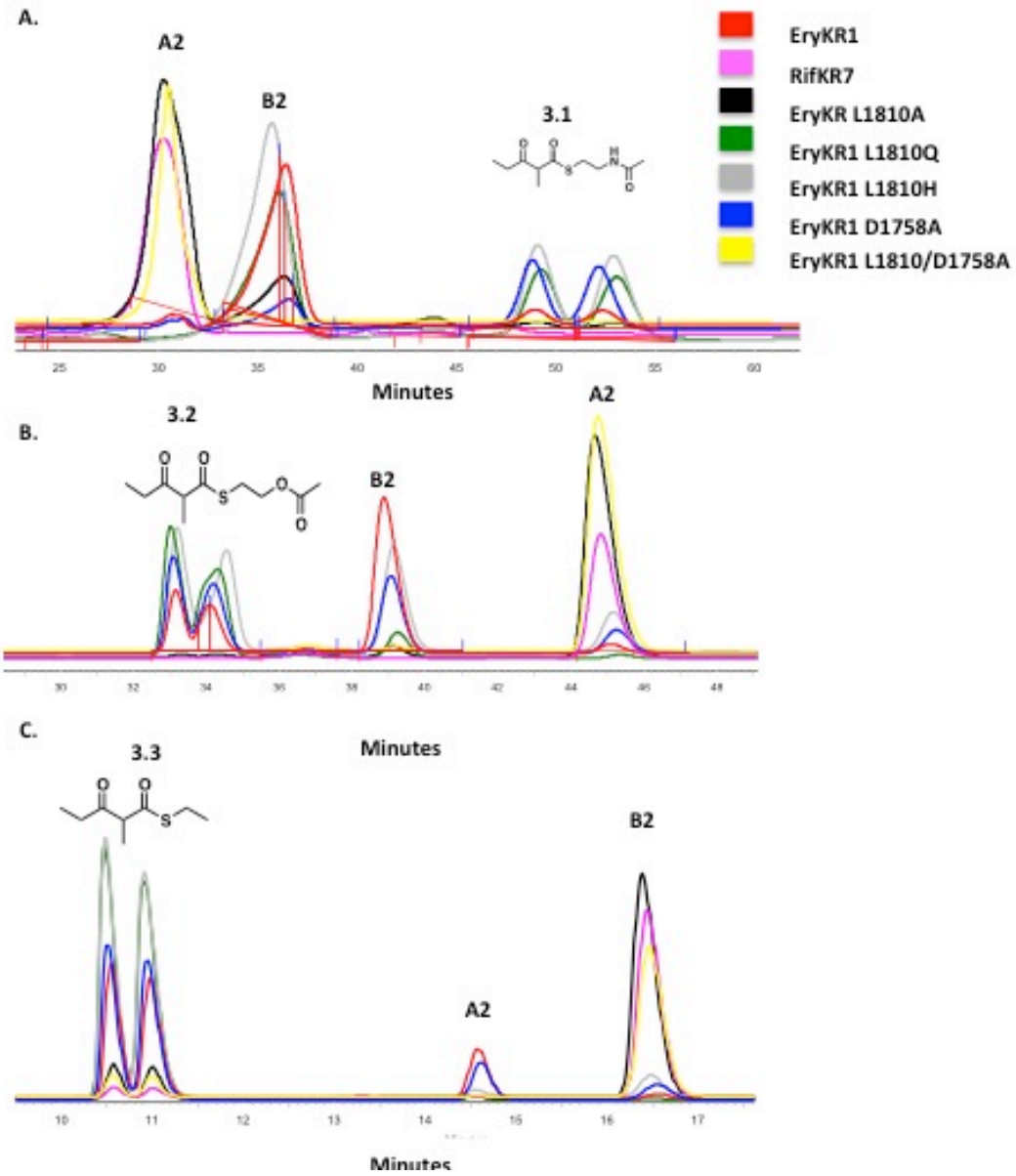


Figure E3.2 Chiral chromatograms of EryKR1 mutant assays. Wild type EryKR1 and RifKR7 are shown for reference

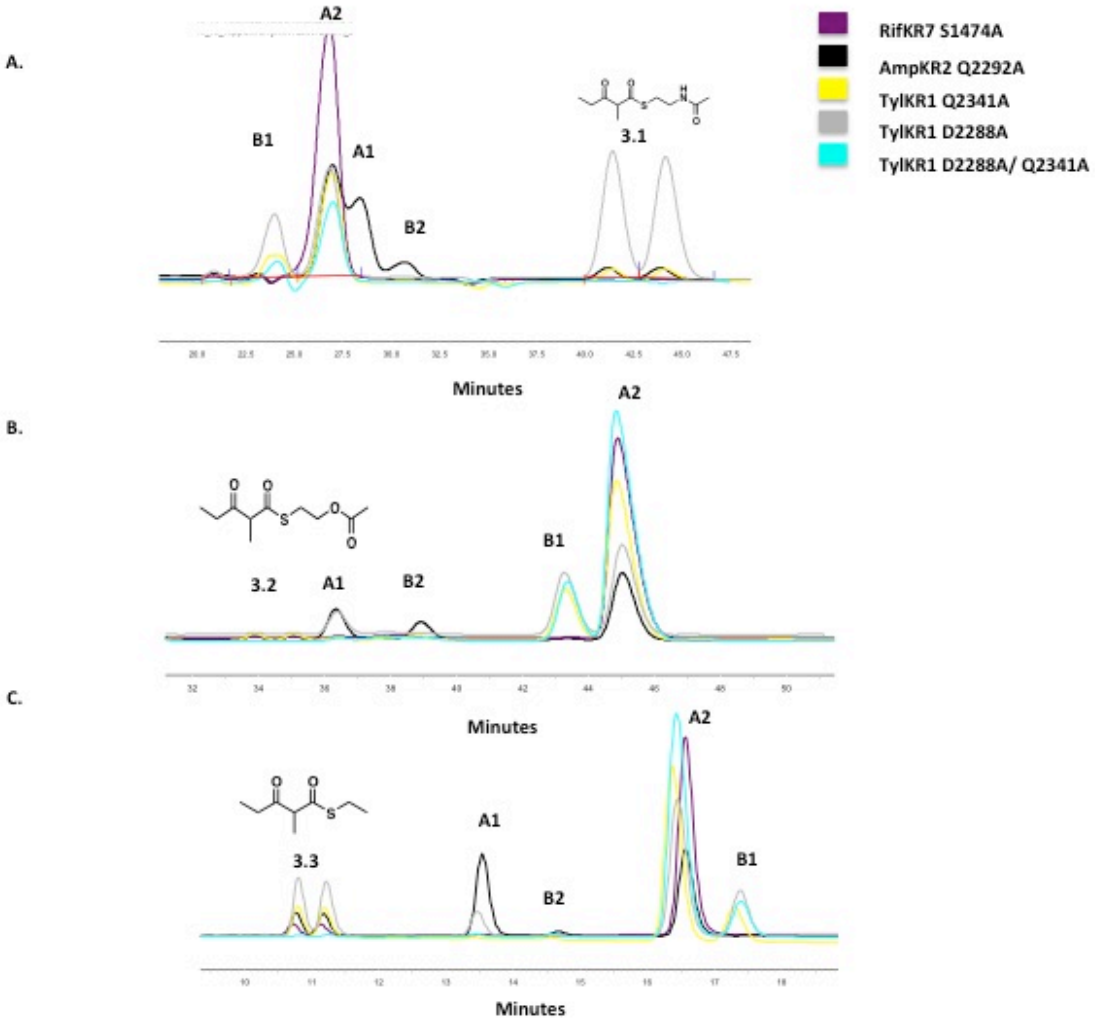


Figure E3.3 Chiral chromatograms of TylKR1, RifKR7, and AmpKR2 alanine mutant assays.

Table E3.1 HPLC conditions for substrates **3.1-3.3.**

Substrate	**Solvent**	**Flow Rate**	**Column (Chiracel)**
3.1	7% ethanol:hexanes	0.8 mL/min	OCH
3.2	2% ethanol:hexanes	0.8 mL/min	OJH
3.3[1]	0.25% isopropanol:hexanes	0.8 mL/min	OJH

Table E3.2 Raw Integrals[a]

KR (Substrate)	B1 (2*R*, 3*R*)	A2 (2*S*, 3*S*)	B2 (S*R*, 3*R*)	A1 (2*R*, 3*S*)	Substrate
EryKR1 (**3.1**)	0 445484	0 0	9214805 23666537	0 0	1469470 956840
EryKR1 (**3.2**)	0 0	1027292 770858	19726215 14191485	0 0	11185389 7968723
EryKR1 (**3.3**)	0 0	0 0	263895 476494	0 0	1016987 2024354
TylKR1 (**3.1**)	21844837 18468769	2133405 1717771	0 0	0 0	0 0
TylKR1 (**3.2**)	26440881 15916294	5621039 3376531	0 461772	1093963 807726	2519523 1484929
TylKR1 (**3.3**)	2949978 3121918	1629280 1681017	74063 78026	83256 86137	839345 1178413
AmpKR2 (**3.1**)	0 0	233136 680406	366565 346325	10402083 14444250	7117744 10956941
AmpKR2 (**3.2**)	160639 278637	1120470 2036427	2753576 4758966	8205203 15085007	4241079 7299425
AmpKR2(**3.3**)	0 0	1119893 202073	1247198 233738	8812280 1603230	4493291 1248737
RifKR7 (**3.1**)	0 470323	15475012 13728778	0 0	0 0	0 423268
RifKR7 (**3.2**)	0 0	29886089 23706392	0 0	0 0	19664 0
RifKR7(**3.3**)	0 0	3116725 2641128	0 0	0 0	312597 272361

Table E3.2

EryKR1 181758A (**3.1**)	0 0	3137430 2813529	16358138 15661836	0 0	918228 0
EryKR1 L1810A (**3.2**)	0 0	5950354 10242816	56160 94210	0 0	1204395 3212731
EryKR1 L1810A (**3.3**)	0 0	53401072 62088435	2536301 2993430	0 0	663022 347471
EryKR1 L1810Q (**3.1**)	0 0	0 0	9123300 10045090	0 0	9046702 4257797
EryKR1 L1810Q (**3.2**)	0 0	369268 725168	2692324 5129830	0 0	24450448 50835477
EryKR1 L1810Q (**3.3**)	0 0	0 0	0 0	0 0	5603593 7945139
EryKR1 L1810H (**3.1**)	0 0	0 0	14625754 7162129	0 0	12109326 26735080
EryKR1 L1810H (**3.2**)	0 0	2459384 1677371	6180311 4225340	0 0	30721895 22531554
EryKR1 L1810H (**3.3**)	0 0	143477 312835	36591 75581	0 0	3284704 10402151
EryKR1 D1758A (**3.1**)	0 0	309678 67129	2606111 1745908	0 0	8477010 6492309
EryKR1 D1758A (**3.2**)	0 0	1678939 1097329	4785579 2579759	0 0	3860198 1498447
EryKR1 D1758A (**3.3**)	0 0	1186544 302256	2040109 520754	0 0	13186106 5240244
EryKR1 L1810A/D1758A (**3.1**)	0 0	7971651 11659601	222041 315671	0 0	293265 46163
EryKR1 L1810A/D1758A (**3.2**)	0 0	10383537 10410108	205691 159205	0 0	0 0
EryKR1 L1810A/D1758A (**3.3**)	0 0	10172575 8275590	61221 46515	0 0	1765948 11600

Table E3.2

TylKR1 Q2341A (**3.1**)	2829087 237764	7321732 853823	0 0	249473 94584	934677 160447
TylKR1 Q2341A (**3.2**)	2474351 2646322	9135642 9946057	0 0	139776 177420	491284 473748
TylKR1 Q2341A (**3.3**)	738336 167537	3590165 806373	39036 8947	32632 5459	872262 394570
TylKR1 D2288A (**3.1**)	2552532 422285	4276178 778821	0 0	0 0	9252685 1848816
TylKR1 D2288A (**3.2**)	3097373 4975899	4924275 7976143	93937 154425	992609 1674712	138423 698819
TylKR1 D2288A (**3.3**)	2657296 433856	7856715 1288805	101167 19890	1074630 178583	4845702 892564
TylKR1 D2288A/Q2341A (**3.1**)	2973482 132245	10808684 648518	0 0	0 0	0 0
TylKR1 D2288A/Q2341A (**3.2**)	2852703 1628102	1396516 8013208	174480 105788	118822 74232	16245 0
TylKR1 D2288A/Q2341A (**3.3**)	114927 197083	736027 1300438	4925 10192	6057 10598	20281 100943
RifKR7 S1474A (**1**)	638489 19455	17411698 1990373	0 0	0 0	0 0
RifKR7 S1474A (**2**)	50070 44112	1188219 10960346	0 0	92692 0	128670 131237
RifKR7 S1474A (**3**)	8493 0	3656946 2590593	0 0	14788 8609	334108 58653
AmpKR2 Q2292A (**1**)	20112 45268	832644 8457910	125251 905308	538268 3135670	1658679 888721
AmpKR2 Q2292A (**2**)	62429 213733	3539596 10430590	857291 2133083	1299603 3583612	252877 110071
AmpKR2 Q2292A (**3**)	0	1608340	79324	1346671	644891

	0	2973851	152171	2393756	1808179

[a] Integrations are reported in arbitrary units.

Table E3.3 Product Ratios (averaged over duplicate runs)

KR (Substrate)	% B1 (2*R*, 3*R*)	%A2 (2*S*, 3*S*)	% B2 (2*S*, 3*R*)	% A1 (2*R*, 3*S*)
EryKR1 (**3.1**)[a]	1	0	99	0
EryKR1 (**3.2**)	0	5	95	0
EryKR2 (**3.3**)	0	0	100	0
TylKR1 (**3.1**)[b]	91	9	0	0
TylKR1 (**3.2**)	79	17	1	4
TylKR1 (**3.3**)	63	34	1.5	1.5
AmpKR2 (**3.1**)[c]	0	3	94	3
AmpKR2 (**3.2**)	1	9	22	68
AmpKR2 (**3.3**)	0	10	11	79
RifKR7 (**3.1**)	1.6	98.4	0	0
RifKR7 (**3.2**)	0	100	0	0
RifKR7 (**3.3**)	0	100	0	0
EryKR1 L1810A (**3.1**)	0	84	16	0
EryKR1 L1810A (**3.2**)	0	95	5	0
EryKR1 L1810A (**3.3**)	0	99	1	0
EryKR1 L1810Q (**3.1**)	0	0	100	0
EryKR1 L1810Q (**3.2**)	0	12	88	0
EryKR1 L1810Q (**3.3**)	0	0	0	0

Table E3.3

EryKR1 L1810H (**3.1**)	0	0	100	0
EryKR1 L1810H (**3.2**)	0	28	72	0
EryKR1 L1810H (**3.3**)	0	80	20	0
EryKR1 D1758A (**3.1**)	0	7	93	0
EryKR1 D1758A (**3.2**)	0	28	72	0
EryKR1 D1758A (**3.3**)	0	37	63	0
EryKR1 L1810A/D1758A (**3.1**)	0	97	3	0
EryKR1 L1810A/D1758A (**3.2**)	0	98	2	0
EryKR1 L1810A/D1758A (**3.3**)	0	99.5	0.5	0
TylKR1 Q2341A (**3.1**)	24	71	0	5
TylKR1 Q2341A (**3.2**)	21	78	0	1
TylKR1 Q2341A (**3.3**)	17	81	1	1
TylKR1 D2288A (**3.1**)	36	64	0	0
TylKR1 D2288A (**3.2**)	34	54	10	2
TylKR1 D2288A (**3.3**)	23	67	1	8
TylKR1 D2288A/ Q2341A (**3.1**)	19	81	0	0
TylKR1 D2288A/ Q2341A (**3.2**)	16	82	1	1
TylKR1 D2288A/ Q2341A (**3.3**)	13	86	0.5	0.5

Table E3.3

RifKR7 S1474A (**3.1**)	2	98	0	0
RifKR7 S1474A (**3.2**)	0.4	98.6	0	1
RifKR7 S1474A (**3.3**)	0.1	99.5	0	0.4
AmpKR2 Q2292A (**3.1**)	1	61	8	30
AmpKR2 Q2292A (**3.2**)	1	63	14	22
AmpKR2 Q2292A (**3.3**)	0	54	2	44

There were deviations from previously published results[1] with **3.1**: [a] Product ratio reported previously: 100% B2 product. [b] Product ratio reported previously: 89% B1 product, 7.9% A2 product, 2.7% A1 product. [c] Product ratio reported previously: 1.3% A2 product, 93.2% A1 product, 5.5% B2 product.

Table E3.4 Summary of Ketoreductase Assays

KR	**Substrate**	**[%] Conversion[a]**	**[%] *de*[b]**	**[%] *ee*[c]**	**Absolute Configuration[d]**
EryKR1	**3.1**	91[e]	98	98	2*S*, 3*R*
	3.2	65	90	90	2*S*, 3*R*
	3.3	25	100	100	2*S*, 3*R*
TylKR1	**3.1**	100	1	83	2*R*, 3*R*
	3.2	93	91	57	2*R*, 3*R*
	3.3	82	93	25	2*R*, 3*R*
AmpKR2	**3.1**	60[f]	93	88	2*R*, 3*S*
	3.2	75	84	35	2*R*, 3*S*
	3.3	67	80	57	2*R*, 3*S*
RifKR7	**3.1**	99	97	97	2*S*, 3*S*
	3.2	100	100	100	2*S*, 3*S*
	3.3	91	100	100	2*S*, 3*S*
EryKR1	**3.1**	98	69	69	2*S*, 3*S*
L1810A	**3.2**	99	91	91	2*S*, 3*S*
	3.3	80	98	98	2*S*, 3*S*
EryKR1	**3.1**	60	100	100	2*S*, 3*R*
L1810Q	**3.2**	10	76	76	2*S*, 3R
	3.3	0	n/a	n/a	n/a
EryKR1	**3.1**	62	100	100	2*S*, 3*R*
L1810H	**3.2**	24	43	43	2*S*, 3*R*
	3.3	4	61	61	2*S*, 3*S*

Table E3.4

EryKR1	**3.1**	24	86	86	2*S*, 3*R*
D1758A	**3.2**	67	44	44	2*S*, 3*R*
	3.3	17	26	26	2*S*, 3*R*
EryKR1	**3.1**	98	95	95	2*S*, 3*S*
L1810A/	**3.2**	100	97	97	2*S*, 3*S*
D1758A	**3.3**	87	99	99	2*S*, 3*S*
TylKR1	**3.1**	90	42	89	2*S*, 3*S*
Q2341A	**3.2**	96	56	97	2*S*, 3*S*
	3.3	77	63	97	2*S*, 3*S*
TylKR1	**3.1**	41	100	27	2*S*, 3*S*
D2288A	**3.2**	97	76	8	2*S*, 3*S*
	3.3	69	80	34	2*S*, 3*S*
TylKR1	**3.1**	100	100	66	2*S*, 3*S*
D2288A/	**3.2**	100	96	63	2*S*, 3*S*
Q2341A	**3.3**	96	97	71	2*S*, 3*S*
	3.1	100	100	95	2*S*, 3*S*
RifKR7	**3.2**	98	98	99	2*S*, 3*S*
S1474A	**3.3**	95	99	99	2*S*, 3*S*
	3.1	92	22	24	2*S*, 3*S*
AmpKR2	**3.2**	98	21	25	2*S*, 3*S*
Q2292A	**3.3**	79	7	7	2*S*, 3*S*

[a] Conversion of total reduced product. [b] *syn: anti* or *anti:syn de* depends on the major product. For wild type EryKR1, AmpKR2, EryKR1 L1810Q, EryKR1 L1810H (with **3.1** and **3.2**), and EryKR1 D1758A the *syn: anti de* was calculated. For RifkR7, TylKR1, EryKR1 L1810H (with **3.3**), EryKR1 D1758A, and EryKR1 L1810A/D1758A, RifKR7 S1474A, AmpKR2 Q2292A, TylKR1 Q2341A, TylKR1 D2288A, and TylKR1 D2288A/Q2341A, the *anti:syn de* was calculated. [c] Enatiomeric excess is calculated for the excess of the major product stereoisomer over all four stereosiomers. [d] Absolute configuration of the major product. [e] Reported previously as 66% conversion.[1] [f] Reported previously as 78% conversion.[1]

SITE DIRECTED MUTAGENESIS

Site directed mutagenesis of EryKR1 used EryKR1 cloned into vector pET28b[1] as a template and used the quickchange method. The following primers were used (mutagenesis sequence indicated in red):

1) For L1810A 5'- gcctttggtgcaccgggt**gc**cggcgggtatgcgccaggcaac-3' and 5'-gttgcctggcgcatacc**gc**cggcacccggtgcaccaaaggc-3'

2) for L1810Q 5'- gcctttggtgcaccgggtc**ag**ggcgggtatgcgccaggcaac-3' and 5'-gttgcctggcgcatacccgcc**ct**gacccggtgcaccaaaggc-3' and

3) for L1810H 5'- gcctttggtgcaccgggtc**a**cggcgggtatgcgccaggcaac-3' and 5'-gttgcctggcgcatacccgccg**t**gacccggtgcaccaaaggc-3' and

4) for D1758A 5’ gcggcggcaaccttggatgccggcaccgtcgatactctg-3’ and 5’ cagagtatcgacggtgccggcatccaaggttgccgccgc-3’. For the double mutant D1758A/L1810A, EryKR1 L1810A in pET28b was used as template.

Site directed mutagenesis of TylKR1 used TylKR1 cloned into vector pET28b[1] as template and used the quickchange method. The following primers were used:

1) for Q380A: 5'-acatggggcaacgccggcgcgggtgcgtacgccgccgccaa-3' and

5'-ttggcggcggcgtacgcacccgcgccggcgttgccccatgt-3'

2) for D327A: 5'-ttccacaccgccgggattctggacgccgcggtgatcgacacgctg-3' and

5'- cagcgtgtcgatcaccgcggcgtccagaatcccggcggtgtggaa

For the double mutant D327A/Q380A, the TylKR1 Q280A in pET28b was used as template.

Site directed mutagenesis of RifKR7 used RifKR7 cloned into vector pET28a[10,11] as template and used the quickchange method. The following primers were used (mutagenesis sequence indicated in red):

1) for S296A: 5'-agcatcttcatgggtgccggcgccggtggttacgcggcagcgaat-3' and

5'-attcgctgccgcgtaaccaccggcgccggcacccatgaagatgct-3'

Site directed mutagenesis of AmpKR2 used AmpKR2 cloned into vector pET28b as template and used the quickchange method. The following primers were used (mutagenesis sequence indicated in red):

for Q380A: 5'-tctggggcagcggtggcgcgcccggctacgccgccgccaa-3' and

5'- ttggcggcggcgtagccgggcgcgccaccgctgccccaga-3'

All mutations were confirmed by DNA sequencing.

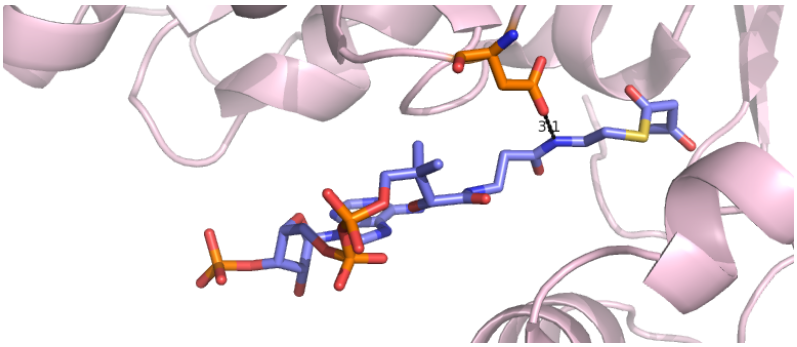


Figure E3.4 Image of the homologous aspartate residue hydrogen bonding with the first amide of the phosphopantetheinyl moiety of acetylacetyl-CoA (analogous to the amide in substrate **3**) in the structure of (***R*)-**3-hydroxybutyryl-coa dehydrogenase From *Ralstonia Eutropha* (pdb code 4N5M).[11]

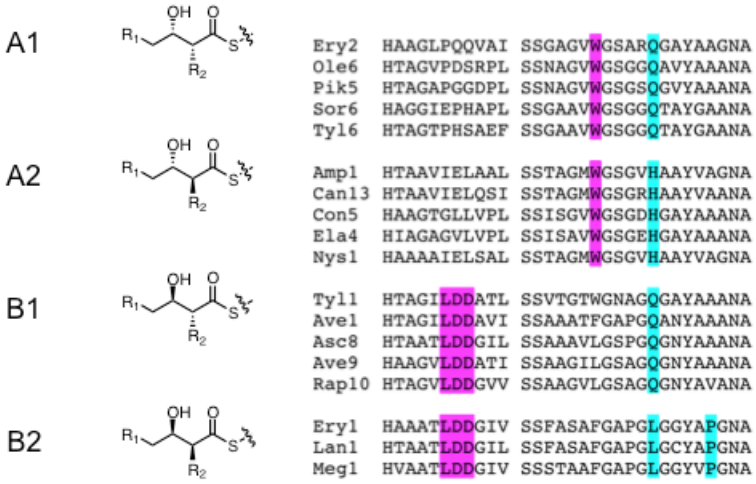
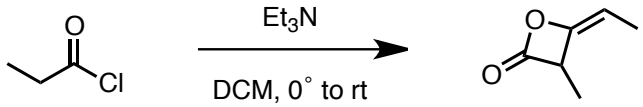
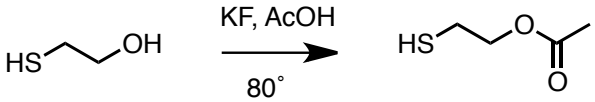


Figure E3.5 Sequence alignment of KRs indicating conserved fingerprint residues. Residues corresponding to hydroxyl stereochemistry (the conserved W of A-type KRs and the LDD of B-type KRs)[12-14] are highlighted in magenta. Residues corresponding to α-substituent stereochemistry are highlighted in cyan.[14]

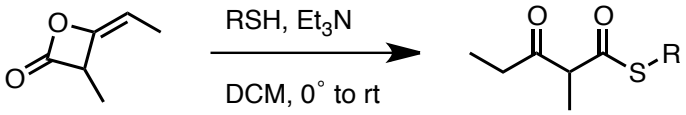
SYNTHETIC METHODS



(E)-4-ethylidene-3-methyloxetan-2-one. Dichloromethane (150 mL) and triethylamine (18.72 mL, 66 mmol, 1 eq) were added over 3Å molecular sieves to a flame-dried flask and cooled to 0°C. Propionyl chloride (19.02 mL, 66 mmol, 1 eq) was added dropwise at 0°C over a period of 90 minutes. The reaction was then stirred at room temperature for 16 hours, concentrated under vacuum, and filtered over celite to remove the triethylammonium chloride salt, affording a yellow oil (2.5 g, 33%). The characterization was in agreement with reported literature data.[2]



2-mercaptoethylacetate. According to a modified literature procedure,[5] 2-mercaptoethanol (0.65 mL, 7.6 mmol, 1 eq) and potassium flouride (0.93 g, 7.6 mmol, 1 eq) were dissolved in acetic acid (15 mL) and heated at 80˚C for 16 hours. The reaction was diluted with water (50 mL), and extracted with ethyl acetate (2x 100 mL). The organic layer was then washed with saturated $NaHCO_3$ (50 mL) and brine (50 mL), dried over Na_2SO_4, evaporated to dryness, and filtered over a plug of silica with 20% ethyl acetate: hexanes to afford a yellow oil (312 mg, 28%). The characterization was in agreement with literature reported data.[3]



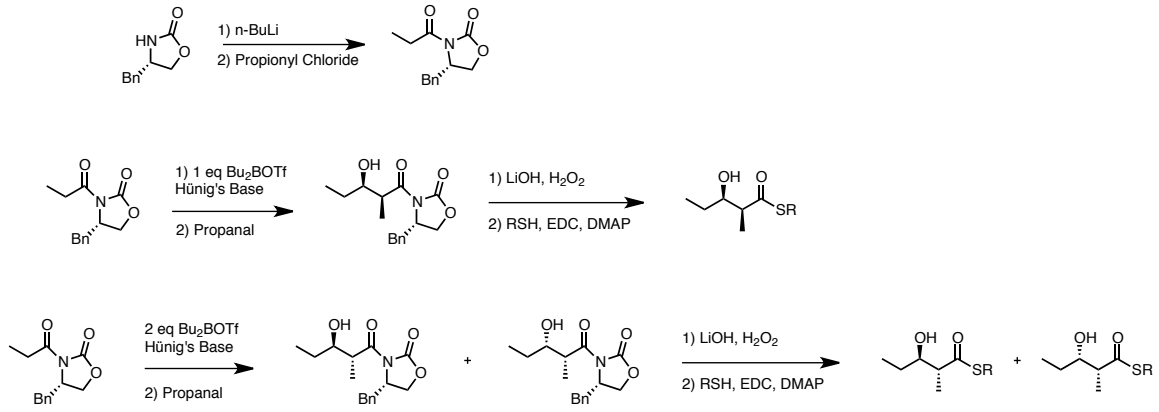
General Synthetic Route for 2-Methyl β-Keto Thioester Substrates (1-3): To a flame dried flask, methyl diketene dimer (427 mg, 3.8 mmol, 1 eq) and thiol (3.8 mmol, 1 eq) were added to dichloromethane (30 mL) at 0°C. Catalytic triethylamine was added, and the reaction warmed to room temperature and stirred for 16 hours. The reaction was then washed with saturated NaCl (1 x 30 mL), extracted in ethylacetate (2 x 50 mL) and concentrated under vacuum to afford the crude product. The crude products were purified via dry flash column chromatography (silica, for 1 50% EtOAC:Hex, for 2 30% EtOAc:Hex, and for 3 EtOAc). An additional semi-preparative HPLC purification step was required for 3 to remove co-eluting impurities: Varian

Microsorb-MV C18 column (250 x 4.6 mm, 5 μm particle size, 100 Å pore size) with a matching Metaguard column, 15-35% B over 30 minutes at 1 mL/min, with mobile phases consisting of water with 0.1% TFA (solvent A) and methanol with 0.1% TFA (solvent B).

(*R,S*)-ethyl 2-methyl-3-oxopentanethioate (1). ^{1}H NMR (400 MHz, $CDCl_3$) δ 3.74 (q, J=7, 1H), 2.91 (q, J=7, 3H), 2.68-2.43 (m, 2H), 1.37 (d, J=7, 3H), 1.26 (t, J=7, 3H), 1.05 (t, J=7, 3H). ^{13}C NMR (100 MHz, $CDCl_3$) δ 205.30, 196.71, 61.14, 35.70, 23.66, 14.44, 13.54, 7.57. HRMS (CI) (m/z) $[M+H]^+$: calcd. for $C_8H_{14}O_2S$: 175.0793, found 175.0792.

(*R,S*) 2-methyl((2-methyl-3-oxopentanoyl)thio)ethyl acetate (2). ^{1}H NMR (400 MHz, $CDCl_3$) δ 4.18 (td, J=6.5 Hz, J=1.3 Hz, 2H), 3.73 (q, J=7 Hz, 1H), 3.14 (t, J=6.5Hz, 2H), 2.65-2.44 (m, 2H), 2.05 (s, 3H), 1.38 (d, 3H), 1.06 (t, 3H, J=7 Hz). ^{13}C NMR (100 MHz,

$CDCl_3$) δ 204.79, 195.80, 170.45, 62.27, 60.91, 34.63, 27.81, 20.57, 13.43, 7.49. HRMS (CI) (m/z) $[M+H]^+$: calcd. for $C_{10}H_{17}O_4S$: 233.0850, found 233.0848.



Scheme E3.1 Route for synthetic standards.

General Synthetic Route for 2-Methyl β-Hydroxy Standards: Standards were synthesized using the method detailed previously by Piasecki et al.[1] Briefly, the (2*S*, 3*R*)-S-methyl 3-hydroxy thioester standards were synthesized via an Evans syn selective aldol reaction with (4,*S*)-Benzyl-3-propionyl-2-oxozolinone and propionaldehyde.[4,5] The (2*R*, 3*R*)-S-methyl 3-hydroxy and (2*R*, 3*S*)-S-methyl 3-hydroxy thioester standards were synthesized as a mixture of diasteromers via the anti selective aldol reaction described by Heathcock et al with (4,*S*)-benzyl-3-propionyl-2-oxozolinone and propionaldehyde. The Heathcock adol uses conditions identical to the classic Evans syn selective aldol, except that two equivalents of dibutylboron triflate are added, and results in a mixture of diasteromers: an anti product with the same hydroxyl stereochemistry the Evans syn product and the non-Evans syn product.[7] Subsequently, the chiral auxiliary was cleaved and the α-methyl β-hydroxy acids were subjected to a Steiglich-type esterification with

the appropriate thiol as described by Boddy et al.[6] The characterizations of the acyl oxazolinone intermediates were in agreement with literature reported data.[1, 4-8]

(2*R*,3*S*)-S-ethyl 3-hydroxy-2-methylpentanethioate. ^{1}H NMR (400 MHz, $CDCl_3$) δ 3.83 (m, 1H), 2.88 (q, J=7, 2H), 2.72-2.66 (m, 1H), 2.38 (br s, 1H), 1.52-1.39 (m, 2H), 1.26 (t, J=7, 3H), 1.21 (d, J=7, 3H), 0.97 (t, J=7, 3H). ^{13}C NMR (100 MHz, $CDCl_3$ δ 204.54, 73.58, 52.69, 27.15, 23.36, 14.79, 11.45, 10.51. ^{13}C (500 MHz, $CDCl_3$): 204.22, 73.43, 52.63, 26.98, 23.14, 14.56, 11.25, 10.29 HRMS (CI) (m/z) $[M+H]^+$ calcd. for $C_8H_{17}O_2S$: 177.0949, found 177.0952.

(2*S*,3*R*)-S-ethyl 3-hydroxy-2-methylpentanethioate and **(2*R*,3*R*)-S-ethyl 3-hydroxy-2-methylpentanethioate** were isolated as an inseparable mixture of diasteromers (4:3 syn:anti dr). ^{1}H NMR (400 MHz, $CDCl_3$) δ syn diasteromer: 3.82 (m, 1H), 2.89 (q, J=7, 2H), 2.72-2.66 (m, 1H), 1.52-1.39 (m, 2H), 1.26 (t, J=7, 3H), 1.21 (d, J=7, 3H), 0.97 (t, J=7, 3H). ^{13}C (100 MHz, $CDCl_3$): 204.26, 75.20, 52.80, 27.10, 24.00, 13.52, 11.54, 9.99δ anti diasteromer: 3.61 (m, 1H), 2.87 (q, J=7, 2H), 2.72-2.66 (m. 1H), 1.57-1.46 (m, 2H), 1.26 (t, J=7, 3H), 1.23 (d, J=3, 2H), 0.98 (t, J=7, 3H). ^{13}C (100 MHz, $CDCl_3$): 204.22, 75.57, 53.32, 27.77, 25.95, 13.11, 10.40. 9.89. HRMS (CI) (m/z) $[M+H]^+$: calcd. for $C_8H_{17}O_2S$: 177.0949, found 177.0954.

2-(((2*R*,3*S*)-3-hydroxy-2-methylpentanoyl)thio)ethyl acetate. ^{1}H NMR (400 MHz, $CDCl_3$) δ 4.32 (t, J=7, 2H), 3.88-3.81 (m, 1H), 3.15 (t, J=7, 2H), 2.76-2.69 (m, 1H), 2.08 (s, 3H), 1.54-1.40 (m, 2H), 1.22 (2, J=7, 3H), 0.98 (t, J=7, 3H). ^{13}C (150 MHz, $CDCl_3$): 202.78, 170.54, 73.34, 62.41, 53.40, 27.41, 29.56, 20.59, 13.70, 10.12. HRMS (CI) (m/z) $[M+H]^+$: calcd. For $C_{10}H_{18}O_4S$: 235.1104, found 235.1106.

2-(((2*S*,3*R*)-3-hydroxy-2-methylpentanoyl)thio)ethyl acetate and **2-(((2*R*,3*R*)-3-hydroxy-2-methylpentanoyl)thio)ethyl acetate** were isolated as an inseparable mixture of diasteromers (3:4 syn: anti dr). (400 MHz, $CDCl_3$) δ syn diasteromer 4.32 (t, J=7, 2H), 3.86-3.81 (m, 1H), 3.15 (t, J=7, 2H), 2.76-2.69 (m, 1H), 2.29 (br d, J=4 HZ, 1H), 2.08 (s, 3H), 1.54-1.40 (m, 2H), 1.22 (d J=7, 3H), 0.98 (t, J=7, 3H). ^{13}C (100 MHz, $CDCl_3$): 204.48, 170.37, 73.20, 62.25, 53.45, 29.39, 27.50, 20.38, 14.44, 9.90 (400 MHz, $CDCl_3$). δ anti diasteromer: 4.18 (t, J=7, 3H), 3.68-3.61 (m, 1H), 3.13 (t, J=7, 2H), 2.78-2.69 (m, 1H), 2.24 (br d, J=7.6), 2.06 (s, 3H), 1.63-1.51 (m, 2H), 1.23 (d J=7, 3H), 0.98 (t, J=7, 3H). ^{13}C (100 MHz, $CDCl_3$): 204.48, 170.34, 74.64, 62.25, 53.70, 28.14, 27.24, 20.34, 14.55, 9.33. HRMS (CI) (m/z) $[M+Na]^+$: calcd. for $C_{10}H_{18}O_4S$: 257.08180, found 257.0821.

REFERENCES

[1] S.K. Piasecki, C.A. Taylor, J.F. Detelich, J. Liu, A. Komsoukaniants, D.R. Siegel, A.T. Keatinge-Clay. *Chem. Biol.* **2011**. *18*, 1331.

[2] K. Sung, S.Y. Wu. *Synth. Comm.* **2001**, 31, 3069.

[3] J.W. Bosco, B.R. Rama Raju, A.K. Saikia. *Synth. Comm.* **2004**. *34*, 2849.

[4] A.K. Evans, J.T. Batroli, L. Shih, L. *J. Am. Chem. Soc.* **1981**. *103*, 2127.

[5] J.R. Gage, D.A. Evans. *Org. Syn.* **1993**, *68*, 83.

[6] K.K. Sharma, C.N. Boddy. *Bioorg. Med. Chem. Lett.* **2006**, *17*, 3034.

[7] B.C. Raimundo, C.H. Heathcock. *Synlett*, 12, 1213-4.

[8] J. Willwatcher, N. Kausch-Busies, A. Fürstner. *Angew. Chem.* **2012**, *48*, 12041.

[9] A.D. Harper, C.B. Bailey, A.D. Edwards, J.F. Detelich, A.T. Keatinge-Clay. *ChemBioChem*. **2012**, *13*, 2000.

[10] C.R. Valenzeno, Y. You, A. Garg, A. Keatinge-Clay, C. Khosla, D.E. Cane. *J. Am. Chem. Soc.* **2010**, *132*, 14697.

[11] J. Kim, J.H. Chang, E. Kim, K. Kim, K. *Biochem. Biophys. Res. Comm.* **2013**, *443*, 783.

[12] P. Caffrey, *ChemBioChem.* **2003**, 4, 654.

[13] R. Reid, M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanthan, J. Carney, D.V. Santi, C.R. Hutichinson, R.A. McDaniel. *Biochemistry*. **2003.** *42*, 72.

[14] A.T. Keatinge-Clay. *Chem. Biol.* **2007**, *14*, 898.

Chapter 4: A *Trans*-Acyltransferase Ketoreductase Reveals Architectural Features and Sequence Elements Unique to Split Bimodules[x]

INTRODUCTION

Complex polyketides comprise a large, pharmaceutically relevant class of natural products biosynthesized by the multienzyme assembly lines, polyketide synthases (PKSs). Modular type I PKSs are comprised of a series of repeating enzymatic domains that each perform a discrete enzymatic function. These domains organize into "modules" that correspond to one round of elongation and reductive processing of a keto group. Type I PKSs are further subdivided into two evolutionarily distinct classes based on whether their acyltransferases (ATs) are embedded within the PKS or exist as free standing protein domains.[1] Of these two PKS classes, the "textbook" rules of polyketide assembly (wherein there is a one-to-one correspondence between protein sequence and metabolite structure, termed "colinearity") are largely based on the dissection of *cis*-AT PKSs, such as the well studied erythromycin PKS.[2,3] In contrast, the biosynthetic logic of the *trans*-AT systems remains far more elusive, both because *trans*-AT PKSs have been identified more recently, and because of their much higher degree of modular variation which frequently incorporates aberrant domain order, inactive domains and repeating domains, which altogether serves to obfuscate the colinear relationship between synthase and product.[2]

---

[x] Portions of this chapter are based on a manuscript in preparation for publication with the coauthors Drew T. Wagner, Darren C. Gay, Hannah R. Manion, and Adrian T. Keatinge-Clay. DTW and CBB are equally contributing co-first authors. DTW solved the PksKR2 crystal structure with the assistance of DCG, CBB performed the functional analysis, and HRM assisted with protein purification. The majority of the structural and bioinformatic analysis was performed by CBB and DTW. DTW, CBB, and ATK analyzed the data and are in the process of writing the manuscript.

Of the more common structural deviations exemplifying the peculiar architecture of *trans*-AT PKSs are modules "split" between two polypeptides (*i.e.* domains that process the same keto group are appear on separate gene products). One of the most prevalent forms of this "splitting" arises when the dehydratase domain (DH) appears on the N-terminus of a downstream polypeptide. These split bimodules conform to one of two domain orderings, denoted by Piel as type A and type B bimodules.[2] Type A split bimodules, correlated with the addition of two carbons in the form of a *cis*-olefin, are comprised of a ketosynthase, ketoreductase, acyl carrier protein, and inactive ketosynthase (KS+KR+ACP+KS^0) at the C-terminal end of the upstream polypeptide and a dehydratase and acyl carrier protein (DH+ACP) at the N-terminal end of the downstream polypeptide. Type B split bimodules, though similar to type A bimodules, are correlated with the addition of four carbons in the form of an α,β-*trans* γ,δ-*cis*-diene and consist of a ketosynthase, ketoreductase, acyl carrier protein, and a second active ketosynthase (KS+KR+ACP+KS) at the C-terminal end of one polypeptide and a dehydratase, acyl carrier protein, and KR (DH+ACP+KR) at the N-terminal end of the subsequent polypeptide (**Figure 4.1A**).[2]

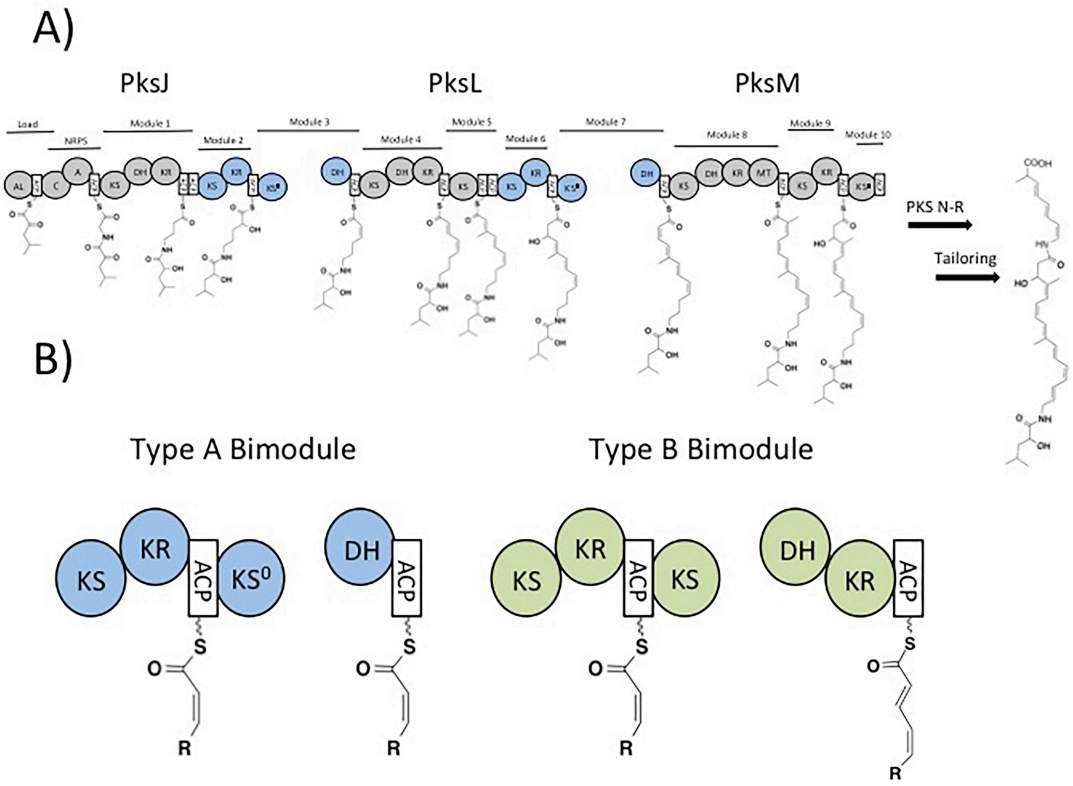


Figure 4.1 A) PksX biosynthetic gene cluster from *Bacillus subtilis*. Two type A bimodules are found, one between the C-terminus of PKSJ and the N-terminus of PkSL and a second between the C-terminus of PkSL and the N-terminus of PkSM. B) Schematic of the two types of split bimodules, the type A bimodule (featuring a non-elogating KS^0 and yielding a *cis* double bond), and the type B bimodule (featuring an elongating KS and yielding an α,β-*trans* γ,δ-*cis*-diene). The timing of the dehydrations is unknown. In type A bimodules, the dehydration could occur on the ACP of the upstream module or the downstream module. For type B bimodules, it is likely that the DH acts twice, once to form a *cis* double bond and once to form a *trans* double bond.

Because there has been little experimental interrogation of split bimodules, their mechanistic features and modular organizations are poorly understood. Indeed, only a

few reported experimental results have uncovered any mechanistic aspects of type A bimodules. Piel and coworkers demonstrated that for both of the type A bimodules in the bacillaene pathway, there was an accumulation of intermediates with mass differences corresponding to the hydrated and dehydrated products on the two ACPs of the bimodules. This result indicates the N-terminal DH as the most likely candidate for catalyzing the dehydration, despite its presence in the downstream polypeptide.[4] A second investigation demonstrated that for type A bimodules, the condensation-inactive KS^0 motif is a crucial feature[5] for polyketide production. Even fewer experimental details regarding the mechanistic aspects of type B bimodules exist. However, chemical and biosynthetic logic suggests that the N-terminal DH likely operates twice: once on the intermediate bound to the ACP of the preceding module and a second time on the intermediate bound to the ACP of the module in which it is located. With regard to chemical logic, the activation of the proton α to the keto moiety is far greater than the proton α to a hydroxy moiety that would result from two sequential dehydrations on the second ACP. With regard to biosynthetic logic, α, β processing seems far more in line with typical keto processing than γ,δ processing.

The polyketide, bacillaene, is produced by a prototypical *trans*-AT PKS, a non-ribosomal peptide synthetase (NRPS)-PKS hybrid endogenous to *B. subtilis* and *B. amyloqiufaciens* FZB42.[2,6–8] Bacillene's polyene structure (**Figure 4.1**) is an example of the collaborative action of tandem pairs of KR and DH domains (two of which are installed by type A split bimodules). In polyketide biosynthesis, *cis* olefins are installed by the *syn*-coplanar dehydration of L-β-hydroxyacyl intermediates, whereas *trans*-olefins arise from D-β-hydroxyacyl intermediates.[9–18] PKS KRs are classified by their stereoselectivity: A-type KRs catalyze the formation of L-β-hydroxyacyl intermediates, whereas B-type KRs catalyze the formation of D-β-hydroxyacyl intermediates. In *cis*-AT

pathways, the KR domains possess signature sequence fingerprints that enable the prediction of the hydroxyl configuration (and thus the olefin geometry installed by the DH). A type KRs contain a conserved tryptophan, whereas B-type KRs contain and a leucine-aspartate-aspartate motif.[11,19,20] The fingerprints found in KRs from *trans*-AT pathways, however, are far less robust. Currently, the only sequence fingerprint correlated with stereochemical outcome in *trans*-AT PKSs is a single aspartate residue, which corresponds to the second D of the "LDD" motif. For *trans*-AT A-type KRs, no sequence fingerprints have been identified; thus assignment of KRs as A-type arises solely from the absence of this diagnostic aspartate.[21–25] As the only crystal structure of a KR from a *trans*-AT pathway is a B-type KR (PksKR1),[25] no structural clues are available that illuminate A-type stereocontrol in *trans* AT pathways.

Here we report the first structural analysis of an A-type KR from a *trans*-AT pathway, the KR from the second module of the bacillaene synthase. In addition, it is the first KR solved that resides within a split bimodule (a type A split bimodule). The 1.98 Å-resolution structure reveals several features that are absent in other structurally characterized KRs. An active site loop reveals an ordered hydrogen-bonding network that may contribute to enforcing A-type stereocontrol. Most strikingly, the structure reveals an exposed hydrophobic helix, which is associated exclusively with the first KRs of split bimodules. Because this helix is present in every type A and B split bimodule KR, we hypothesize that it may be involved in protein-protein recognition events to recruit the upstream DH. Bioinformatic analysis indicates that this helix corresponds to an insertion within the protein sequence, as well as several other diagnostic sequence motifs. Based on the analysis of the KR types present within split bimodules, it appears that type A split bimodules always generate α,β-*cis* olefins, whereas type B split bimodules always generate α,β-*trans*- and γ,δ-*cis* dienes, which has implications for the reassignment of

biosynthetic intermediates in known pathways. We also envision that several of these notable sequence elements will have diagnostic utility in future genome mining applications.

OVERALL STRUCTURE

The crystal structure of PksKR2 was solved to 1.98 Å using molecular replacement with PksKR1 (PDB code 4J1Q) from the bacillaene synthase (from *B. subtilis sp. 168*) as a search model.[25] The construct encoding the polypeptide fragment, PksKR2, spanned residues 3954–4459 of the PksJ protein, with the first 14 and final 33 residues absent in the X-ray structure. The remainder of the structure, aside from the omission of residues 69-73 and 207-210 due to poor electron density, was complete. PksKR2 crystallized in a complex with an $NADP^+$ cofactor in space group $P2_12_12_1$, with a single monomer in the asymmetric unit (**Table 4.1**).

Like previous KR structures, PksKR2 is comprised of two domains: an N-terminal structural subdomain and a C-terminal catalytic subdomain. Both domains contain the Rossmann fold common to short-chain dehydrogenase/reductase (SDR) enzymes, however the structural subdomain lacks the $NADP^+$ cofactor. Clear density was observed for the entire $NADP^+$ cofactor, which makes multiple contacts with the enzyme (**Figure 4.2c**). The active site of the catalytic subdomain includes the critical tyrosine, Y386, which acts as a general acid, as well as the conserved lysine and serine residues (S347 and K371). When comparing the structure to PksKR1 (PDB code: 4J1Q), a high degree of structural similarity between the two proteins is observed with an almost complete overlay of secondary structure (aside from a few structural elements; **Figure 4.2A**), as exemplified by an overall R.M.S.D. of 1.9 Å.

Table 4.1 Data collection and refinement statistics.

Data collection	**PksKR2**
Wavelength (Å)	1.0332
Space group	$P2_12_12_1$
Cell dimensions, *a*, *b*, *c* (Å)	46.1, 84.9, 134.3
Resolution (Å)	50.00–1.98
R_{merge}	0.082 (0.681)
$I/\sigma(I)$	21.8 (1.76)
No. of reflections	34894
Completeness (%)	97.1 (87.2)
Redundancy	6.4 (5.6)
Wilson B value (Å^2)	40.1
Refinement	
Resolution (Å)	71.74–1.98
No. of reflections	34894
R_{work}/R_{free}	0.204/0.236
No. of atoms	
Protein	3565
Water	16
Average *B* factors (Å^2)	
Protein	40.52
Water	30.18
RMS deviations	
Bond lengths (Å)	0.024
Bond angles (°)	1.766
Ramachandran Statistics (%)	
Preferred Regions	97.76
Allowed Regions	2.24
Outliers	0

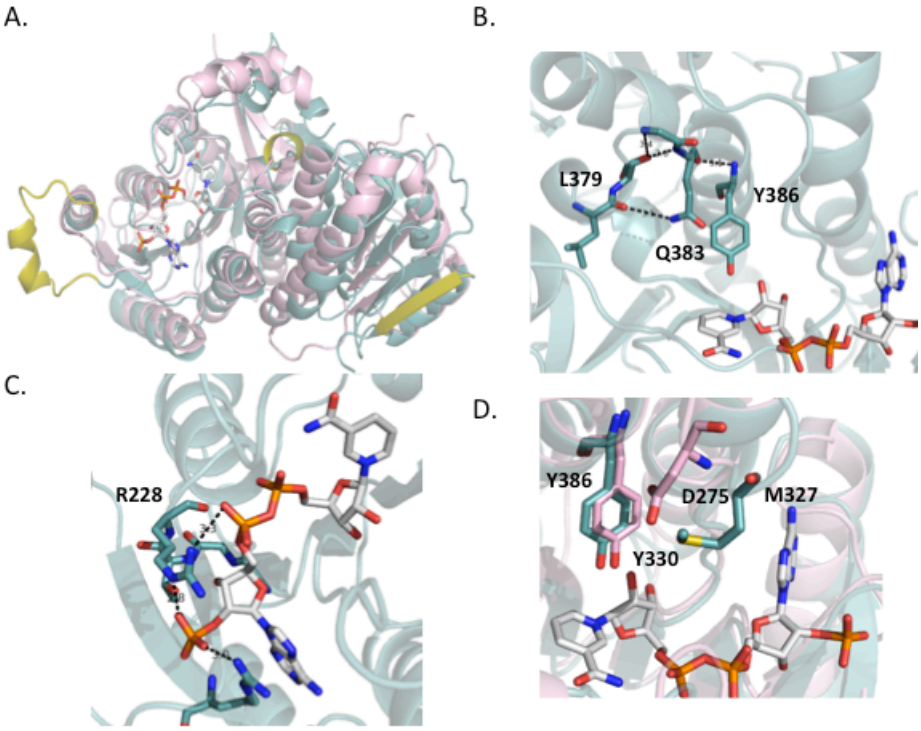


Figure 4.2 New structural elements of PksKR2. A) Structural overlay of PksKR2 and PksKR1 (4j1q). New structural features (including the insertion helix, an additional β-strand, and a new ordered loop) are shown in gold. B) The hydrogen bonding network found in the new active site ordered loop region preceding to the catalytic tyrosine, featuring Q383 and various backbone carbonyls, including L379. C) The NADPH binding region has a new ionic interaction with R228. D) Overlay of PksKR1 and PksKR2 in the "LDD loop" region, indicating that in PksKR1, the diagnostic D275 is poised to interact with the elongating intermediate, whereas in PksKR2 M327 is in this position. Additionally, the LDD loop of PksKR1 is larger by approximately 3 residues, which orients the flexible loop to interact with the substrate in a different fashion.

DEVIATIONS FROM OTHER KETOREDUCTASE STRUCTURES

The structure of PksKR2 reveals several features not observed in other KR structures. The conserved motif of the dinucleotide binding site present in both A and B type KRs from *cis* and *trans*-AT pathways, is altered in these KRs compared to all other

types (typical consensus sequence: GGTGGTGxLG; split bimodule KR consensus sequence: GGTRGLG[A,L]). In particular, this conserved arginine (R228) makes ionic interactions with both the 5' and 2' phosphates of the $NADP^+$ cofactor (**Figure 4.2c**). Additionally, PksKR2 contains an extra β-sheet in comparison to PksKR1 in the structural subdomain and a new ordered loop in the active site region appears in PksKR2, which, by sequence, appears to be absent *trans*-AT KRs that are not embedded within split bimodules (**Figure 4.2A-B, Figure 4.3**). The most striking difference between PksKR2 and PksKR1 is the presence of an additional surface helix of approximately 15 residues located approximately 20 residues downstream of the dinucleotide-binding site within the C-terminal catalytic subdomain of PksKR2. This inserted helical region sharply protrudes from structure and is not present in any of the other nine KRs solved to date (**Figure 4.2A**).

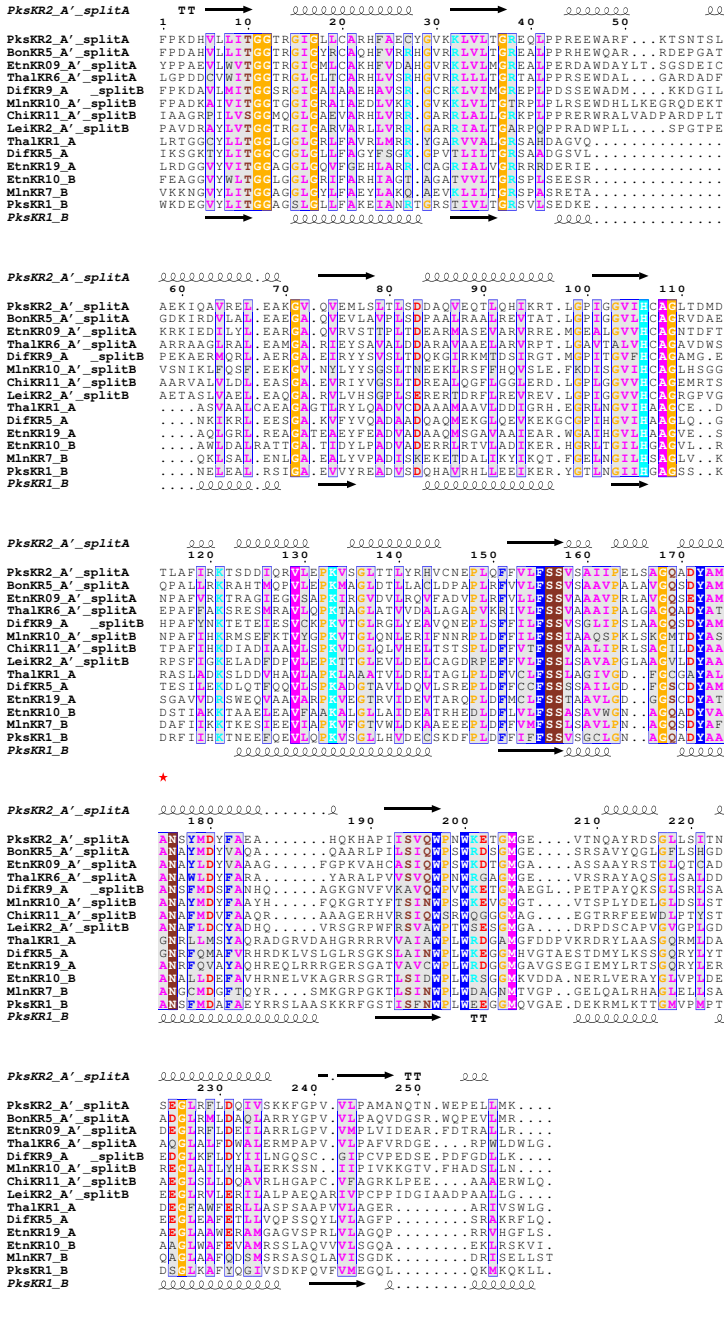


Figure 4.3 Sequence alignment of the catalytic subdomain of *trans*-AT KRs. Split bimodule KRs are denoted as A'. The diagnostic B-type D is indicated by a red asterisk.

STRUCTURAL FEATURES THAT MAY CONTRIBUTE TO A-TYPE STEREOCONTROL

Several structural features of PksKR2 illuminate how A-type stereocontrol may be enforced. The structural dissection of PKS KRs has revealed that all reductase-competent KRs bind NADPH in the same orientation, transferring the 4-*pro*-*S*-hydride to the β-carbonyl of the polyketide intermediate.[11,12,25–29] Thus, KR stereocontrol arises solely from the substrate orientation, which is dictated by the stereoelectronic environment of the active site. One of the residues hypothesized to interact with the β-keto moiety is the third D of the LDD motif. This D is located on a loop that appears directly above the catalytic tyrosine, termed the "LDD loop.[11,30,31] It has been suggested that this conserved D hydrogen bonds with the phosphopantetheinyl arm,[11,12] thus guiding the intermediate to expose the appropriate face to the NADPH for B-type ketoreduction. In support of this hypothesis, this hydrogen-bonding interaction was observed between an aspartate homologous to the invariant aspartate of the LDD loop and the first pantetheinyl amide in a structure of a related enzyme, PhaB from *Ralstonia eutropha* H16.[32] As PksKR1 is a B-type KR, it possesses this conserved aspartate (D275). In contrast to PksKR1, the LDD loop region of PksKR2 appears to be expanded by approximately three residues (**Figure 4.2D, Figure 4.3**). Indeed, by sequence analysis, KRs associated with split bimodules are always A-type and typically appear to have about 3-4 additional residues in the LDD loop region in comparison to other KRs from *trans*-AT pathways (**Figure 4.3**). The residue M327 of PksKR2 roughly overlays with D275 of PksKR1 in PksKR2 suggesting the possibility that the combination of additional residues and the steric bulk of M327 positions this loop such that it precludes the facial orientation that leads to B-type ketoreduction in split bimodules.

PksKR2 also possesses a conserved glutamine residue (Q383) three residues prior to the catalytic tyrosine, which appears roughly analogous to a conserved glutamine

residue observed in two solved *cis* AT A-type KRs (the second KR from the amphotericin PKS, AmpKR2 and the first KR from the phospholactomycin PKS, Plm1).[28,30] Mutagenesis and modeling studies with AmpKR2 suggest that this conserved glutamine residue interacts with the substrate, impacting stereocontrol.[28,33,34] Directly preceding the catalytic tyrosine is a new structured loop with an ordered hydrogen-bonding network that appears to position both Y386 and Q383. This structured loop provides numerous hydrogen bonding interactions with the backbone carbonyls and appears to be dependent on an invariant proline that precedes the glutamine by 6 residues (P377 in PksKR2; **Figure 4.2B, Figure 4.3**). Of particular note, the backbone carbonyl of L379 (which is a highly conserved residue in split bimodule KRs; **Figure 4.3**) appears to play a critical role orienting Q383 in this hydrogen-bonding network. Thus, the conformation of Q383 is likely stabilized in such a way where it is primed to interact with the β-keto substrate, guiding the substrate to adopt the correct facial orientation for A-type ketoreduction. Oddly, by inspection of the sequence alignment, the residues correlating to this ordered loop are absent in A-type KRs that are not harbored within split bimodules (**Figure 4.3**), further highlighting the evolutionary divergence.

STEREOCHEMICAL ASSAY

To further confirm the assignment of PksKR2 as an A-type KR, we incubated PksKR2 with a common substrate for KR activity assays (β-keto *N*-acetyl cysteamine thioester, **4.1**) in the presence of an NADPH regeneration system (*B. subtilis* glucose dehydrogenase and glucose) and analyzed the products via chiral HPLC (**Figure 4.4**).[25,35] Although it is possible to observe results that are artifactual due to the truncated nature of *N*-acetyl cysteamine substrates,[36] typically the ketoreduction products of these analogs corroborate with the reduction observed on their natural substrates (this trend

appears to hold in particular with KRs that are found early in the pathway and naturally reduce smaller substrates such as PksKR2).[37] For comparison, identical ketoreduction assays were performed with two characterized *cis*-AT KRs: the first KR from the tylosin synthase (TylKR1), which has been shown generate the *R* stereoisomer (**4.1A**) with 100% *ee* on substrate **4.1**, and the fifth KR from the mycolactone synthase (MycKR5), which has been shown to generate the *S* stereoisomer (**4.1B**) with 93% *ee* on substrate **4.1**.[37] In agreement with the bioinformatic identification of PksKR2 as an A-type KR, the major product of PksKR2 co-eluted with the major ketoreduction product of MycKR5, revealing its identity to be the *S* stereoisomer (**1b**), whereas the minor product of PksKR2 co-eluted with TylKR1 revealing its identity to be the *R* stereoisomer (**1a**) (the *ee* for PksKR2 was 64%) (**Figure 4.3**).

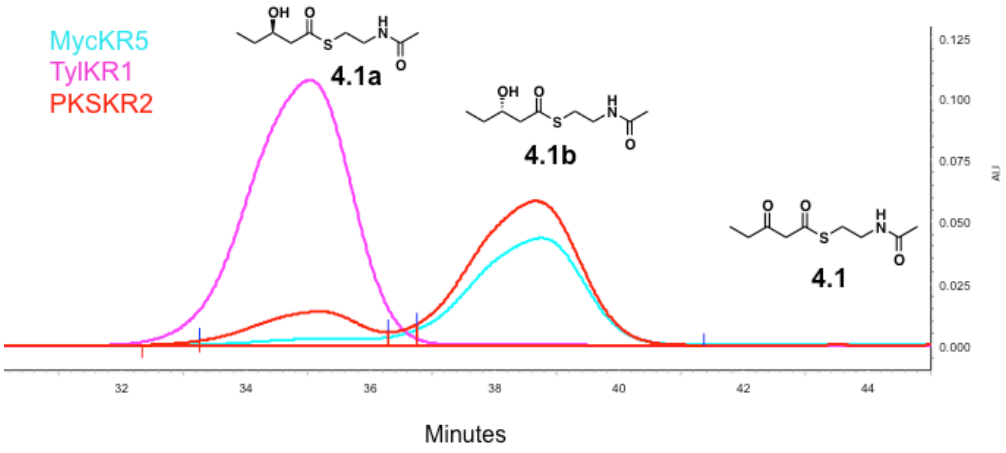


Figure 4.4 Stereochemical characterization of PKSJKR2 with an *N*-acetyl cysteamine-bound substrate analog. Red corresponds to the reduction of **4.1** by PKSJKR2, cyan to the reduction of **4.1** by TylKR1 (a B type KR), and magenta to the reduction of **4.1** by MycKR5 (an A type KR).

BIOINFORMATIC ANALYSIS REVEALS SEQUENCE SIGNATURES OF SPLIT BIMODULE KRS

In light of the unique structural features of PksKR2 compared to PksKR1 and previous *cis*-AT structures, we began a focused bioinformatic investigation to compare these novel features to KRs from other *trans*-AT pathways. An extended multiple sequence alignment revealed that these split bimodule KRs are well distributed across *trans*-AT systems, occurring in most of the better characterized pathways including difficidin, leinamycin, etnangien, the thailandamides, macrolactin, oxazolamycin, kirromycin, bongkrekic acid and pederin. The alignment also revealed that all A-type KRs fall into one of two distinct groups, with one is generally found in the context of canonical PKS module organization reminiscent of *cis*-AT pathways and the other almost always found in modules that are split or structurally anomalous in some other fashion. The traditional subclass of A-type KR shows a high degree of sequence similarity to B-type *trans*-AT KRs (**Figure 4.3**), with the only clear distinguishing residue being the

previously mentioned diagnostic aspartate residue.[21–25] Aside from this single residue, the consensus sequences of B type and the majority of A type KRs are essentially identical (**Figure 4.3**), rendering the sequence dissimilarity of split bimodule A type KRs all the more remarkable.

Several of the deviations found in this unusual KR class are conserved enough at the sequence level to be diagnostically important. Included among these distinguishing elements is the ~15-residue insertion helix lying approximately 25 residues C-terminal to the NADPH binding motif with the consensus sequence LPPRxEW (in PksKR2, the insertion corresponds to residues L255-L271) (**Figure 4.2A**). Additionally, there is a diagnostic fingerprint that appears directly above the LDD loop region with consensus sequence [P,L]AF[L,I,V,A]RK (residues A331-K335 in PksKR2) (**Figure 4.3**). Also, the aforementioned nicotinamide-binding motif has a G to R substitution (**Figure 4.2C, Figure 4.4**) and although noted previously in the literature, it has not to our knowledge been associated with type A and type B split bimodules.[38]

EXCEPTIONS: KRS HARBORING THESE MOTIFS THAT APPEAR OUTSIDE THE CONTEXT OF A SPLIT BIMODULE

We uncovered no example of a type A or type B bimodule not containing a KR with these motifs as the penultimate domain (excluding ACPs) in the upstream protein. However, a few examples of KRs harboring these atypical motifs were found outside the context of type A and type B bimodules. The eighth KR from the oxazolomycin pathway (OzmKR8) contains the motifs detailed above, even though it falls within a canonical modular organization wherein the DH is present within the same polypeptide.[39] The second KR from the myxovirescin pathway (TAKR2) also possesses these motifs, despite being outside the context of a split bimodule. However, TAKR2 has an unusual modular organization as it appears C-terminal to an enoyl-CoA hydratase (ECH) and N-terminal

to a KS^0, which are both domains associated with the isoprenoid-like β-branching activity.[2,40–42] However, the functions of these domains remain mysterious, as β-branching does not occur in TA_Mod2.[2,38,42] The second KR from the pederin pathway (PedKR2), also possesses the features described above and occurs C-terminal to an uncommon pyran synthase domain, which cyclizes the polyketide to form a pyran moiety.[43] In both of these cases, then, it is possible that the structural features of split bimodule KRs have been co-opted to mediate uncommon domain activities. The chivosazol PKS also harbors a KR with these unique features, outside of the context of a split bimodule (ChiKR14 in ChiF). However, it appears that ChiDH15 performs two dehydrations, dehydrating the hydroxyls installed by ChiKR14 and ChiKR15, yielding an α,β-*trans*- γ,δ-*cis*-diene.[44] Given the presence of the sequence signatures in ChiKR14 and the double dehydrating activity of ChiDH15, it is tempting to hypothesize that this ChiF bimodule may be a recent descendant from a type B split bimodule that evolved to merge the two polypeptides.

PROTEIN-PROTEIN INTERACTIONS IN SPLIT BIMODULES

Split modules necessitate inter-protein interactions to bridge the vertically adjacent polypeptides. Given the strict inclusion of these unprecedented KR features in both bimodule types, one hypothesis for their distinct structural and sequence features may be their involvement in protein-protein recognition. Since *trans*-AT systems lack the N- and C-terminal docking domains (NDD and CDD regions) that link adjacent PKS polypeptides in the *cis*-AT systems,[45–47] adjacent proteins must rely on inter-domain binding contacts as the sole mediators of synthase assembly. As the insertion helix sequence is a conserved element and contains several surface-exposed hydrophobic residues, we hypothesize that it may have a role in mediating protein-protein interactions.

One hypothesis is that the exposed helix may be involved in recruiting the DH from the next polypeptide. In type B bimodules, such an interaction would be necessary, as the DH is likely visited by the ACP from both the downstream and upstream polypeptides. In type A bimodules, it is unknown whether the DH performs the dehydration on the upstream or downstream ACP. Regardless of the timing of the dehydration events, clearly these two bimodular organizations are evolutionarily related, and thus likely rely on similar inter-peptide interactions to ensure chain transfer to the next polypeptide. Notably, the structure of N-terminal DH from a type B bimodule from the difficidin PKS, Dif10 (unpublished results, Dr. Jia Zeng) reveals that N-terminal DHs from split bimodules are truncated relative to embedded DHs. This truncation exposes a hydrophobic region along the surface that is a plausible binding partner for the insertion helix.

REASSIGNMENT OF INTERMEDIATES IN KNOWN PATHWAYS

In our bioinformatic investigations we discovered an invariant pattern in type B bimodules with regard to the KR types present. In the first module of the bimodule, the KR is invariantly A-type and possesses the unique features associated with split bimodule A-type KRs. The KR of the second module always possesses the signature aspartate residue, indicative of B-type reduction. With the biosynthetic logic detailed earlier, the resultant dehydration products would be an α,β-*trans*- γ,δ-*cis*-diene. Although the majority of annotated *trans*-AT metabolites and their corresponding intermediates are in agreement with this prediction,[2] there are a few examples for which the intermediates resulting from type B bimodules are annotated as having α,β-*cis,* γ,δ-*cis* or α,β-*trans,* γ,δ-*trans* intermediates. We believe that these cases wherein α,β-*trans*-γ,δ-*cis* are not annotated as such in the literature are likely the result of either misannotation, unstable

polyene structures, or the action of a downstream isomerase. For example, in both the difficidin and macrolactin pathways, there are intermediates arising from type B bimodules which are typically denoted in the literature as having α,β-*cis*- γ,δ-*cis*-dienes (in difficidin, the type B bimodule is present between DifI and DifJ, and in macrolactin, the type B bimodule is present between MlnF and MlnH).[2,48–50]

For each of these cases, inspection of the KR sequences harbored within the type B bimodule yielded the aforementioned pattern of an A-type KR with the distinctive sequence motifs detailed above followed by a B-type KR suggesting that the intermediates would have α,β-*trans*- γ,δ-*cis*-diene geometries rather than the variant diene geometries denoted in the literature.[2] The difficidin PKS presents an interesting case wherein a type B bimodule (between DifI and DifJ, modules 10 and 11) is immediately followed by a type A bimodule (between DifJ and DifK, module 12). DifK also harbors an unusual DH following the N-terminal DH of module 12, which currently has no annotated function (**Figure 4.5A**). Presumably, the double bond geometries assigned in the literature of difficidin intermediates in modules 10-12 are inferred from the geometry of the triene in the final metabolite. However, in addition to the bioinformatic discordance of a α,β-*cis*- γ,δ-*cis*-diene in the type B bimodule, the type A bimodule present between DifJ and DifK is annotated as a *trans* olefin (as mentioned previously, type A bimodules are believed to always generate *cis* olefins[2]). The disparity between the triene geometry present in the final metabolite and the preceding intermediate triene geometry can be explained by the presence the additional DH domain in DifK. Inspection of the sequence of the second consecutive DH of DifK revealed that in place of the active site aspartate, there is an asparagine. This aspartate to asparagine substitution is associated with isomerase activity: this sequence signature is found both in DH-like domains that have pyran synthase activity[43] as well as "shift" enoyl isomerases

which isomerize olefins from the α,β position to the β,δ position (Gay et al, 2014; Mouldenhauer, 2010; Kusebach et al, 2010). Thus the disparity between the final triene geometry in the metabolite and the bioinformatic predictions is explainable by this DH-like isomerizing domain (**Figure 4.5B**).

Macrolactin, at first inspection presents a similar case, in that the intermediates annotated as having α,β-*cis,* γ,δ-*cis* geometry, along with the presence of an additional dehydratase present on the C-terminus of MlnF.[2] The C-terminal dehydratase also possesses the diagnostic aspartate to asparagine mutation. However, inspection of the final metabolite corroborates with the bioinformatic prediction of the diene geometries from a type B bimodule (α,β-*trans*- γ,δ-*cis*-diene), suggesting that the intermediate structure was misannotated in the literature (**Figure 4.5C-D**). Thus, it is possible that the additional dehydratase in this case is present for structural reasons, perhaps providing a bridging interaction with the N-terminal KR of the upstream polypeptide, rather than isomerase activity. Thus, in all three cases detailed above, it is likely that the diene geometry forms as α,β-*trans*- γ,δ-*cis,* from the type B bimodule as opposed to the variant geometries shown in the literature (**Figure 4.5**).

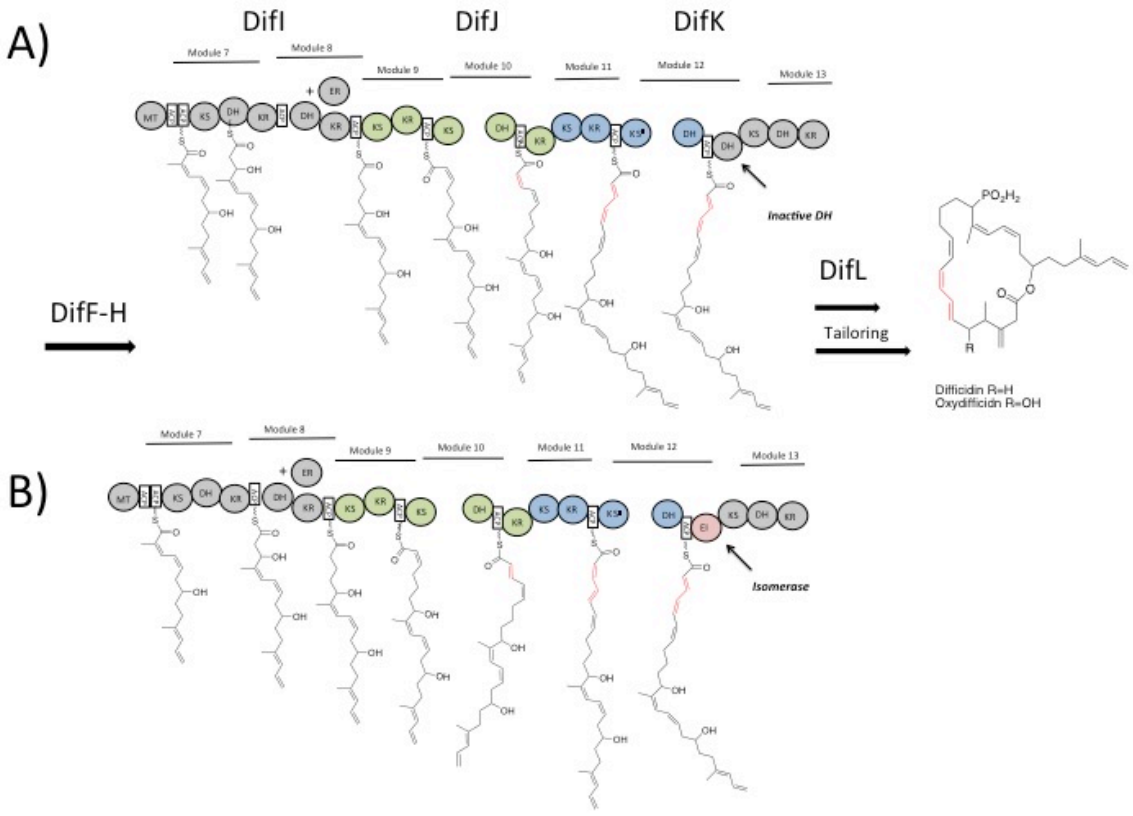


Figure 4.5 A) Current assignment of intermediates in the difficidin pathway B) Reassignment of intermediates in the difficidin pathway using sequence motifs.

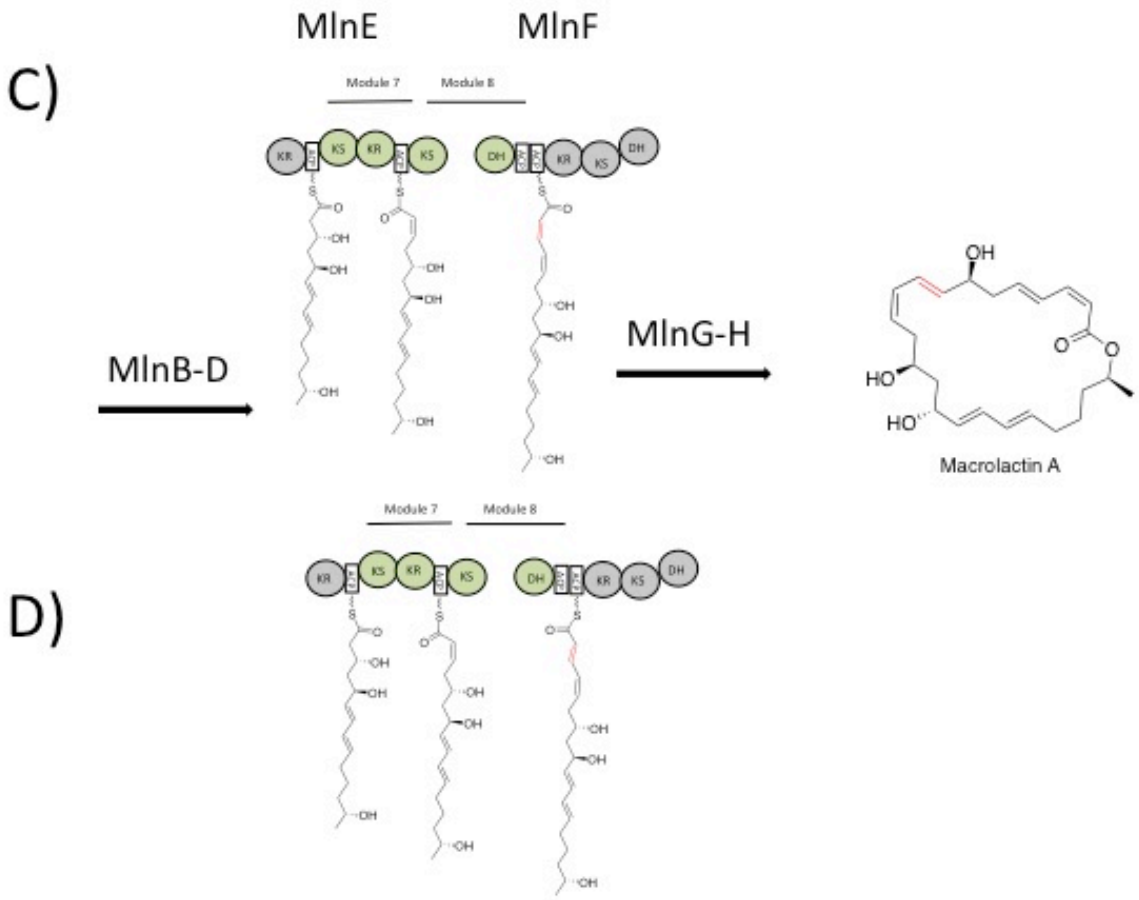


Figure 4.5 C) Current assignment of the macrolactin pathway D) Reassignment of the macrolactin pathway using sequence motifs. [2]

In addition to the reassignment of PKS intermediates, our bioinformatic analysis of split bimodules revealed an example of reassigning the domain ordering of an entire PKS based on these new sequence elements. Specifically, our efforts uncovered a polypeptide from the basiliskamide PKS, a pathway reported by Cichewicz and coworkers from the genome of *B. laterosporus* strain PE36 isolated from a feral pig nostril. Chichewicz and co-workers' proposed domain order appears to us to lack a DH domain in module one along with a product requiring 3 dehydrations when only 2 are

present (experimental section). Due to the presence of the split module KR sequence motifs, and the presence of an active C-terminal KS and N-terminal DH, we believe it more likely that the basiliskamide PKS type B bimodule perform a double dehydration in its terminal module to account for all dehydrations observed in the final product. This would explain why there is one fewer DH than one would expect from the sequence-structure colinear relationship, which was missed in the initial assignment of the biosynthetic pathway (experimental section). [51]

GENOME MINING APPLICATIONS OF KR SEQUENCE SIGNATURES

To test the predictive power of these KR sequence fingerprints and evaluate their prevalence within the bacterial kingdom, we performed a BLAST search using a 148 consensus sequence spanning the C-terminal catalytic domain of the KR. Several new protein sequences were returned from uncharacterized systems of under-studied bacteria including *Brevibacillus laterosporus* (PE36), *Paenibacillus taiwanensis*, and *Salinibacillus aidingensis* MSP4 (experimental section). The sequences encoding these split bimodule KRs were obtained and submitted for analysis using the protein homology algorithm, antiSMASH.[52–54] The resulting predicted domain order of these orphan sequences was indicative of a modular structure consistent with either type A or type B bimodules, characterized by a C-terminal KS or KS^0 and bridging N-terminal DH. Thus one might envision that the novel sequence features presented in this report might be useful diagnostic regions for software programs that aid in the identification of orphan gene clusters. This may be particularly useful in the identification of the double dehydrations of type B bimodules, as such DH activity deviates from canonical PKS colinearity.

CONCLUSIONS AND OUTLOOK

The recent crystallographic dissection of *trans*-AT PKSs has revealed a great deal about the complex, peculiar architecture seen at both the domain and modular level that further illustrates the divergent nature of these systems. The structure of PksKR2 contains several unprecedented structural elements and conserved sequence motifs that strongly correspond to the unique modular architectures of split bimodules. The near invariant inclusion of these KR sequence and structural elements in split bimodules suggests a role in mediating the docking of adjacent PKS polypeptides (possibly to the upstream DH), a process that remains a lingering mystery for the trans-AT systems. Additionally, we suggest that the fingerprints presented herein will be useful in the characterization of PKS gene clusters and their corresponding metabolites.

ACKNOWEDGEMENTS

Instrumentation and technical assistance for crystallographic work were provided by Dr. Art Monzingo and the Macromolecular Crystallography Facility, with financial support from the College of Natural Sciences, the Office of the Executive Vice President and Provost, and the Institute for Cellular and Molecular Biology at the University of Texas at Austin. The Advanced Light Source (ALS) is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under contract no. DE-AC02-05CH11231. We thank the NIH (GM106112) and the Welch Foundation (F-1712) for supporting this research (A.T.K.)

REFERENCES

[1] C. Hertweck, *Angew. Chem.* **2009**, *48*, 4688.

[2] J. Piel, *Nat. Prod. Rep.* **2010**, *27*, 996.

[3] C. Khosla, Y. Tang, A. Y. Chen, N. a Schnarr, D. E. Cane, *Annu. Rev. Biochem.* **2007**, *76*, 195.

[4] J. Moldenhauer, X.-H. Chen, R. Borriss, J. Piel, *Angew. Chem.* **2007**, *46*, 8195.

[5] A. K. El-Sayed, J. Hothersall, S. M. Cooper, E. Stephens, T. J. Simpson, C. M. Thomas, *Chem. Biol.* **2003**, *10*, 419.

[6] P. D. Straight, M. A. Fischbach, C. T. Walsh, D. Z. Rudner, R. Kolter, *Proc. Natl. Acad. Sci., U. S. A.* **2007**, *104*, 305.

[7] X.-H. Chen, J. Vater, J. Piel, P. Franke, R. Scholz, K. Schneider, A. Koumoutsi, G. Hitzeroth, N. Grammel, A. W. Strittmatter, et al., *J. Bacteriol.* **2006**, *188*, 4024.

[8] R. A. Butcher, F. C. Schroeder, M. A. Fischbach, P. D. Straight, R. Kolter, C. T. Walsh, J. Clardy, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 1506.

[9] R. Reid, M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D. V Santi, C. R. Hutchinson, R. McDaniel, *Biochemistry* **2003**, *42*, 72.

[10] P. Caffrey, *ChemBioChem* **2003**, *4*, 654.

[11] A. T. Keatinge-Clay, *Chem. Biol.* **2007**, *14*, 898–908.

[12] S. A. Bonnett, J. R. Whicher, K. Papireddy, G. Florova, J. L. Smith, K. A. Reynolds, *Chem. Biol.* **2013**, *20*, 772.

[13] A. Keatinge-Clay, *J. Mol. Biol.* **2008**, *384*, 941.

[14] D. L. Akey, J. R. Razelun, J. Tehranisa, D. H. Sherman, W. H. Gerwick, J. L. Smith, *Structure* **2010**, *18*, 94.

[15] J. Wu, K. Kinoshita, C. Khosla, D. E. Cane, *Biochemistry* **2004**, *43*, 16301.

[16] X. Guo, T. Liu, C. R. Valenzano, Z. Deng, D. E. Cane, *J. Am. Chem. Soc.* **2010**, *132*, 14694.

[17] C. R. Valenzano, Y.-O. You, A. Garg, A. Keatinge-Clay, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.* **2010**, *132*, 14697.

[18] D. Gay, Y.-O. You, A. Keatinge-Clay, D. E. Cane, *Biochemistry* **2013**, *52*, 8916.

[19] S. J. Kodumal, K. G. Patel, R. Reid, H. G. Menzella, M. Welch, D. V Santi, *Proc. Natl. Acad. Sci., U. S. A.* **2004**, *101*, 15573.

[20] P. Caffrey, *Chem. Biol.* **2005**, *12*, 1060.

[21] D. Janssen, D. Albert, R. Jansen, R. Müller, M. Kalesse, *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 4898.

[22] D. Menche, F. Arikan, O. Perlova, N. Horstmann, W. Ahlbrecht, S. C. Wenzel, R. Jansen, H. Irschik, R. Müller, *J. Am. Chem. Soc.* **2008**, *130*, 14234.

[23] K. Ishida, T. Lincke, C. Hertweck, *Angew. Chem.* **2012**, *51*, 5470.

[24] N. Moebius, C. Ross, K. Scherlach, B. Rohm, M. Roth, C. Hertweck, *Chem. Biol.* **2012**, *19*, 1164.

[25] S. K. Piasecki, J. Zheng, A. J. Axelrod, M. E Detelich, A. T. Keatinge-Clay, *Proteins* **2014**, *82*, 2067.

[26] J. Zheng, A. T. Keatinge-Clay, *MedChemComm* **2012**. *4*, 34.

[27] J. Zheng, S. K. Piasecki, A. T. Keatinge-Clay, *ACS Chem. Biol.* **2013**, *8*, 1964–1971.

[28] J. Zheng, C. A. Taylor, S. K. Piasecki, A. T. Keatinge-Clay, *Structure* **2010**, *18*, 913.

[29] A. T. Keatinge-Clay, R. M. Stroud, *Structure* **2006**, *14*, 737.

[30] S. A. Bonnett, J. R. Whicher, K. Papireddy, G. Florova, J. L. Smith, K. A. Reynolds, *Chem. Biol.* **2013**, *20*, 772.

[31] J. Zheng, A. T. Keatinge-Clay, *Med.Chem.Commun.* **2013**, *4*, 34.

[32] J. Kim, J. H. Chang, E.-J. Kim, K.-J. Kim, *Biochem. Biophys. Res. Commun.* **2014**, *443*, 783.

[33] J. Zheng, S. K. Piasecki, A. T. Keatinge-Clay, *ACS Chem. Biol.* **2013**, *8*, 1964.

[34] M. L. Mugnai, Y. Shi, A. T. Keatinge-Clay, R. Elber, *Biochemistry* **2015**, *54*, 2346.

[35] S. K. Piasecki, C. A. Taylor, J. F. Detelich, J. Liu, J. Zheng, A. Komsoukaniants, D. R. Siegel, A. T. Keatinge-Clay, *Chem. Biol.* **2011**, *18*, 1331.

[36] M. Häckh, M. Müller, S. Lüdeke, *Chemistry* **2013**, *19*, 8922.

[37] S. K. Piasecki, C. A. Taylor, J. F. Detelich, J. Liu, J. Zheng, A. Komsoukaniants, D. R. Siegel, A. T. Keatinge-Clay, *Chem. Biol.* **2011**, *18*, 1331.

[38] V. Simunovic, R. Müller, *ChemBioChem* **2007**, *8*, 497.

[39] C. Zhao, J. M. Coughlin, J. Ju, D. Zhu, E. Wendt-Pienkowski, X. Zhou, Z. Wang, B. Shen, Z. Deng, *J. Biol. Chem.* **2010**, *285*, 20097.

[40] T. W. Geders, L. Gu, J. C. Mowers, H. Liu, W. H. Gerwick, K. Håkansson, D. H. Sherman, J. L. Smith, *J. Biol. Chem.* **2007**, *282*, 35954.

[41] C. T. Calderone, W. E. Kowtoniuk, N. L. Kelleher, C. T. Walsh, P. C. Dorrestein, *Proc. Natl. Acad. Sci., U. S. A.* **2006**, *103*, 8977.

[42] C. T. Calderone, D. F. Iwig, P. C. Dorrestein, N. L. Kelleher, C. T. Walsh, *Chem. Biol.* **2007**, *14*, 835.

[43] P. Pöplau, S. Frank, B. I. Morinaka, J. Piel, *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 13215.

[44] O. Perlova, K. Gerth, O. Kaiser, A. Hans, R. Müller, *J. Biotechnol.* **2006**, *121*, 174.

[45] K. J. Weissman, *ChemBioChem* **2006**, *7*, 485.

[46] K. J. Weissman, *ChemBioChem* **2006**, *7*, 1334.

[47] J. Yan, S. Gupta, D. H. Sherman, K. A. Reynolds, *ChemBioChem* **2009**, *10*, 1537.

[48] K. E. Wilson, J. E. Flor, R. E. Schwartz, H. Joshua, J. L. Smith, B. A. Pelak, J. M. Liesch, O. D. Hensens, *J. Antibiot. (Tokyo).* **1987**, *40*, 1682–91.

[49] S. B. Zimmerman, C. D. Schwartz, R. L. Monaghan, B. A. Pelak, B. Weissberger, E. C. Gilfillan, S. Mochales, S. Hernandez, S. A. Currie, E. Tejera, *J. Antibiot. (Tokyo).* **1987**, *40*, 1677.

[50] K. Schneider, X.-H. Chen, J. Vater, P. Franke, G. Nicholson, R. Borriss, R. D. Süssmuth, *J. Nat. Prod.* **2007**, *70*, 1417.

[51] T. Weber, K. J. Laiple, E. K. Pross, A. Textor, S. Grond, K. Welzel, S. Pelzer, A. Vente, W. Wohlleben, *Chem. Biol.* **2008**, *15*, 175.

[52] J. Piel, *Nat. Prod. Rep.* **2010**, *27*, 996.

[53] C. M. Theodore, B. W. Stamps, J. B. King, L. S. L. Price, D. R. Powell, B. S. Stevenson, R. H. Cichewicz, *PLoS One* **2014**, *9*, e90124.

[54] K. Blin, M. H. Medema, D. Kazempour, M. A. Fischbach, R. Breitling, E. Takano, T. Weber, *Nucl. Acids Res.* **2013**, *41*, W204.

[55] M. H. Medema, K. Blin, P. Cimermancic, V. de Jager, P. Zakrzewski, M. A. Fischbach, T. Weber, E. Takano, R. Breitling, *Nucl. Acids Res.* **2011**, *39*, W339.

[56] T. Weber, K. Blin, S. Duddela, D. Krug, H. U. Kim, R. Bruccoleri, S. Y. Lee, M. A. Fischbach, R. Muller, W. Wohlleben, et al., *Nucl. Acids Res.* **2015**, *43*, W237.

Experimental Section for Chapter 4

Clonging, Protein Expression, and Purification

The DNA encoding PKSKR2 (from *pksJ*, accession number NP_389598) was amplified from *B. subtilis str. 168* gDNA with primers:

5'- GCGGCCTGGTGCCGCGCGGCTCTAGCTCAGAAAGAGACAAAAAAGAACTG -3'

and

5'- GTGGTGGTGGTGGTGGTGATGTTATCCATCACATTTGGAAACTTC-3' (black corresponds to the annealing region, and red corresponds to the junction for ligation independent cloning) and cloned into expression plasmid pGAY28b.[1] The plasmid was then transformed into *E. coli* BL21(DE3). The cells were grown to an OD_{600} of 0.5 in Luria broth containing 50 mg/L kanamycin at 37 °C. The temperature was then dropped to 15 °C prior to inducing protein expression with 0.5 mM IPTG, and grown for an additional 16 hr. Cells were collected via centrifugation at 4,000 x g for 15 min, resuspended in lysis buffer (400 mM NaCl, 50 mM HEPES pH 7.5, 10% glycerol), and lysed by sonication. Cellular debris was removed by centrifugation at 30,000 x g for 30 min and the cell lysate was then poured over Ni-NTA resin (Qiagen) pre-equilibrated with lysis buffer. Bound protein was washed in 40 ml of lysis buffer containing 15 mM imidazole and eluted with lysis buffer containing 150 mM imidazole. An equilibrated Superdex 200 gel filtration column was used to buffer exchange the protein solution into 150 mM NaCl and 10 mM HEPES pH 7.5 and the protein was concentrated to 11 mg/mL prior to crystallization trials.

Ketoreductase Stereochemical Assays

Ketoreduction assays were modified from a method described previously,[2,3] with 5 mM substrate **3.1**, 125 mM HEPES pH 7.5, 250 mM NaCl, 200 mM D-glucose, 10%

v/v glycerol, 200 μM $NADP^+$, 1 μM glucose dehydrogenase (from *B. subtilis*), and 5 μM ketoreductase to a total volume of 500 μL The reactions were incubated overnight at room temperature (25 °C), extracted with ethyl acetate (3 x 500 μL) and dried in a speedvac. All samples were re-suspended in ethanol prior to chromatographic analysis. Chiral chromatography was performed with a ChiraCel OC-H column (250 x 4.6 mm) with a Beckman Coulter System Gold 126 pump and a System Gold 166 PDA detector equipped with a 20 μL loop. Absorbance was monitored at 235 nm. The solvent system used was 7% ethanol in hexanes (measured using volumetric flasks) at 0.8 mL/min. Substrate **3.1** was synthesized as described previously.[4] Enantiomeric excess was determined using peak area integrations.

CRYSTALLIZATION AND STRUCTURE DETERMINATION

Crystals of PksKR2 grew in 2 days by sitting drop vapor diffusion at 22 °C. Drops were formed by mixing 2 μL protein solution (9 mg/mL PksKR2, 150 mM NaCl, 10 mM HEPES, pH 7.5) with 1 μL crystallization buffer (sodium citrate, 0.1 M HEPES pH 7.0). Crystals were frozen in liquid nitrogen after a 20-min soak in the crystallization buffer modified with 10% (v/v) ethylene glycol. Diffraction data, collected at ALS Beamline 5.0.3, were processed by HKL2000.[5] The structure was solved to 1.98 Å resolution by molecular replacement with PhaserMR[6] in the CCP4 suite,[7] using the KR1 monomer from the PksX synthase of *B. subtilis* (PDB code: 4J1Q)[3] as the search model. The model generated from the molecular replacement solution was used to iteratively build into the remaining electron density map with Coot[8] and was refined with Refmac5.[9]

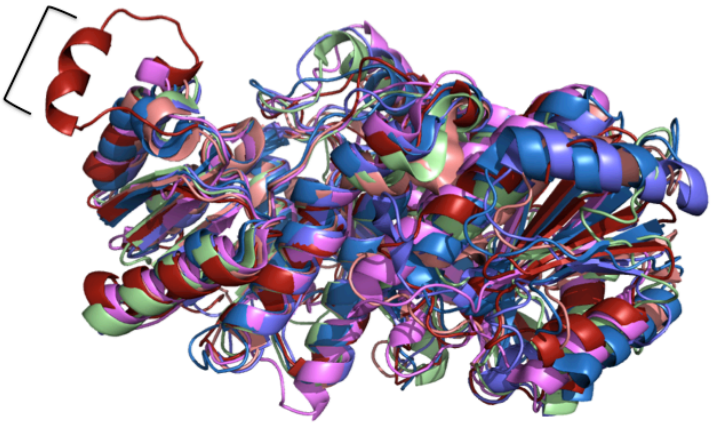


Figure E4.1 Structural overlay of PKSJKR2 (red), PKSJKR1 (dark purple; pdb code 4j1q), PlmKR1 (green; pdb code 4hxy), TylKR1 (light pink; pdb code 2fr1), EryKR1 (blue; 2z5L), and AmpKR2 (magenta; 3mjv). The insertion helix unique to A' KRs is indicated by brackets.

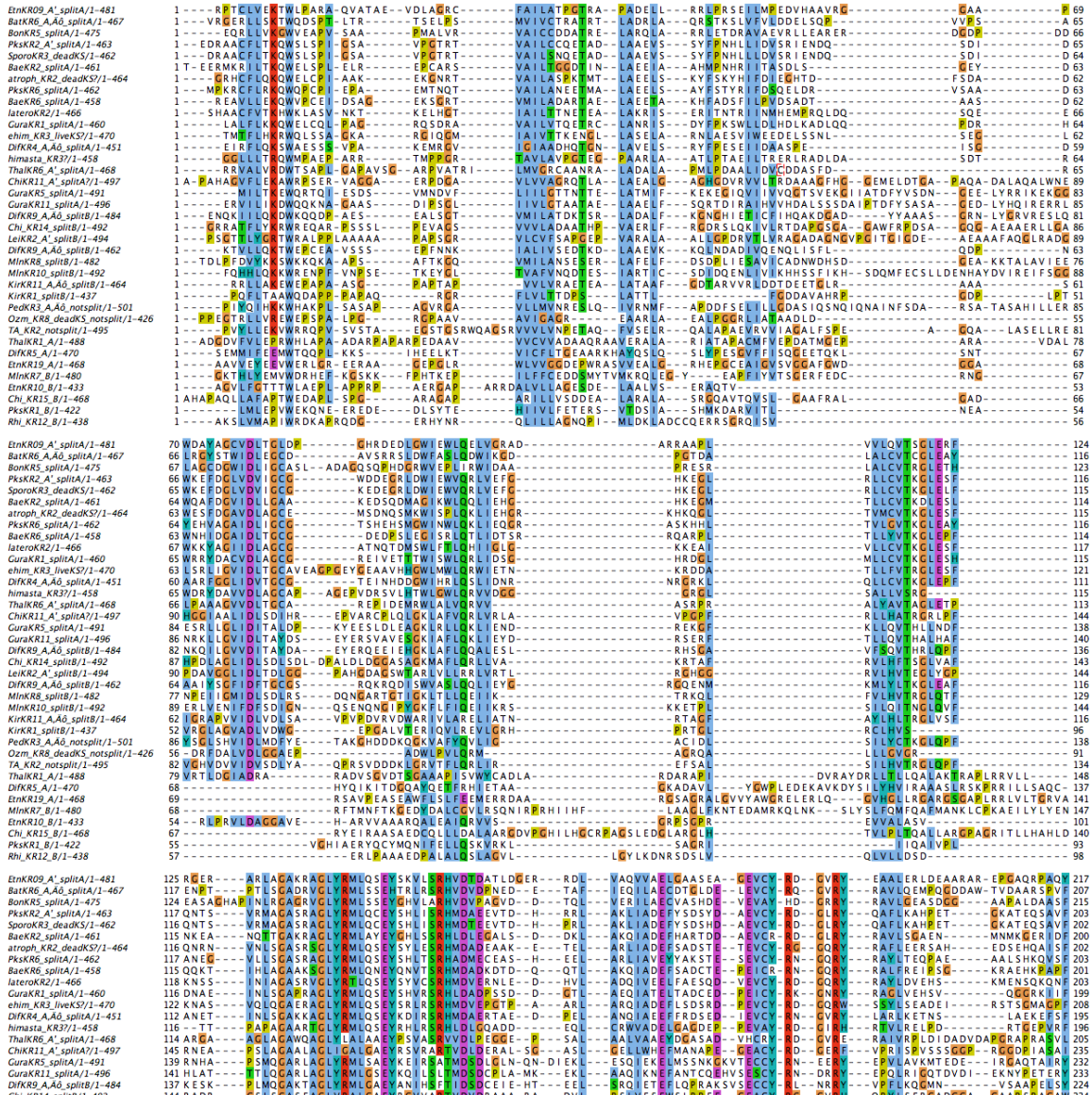


Figure E4.2

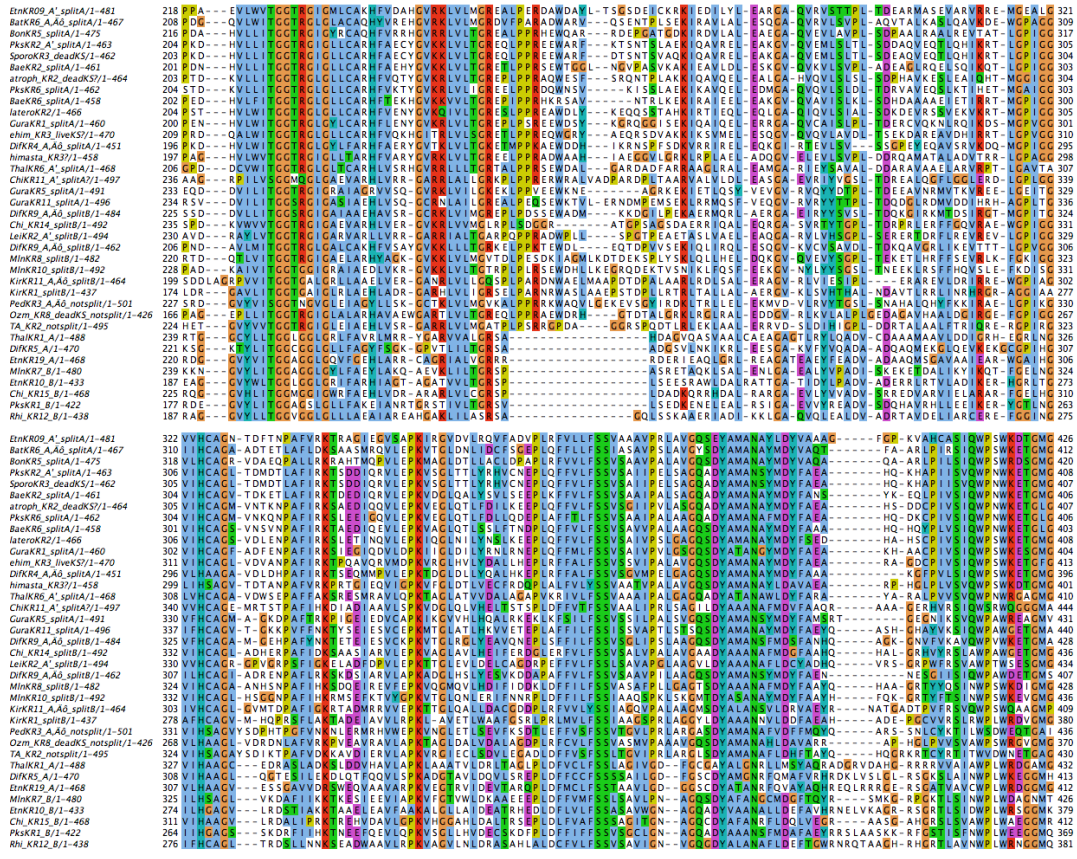


Figure E4.2 Multiple Sequence Alignment of KR domains from split bimodules. SplitA and splitB indicates type A and type B bimodule KRs respectively. A or B indicates A type or B type ketoreduction respectively. Etn = etnangien PKS, Bat = batumin PKS, Bon = bongkrecic acid PKS, Pks = bacillaene PKS (*B. subtilis*), Bae = bacillaene PKS (*B. amyloliquefaciens)*, Dif = difficidin PKS, Ozm = oxazolomycin PKS, Rhi = rhizoxin PKS, Mln = macrolatin PKS, Lei = leinamycin PKS, Ped = pederin PKS, TA = antibiotic TA PKS, Thal = thailandamide PKS, Kir = kirromycin PKS, Chi = chivosazol PKS,

Gura = Geobacter uraniumreducens polyketide PKS. Sporo, atroph, latero, ehim & himasta represent uncharacterized synthases.

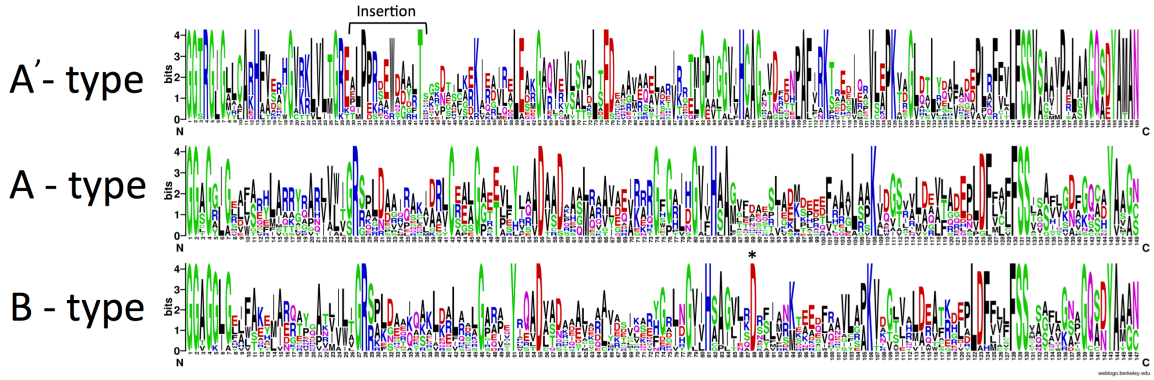


Figure E4.3 Sequence LOGOS of catalytic subdomains show sequence dissimilarity between A' KRs and all other trans-AT keteoreductases. Sequence logos for PksKR2's A type subclass (A'), regular A type (A) and B type *trans*-AT KRs. Logos comprise KR domains from the bacillaene, bongkrekic acid, chivosazol, difficidin, etnangien, oxazolomycin, rhizoxin and thailandamide synthases. The asterisk indicates conserved aspartate fingerprint associated with B type reduction.

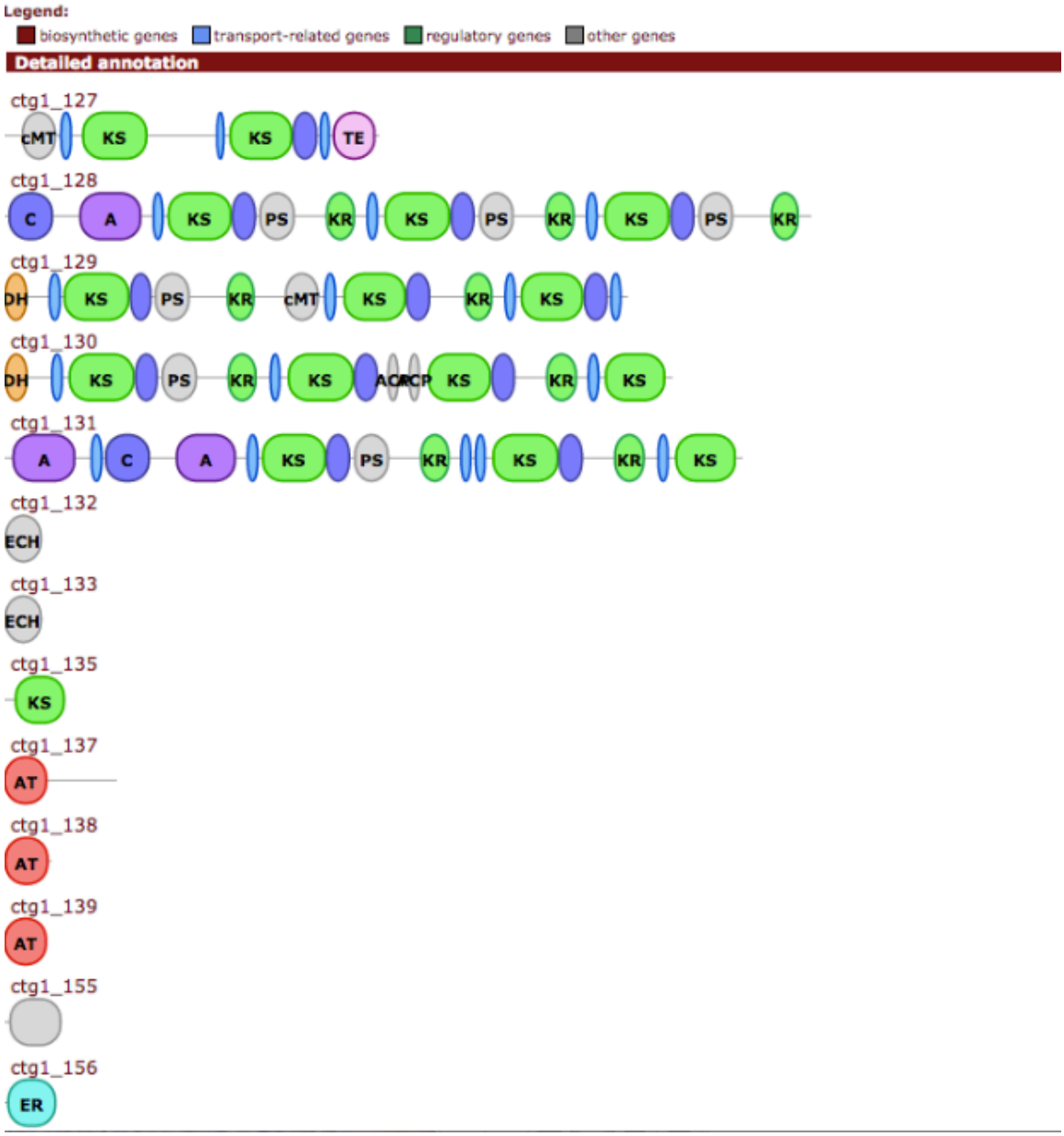


Figure E4.4 Orphan cluster from *Salinibacillus aidingensis* MSP4 (NCBI accession code: APIS01000007) anti-SMASH output.[10–12] The KR present in ctg_1_131 is an A' type KR. The KS present in ctg1_131 has sequence motifs indicating that it is condensation-incompetent (the HGTG motif is absent).[13] Taken together, this indicates that it is a type A bimodule (no double dehydration occurs).

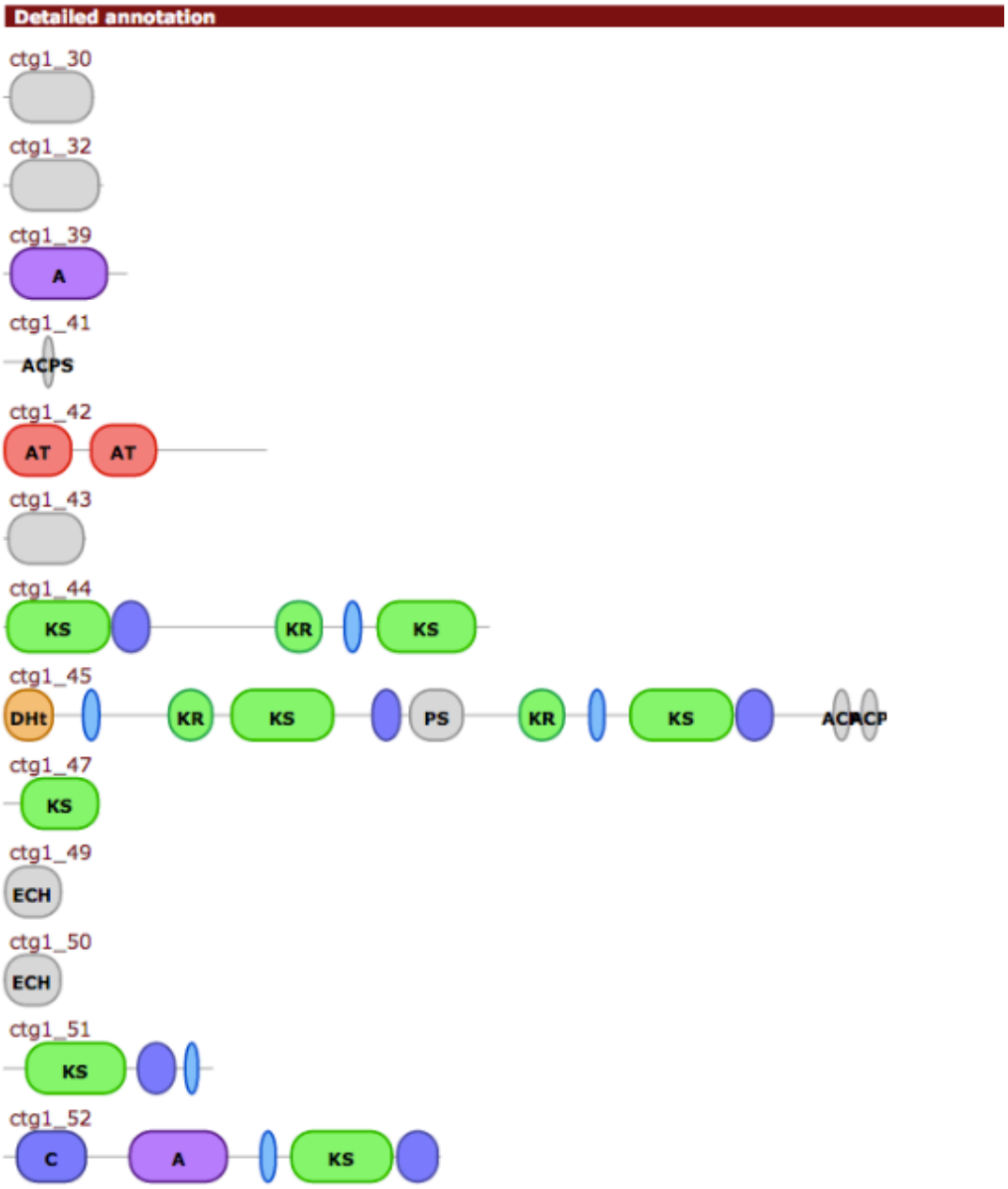
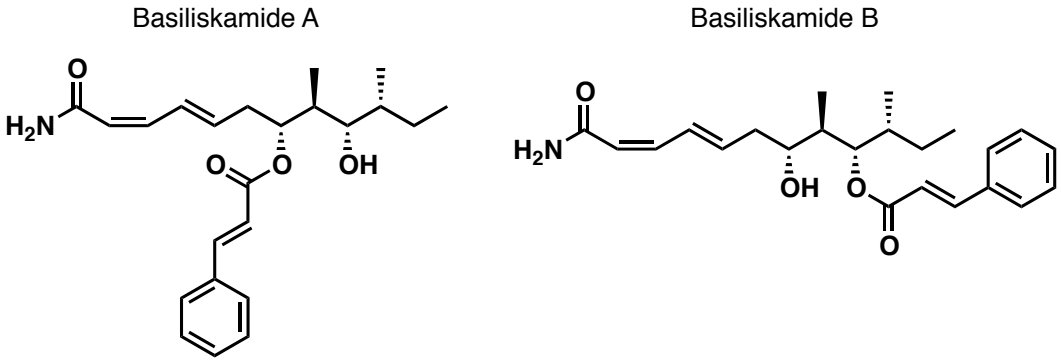


Figure E4.5 Orphan cluster from *Paenibacillus taiwanensis* (NCBI accession code: KE384307) anti-SMASH output. The KR present in ctg1_44 is an A' KR type, and the KS present has sequence motifs indicating that it is condensation-competent (the HGTG motif is present).[13] The KR in ctg 1_45 is a B-type KR by sequence. Taken together, this indicates that a type B bimodule is present, and that a double dehydration should be assigned to this module.

A.



B.

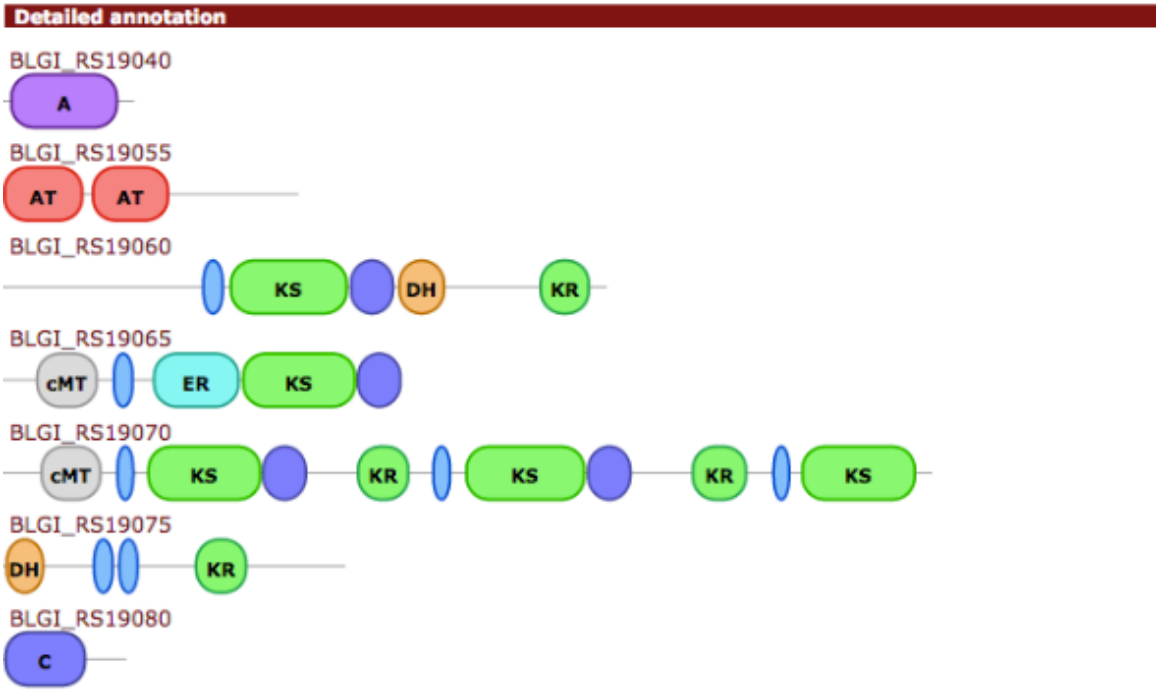


Figure E4.6 A) Structures of basiliskamides A and B. **B)** anti-SMASH output of the cluster.

C.

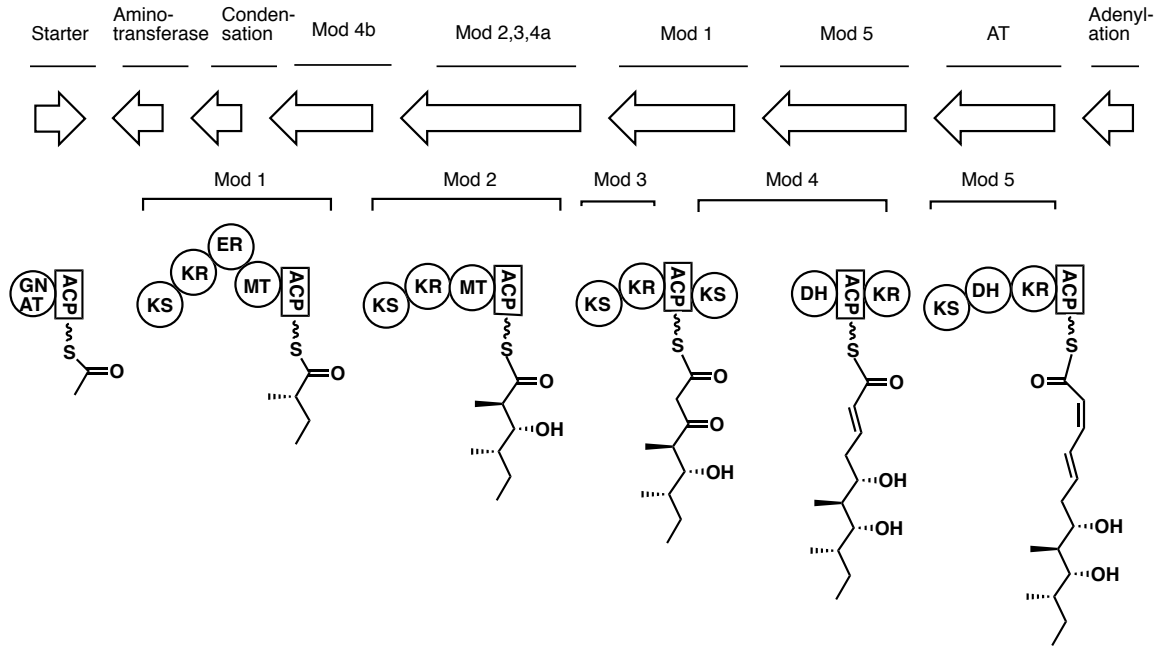


Figure E4.6 C) Anti-SMASH interpretation suggested by Chichewitz and co-workers, based on a collinear interpretation of DH activity.

D.

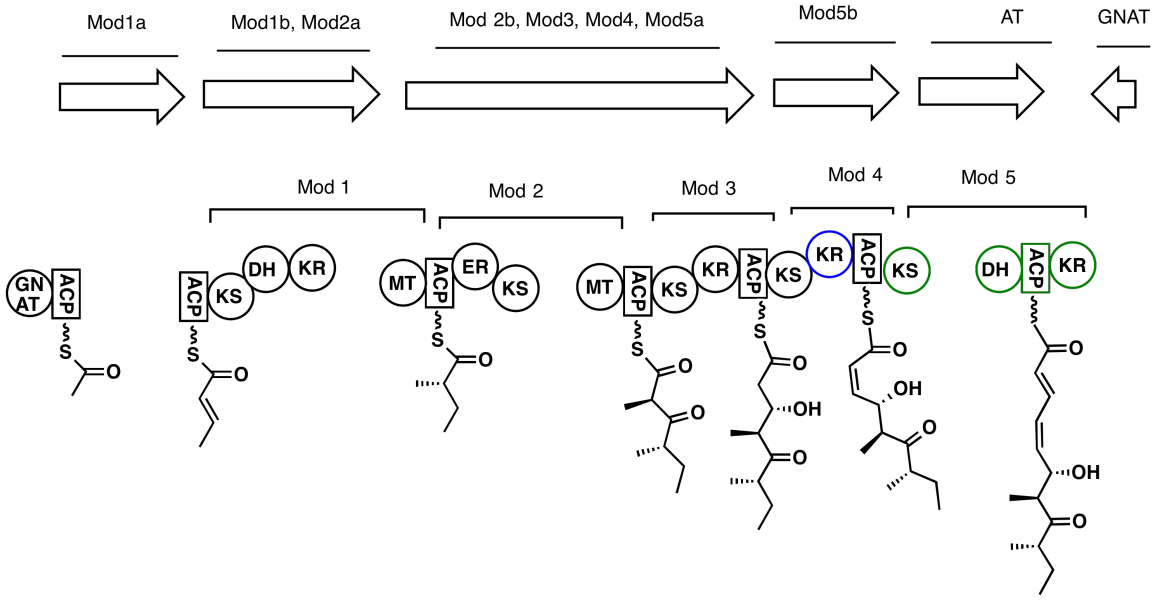


Figure E4.6 Revised analysis of the anti-SMASH analysis of Brevibacillus laterosporus (NCBI accession code NZ_CAGD00000000) GI-9 type B bimodule (PKS genes: WP_035319071.1,WP_035319075.1,WP_035319073.1,WP_003344916.1), suggested by Chichewiz and coworkers[14] to be the putative biogenic origins of basiliskamides and B. **D)** Revised anti-SMASH interpretation based on the presence of the double dehydrating type B bimodule (highlighted in green).

Corroborating our assignment is the presence of an A' type KR in Mod 4, as well as a B-type KR in Mod5, suggesting *cis* followed by *trans* dehydration (**Figure E4.6**). The active KS of the split Mod5 is also diagnostic of a double-dehydrating type B bimodule. Additional features that we believe are supportive of our assignment are a) that the genes of the operon are not scrambled in terms of modular order, b) and DH activity from module 1 (which is necessary for ER activity) is missing in the previous assignment. Indeed, this pathway has three putative dehydrations, but only two DHs present, which is further evidence for a double dehydrating DH. Our assignment has a

couple features that are not accounted for: the final metabolites' structures suggest a ketoreduction step in what would putatively occur module 2. We believe this is explained either via non-colinear action of KR3 (which is an A-type KR, consistent with the structure of the final metabolite), or by a downstream tailoring enzyme. Consistent with the latter hypothesis, an uncharacterized NADPH utilizing oxidoreductase appears in the cluster after the PKS. The other feature is that the double bond geometry deviates from the final metabolite, possibly also explained by downstream tailoring enzymes.

REFERENCES

[1] G. Gay, D. T. Wagner, A. T. Keatinge-Clay, D. C. Gay, *Plasmid* **2014**, *76C*, 66.

[2] S. K. Piasecki, C. A. Taylor, J. F. Detelich, J. Liu, J. Zheng, A. Komsoukaniants, D. R. Siegel, A. T. Keatinge-Clay, *Chem. Biol.* **2011**, *18*, 1331.

[3] S. K. Piasecki, J. Zheng, A. J. Axelrod, M. E Detelich, A. T. Keatinge-Clay, *Proteins* **2014**, *82*, 2067.

[4] S. K. Piasecki, C. A. Taylor, J. F. Detelich, J. Liu, J. Zheng, A. Komsoukaniants, D. R. Siegel, A. T. Keatinge-Clay, *Chem. Biol.* **2011**, *18*, 1331.

[5] Z. Otwinowski, W. Minor, *Methods Enzymol.* **1997**, *276*, 307.

[6] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Crystallogr.* **2007**, *40*, 658.

[7] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, et al., *Acta Crystallogr. D. Biol. Crystallogr.* **2011**, *67*, 235.

[8] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr. D. Biol. Crystallogr.* **2010**, *66*, 486–501.

[9] G. N. Murshudov, A. A. Vagin, E. J. Dodson, *Acta Crystallogr. D. Biol. Crystallogr.* **1997**, *53*, 240.

[10] K. Blin, M. H. Medema, D. Kazempour, M. A. Fischbach, R. Breitling, E. Takano, T. Weber, *Nucl. Acids Res.* **2013**, *41*, W204.

[11] T. Weber, K. Blin, S. Duddela, D. Krug, H. U. Kim, R. Bruccoleri, S. Y. Lee, M. A. Fischbach, R. Muller, W. Wohlleben, et al., *Nucl. Acids Res.* **2015**, *43*, W237.

[12] M. H. Medema, K. Blin, P. Cimermancic, V. de Jager, P. Zakrzewski, M. A. Fischbach, T. Weber, E. Takano, R. Breitling, *Nucl. Acids Res.* **2011**, *39*, W339.

[13] T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck, J. Piel, *Nat. Biotechnol.* **2008**, *26*, 225.

[14] C. M. Theodore, B. W. Stamps, J. B. King, L. S. L. Price, D. R. Powell, B. S. Stevenson, R. H. Cichewicz, *PLoS One*. **2014**, *9*, e90124.

Chapter 5: Mechanically Modulating the Photophysical Properties of Fluorescent Proteins Using Mechanical Force[xi]

INTRODUCTION

Polymer mechanochemistry[1–4] is a rapidly growing field of study wherein mechanical energy is harnessed to drive useful chemical transformations,[5–10] many of which are otherwise inaccessible. Apart from their fundamental interest, mechanochemical phenomena can be applied toward the development of novel stress-sensing materials with the capacity to report damage quantitatively and qualitatively. Salient examples of mechanically facilitated transformations that have been exploited within such materials include the electrocyclic ring opening of spiropyran derivatives,[11–13] formal [4+2] cycloreversions of anthracene derived Diels-Alder adducts,[14,15] and formal [2+2] cycloreversions of 1,2-dioxetanes.[16] Collectively, these systems report stress either through mechanochromism[11–13] or mechanoluminscence,[14,15] which enables quantification of mechanical damage using standard optical spectroscopies. Unfortunately, such stimulus responsive materials typically require tedious chemical syntheses; consequently, tuning their mechanochemical reactivity (*e.g.*, through chemical diversification of mechanically labile scaffolds) can present a significant impediment to the development of new force-responsive sensors.

Recently, attention has been directed toward harnessing the biosynthetic machinery of living organisms to access mechanically sensitive biomolecules (*i.e.*,

---

[xi] Portions of this chapter were reproduced from: Brantley, J. N.; **Bailey, C. B**.; Cannon, J. R.; Clark, K. A.; Vanden Bout, D. A.; Brodbelt, J. S.; Keatinge-Clay, A. T.; Bielawski, C. W. *Angew. Chem. Int. Ed.* 2014, 20, 5188. DOI: 10.1002/anie.201306988. JNB and CBB prepared and tested the protein composites. JRC and JSB performed the mass spectrometry studies. KAC and DAVB assisted with fluorescence microscopy. ATKC and CWB helped design and evaluate the experimental results. All authors contributed to the writing of the original text and figure preparation. The majority of the text was written by JNB, CBB, AKC, and CWB.

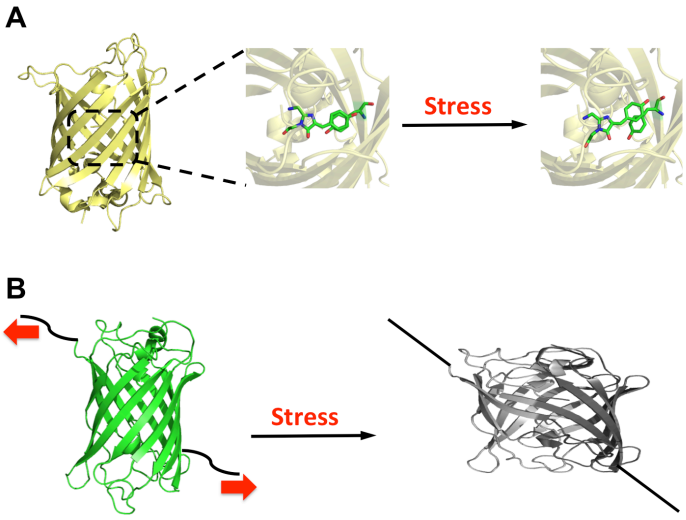
"biomechanophores").[18] While nature is replete with examples of force responsive systems,[20] there is a relative dearth of reports wherein biomolecules are used for mechanochemical applications. Indeed, although the modulation of enzymatic activity through mechanical stress has been reported,[18–21] few efforts have been directed toward developing biomechanophores that report mechanical stress through optical output. We envisioned that polymeric materials containing fluorescent proteins could serve as useful classes of stress-sensing biocomposites. Fluorescent proteins,[22] which are ubiquitous within the purview of the biochemical sciences, can be modified *via* site-selective mutagenesis[24] to precisely alter their structural and photophysical properties. Additionally, fluorescent proteins have been extensively optimized to achieve high stability and high levels of recombinant overexpression.

The photophysical properties of the canonical fluorescent protein, green fluorescent protein (GFP), stem from a 4-(*p*-hydroxybenzylidene)imidazolidin-5-one chromophore located in the center of the protein's β-barrel structure.[22] Genetic mutations that alter the structure of the chromophore (as in the case of cyan fluorescent protein or blue fluorescent protein) or local residues that impact the stereoelectronic environment surrounding the chromophore (as in the case of yellow fluorescent protein) give rise to a vibrant array of proteins with unique emissive properties.[22] The fluorescence of all photoemissive protein variants is highly dependent on proper folding of the protein;[22,24,25] as such, mechanical perturbation of the β-barrel structure results in modulation of any associated photophysical properties.[26–28] Yellow fluorescent protein (YFP) is particularly attractive for use as a biomechanophore, as the yellow fluorescence results from a weak arene interaction between the chromophore and tyrosine 203 (mutated from threonine in the parent GFP).[25] Gruner and colleagues reported that pressurized crystals of the YFP variant, citrine, exhibited a gradual hypsochromic shift in fluorescence as the pressure

was increased from 0 to 360 MPa at low temperatures (77 K).[29,30] The fluorescence was subsequently found to bathochromically shift upon reducing the pressure and warming the crystals to 180 K. While this work clearly revealed the potential to develop YFP as a mechanosensor, there have been no reports to date wherein an analogous modulation of YFP fluorescence was harnessed for applications in mechanically responsive materials. In addition, surprisingly few efforts have been directed toward developing stress-sensing materials that employ fluorescent proteins. For example, Clark and colleagues utilized Förster resonance energy transfer (FRET) between YFP and cyan fluorescent protein (CFP) to develop stress reporting poly(acrylamide) composites.[30,32] Stretching these materials under uniaxial strain resulted in increased FRET interactions between YFP and CFP near micro-cracks that formed within the material, as determined by fluorescence confocal microscopy and fluorescence lifetime imaging (FLIM).[32] Bruns and colleagues more recently reported that eYFP could serve as a mechanically sensitive link between glass substrates and epoxy resins, where delamination of the resin resulted in denaturation of the protein and subsequent fluorescence quenching.[33] While these examples elegantly demonstrated that fluorescent proteins could be adapted for applications in stress reporting, there have been no reports showcasing diverse and tunable mechanochemical responses from biocomposite materials containing fluorescent proteins, which are features that are expected to be valuable for the design of precisely tailored force-sensing materials. Here, we report the facile preparation of biocomposite materials containing either: 1) an enhanced YFP (eYFP[34]) that exhibits shifts in λ_{em} under mechanical stress, or 2) a genetically modified GFP (GFPuv[35,36]) that exhibits fluorescence quenching under the action of mechanical force.

MECHANICAL ACTIVATION OF EYFP BIOCOMPOSITES

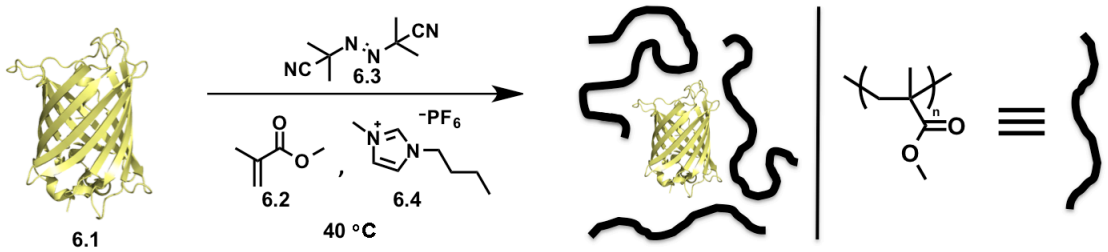
As eYFP was predicted to exhibit greater mechanical sensitivity than GFPuv,[26] our initial efforts were directed toward the development of eYFP-containing biocomposites. We reasoned that embedding eYFP within a polymeric matrix and subjecting the resulting material to bulk compression would elicit the desired photophysical modulation, as local areas of high pressure generated during material compression could disrupt the arene interaction responsible for yellow fluorescence *via* subtle distortions of the protein's structure (**Scheme 5.1**).



Scheme 5.1 Schematic representations of proposed mechanical activations of fluorescent proteins. (A) Compression of composite materials containing eYFP distorts the arene interaction between the chromophore and tyrosine 203. (B) The incorporation of cysteine residues at strategic sites in GFPuv facilitates the covalent attachment of polymer chains to the protein; subsequent compression of the composite mechanically denatures GFPuv and quenches the protein's fluorescence.

To test the aforementioned hypothesis, we overexpressed hexahistidine-tagged eYFP in *E. coli* BL21 (DE3) and subsequently purified the isolated protein by nickel

affinity chromatography. As shown in **Scheme 5.2**, poly(methyl methacrylate) (PMMA) composites were prepared by adding eYFP (**5.2**) directly to a mixture of methyl methacrylate (**A.2**), azobisisobutyronitrile (AIBN, **5.3**), and the plasticizer, 1-butyl-3-methylimidazolium hexafluorophosphate (BMIM-PF_6, **5.4**), at 40 °C (see experimental section for additional details). The stability of eYFP under these relatively harsh conditions was remarkable, and the composite materials isolated following consumption of the free monomer exhibited strong fluorescence (λ_{ex} = 485 nm; λ_{em} = 540 nm; Φ = 0.64) and relatively uniform protein distribution. Although significant denaturation of eYFP was observed upon dissolution of the isolated biocomposites in tetrahydrofuran (as evidenced by fluorescence quenching), this obstacle to material processing was circumvented by cutting and polishing the composites to afford specimens with defined geometries. The BMIM-PF_6 additive, which is known to serve as a highly effective plasticizer of acrylate derived polymers,[37] allowed precise modulation of the physical properties exhibited by the composites. Specifically, the addition of BMIM-PF_6 enabled the glass transition (T_g) of PMMA to be reduced to approximately 40 °C (as determined by differential scanning calorimetry), which was found to be beneficial for sample processing.



Scheme 5.5 Synthesis of eYFP-containing biocomposites. General conditions: eYFP (**5.1**; 1.0 equiv), MMA (**5.2**; 2.6 x 10^5 equiv), AIBN (**5.3**; 1.3 x 10^3 equiv), and BMIM-PF_6 (**5.4**; 3.1 x 10^4 equiv) were combined in a single vessel under N_2 and heated to 40 °C.

To test the mechanical sensitivity of our composite materials, a 50 mg sample was mounted in a hydraulic press and subjected to compression at incrementally increasing pressures (0 – 360 MPa) for periods of 45 s, after which time the solid state fluorescence was measured. As shown in **Figure 5.1** (left), the λ_{em} of the sample gradually shifted from 539 nm at 0 MPa to approximately 534 nm at 360 MPa (a hypsochromic shift commensurate with that previously reported by Gruner[30]). Frictional heating during compression appeared to contribute to the overall response of the material through thermal denaturation of the protein (as evidenced by a reduction in fluorescence intensity; see experimental section for additional details). Importantly, though, the observed change in λ_{em} correlated monotonically with the applied force and was, thus, consistent with a mechanical process.[1-4] Compressing the composites for 1 h did not cause their λ_{em} to shift beyond what was measured after compression for 45 s at the same pressure. Collectively, these data suggested to us that mechanical forces generated upon compressing the composites were inducing subtle distortions of the protein's chromophore.

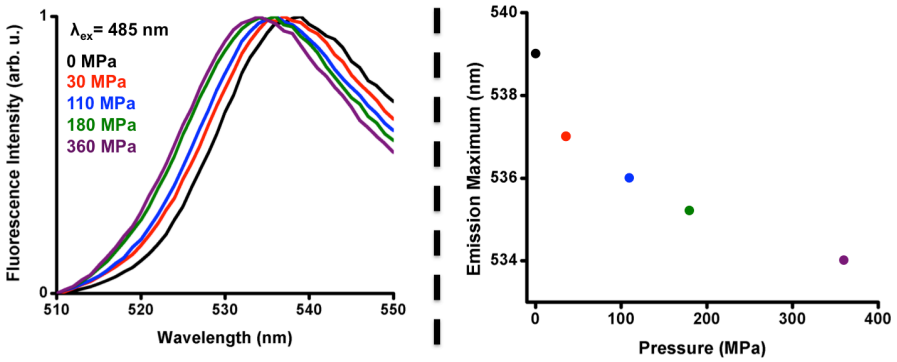
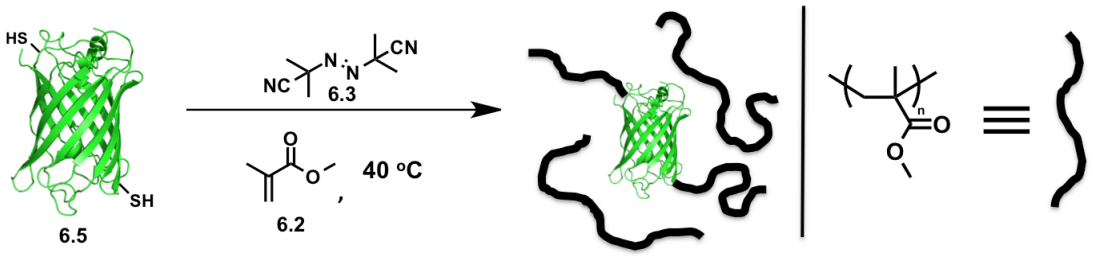


Figure 5.1 (Left) Compression of PMMA composites containing eYFP caused the λ_{em} to gradually undergo a hypsochromic shift. Normalized fluorescence intensities are shown. (Right) The fluorescence maxima of the compressed PMMA composite containing eYFP plotted as a function of applied pressure.

MECHANICAL ACTIVATION OF GFPUV BIOCOMPOSITES

Having established a straightforward preparation of ratiometric stress sensors, we sought to explore the mechanochromism of another fluorescent protein to realize an intensiometric stress reporter. Such reporters are particularly valuable because they exhibit changes in their optical properties as a function of applied load that facilitate rapid assessment of mechanical damage. As mechanical unfolding of GFP has previously been shown to quench the protein's fluorescence,[26] efforts were directed toward expanding the results of these atomic force microscope (AFM) pulling experiments to bulk materials. While the mechanical response of eYFP presumably resulted from the distortion of a weak, local interaction, we reasoned that modulation of GFP's photophysical properties would involve more global phenomena (*i.e.*, complete denaturing of the protein); thus, we surmised that significantly higher forces would be required to mechanically denature the protein. As such, we hypothesized that GFP would need to be chemically cross-linked to the polymer matrix to sufficiently harness the mechanical forces generated during compression to achieve the desired fluorescence

quenching. Guided by the work of Dietz, Rief, and Lorimer,[26] which revealed that the *N*-terminal β-sheet in GFP is mechanically labile, we concluded that polymeric appendages should be introduced on opposing sides of the aforementioned β-sheet in order to direct mechanical forces to this putative "Achilles' heel" (*i.e.*, the most mechanically labile structural element within the protein). The strategic incorporation of cysteine residues within the polypeptide backbone was predicted to facilitate the desired polymer ligation, as Bowman and Cramer have shown that thiyl radicals (which can be generated from thiols under free radical polymerization conditions) react efficiently with propagating acrylate radicals (provided the initial thiol concentration is relatively low).[38]



Scheme 5.3 Synthesis of mechanically active GFPuv-containing biocomposites. General conditions: GFPuv(Y39C/D103C) (**A.5**; 1.0 equiv), MMA (**A.2**; 2.6 x 10^5 equiv), and AIBN (**A.3**; 1.3 x 10^3 equiv) were combined in a single vessel under N_2 and heated to 40 °C.

Site-directed mutagenesis was performed to incorporate cysteine residues on opposing sides of the β-barrel (replacing tyrosine 39 and aspartate 103) in GFPuv, as the attachment of polymer chains at these sites could direct mechanical forces to the aforementioned β-sheet and induce mechanical denaturation (and, consequently, fluorescence quenching). GFPuv was selected due to its high stability in bacterial expression systems and its stronger fluorescence signal than wild-type GFP.[34,36] The

resultant hexahistidine-tagged GFPuv(Y39C/D103C) double mutant was overexpressed in *E*. coli BL21 (DE3) and purified by nickel affinity chromatography. As shown in **Scheme 5.3**, PMMA composites were prepared by polymerizing methyl methacrylate (**5.2**) in the presence of GFPuv(Y39C/D103C) (**5.5**; see experimental section for additional details). Gel-permeation chromatography (GPC) visualized with ultraviolet-visible detection at 280 nm (a λ_{max} of tyrosine) revealed that the resulting polymeric material displayed an increased absorbance at this wavelength relative to a PMMA homopolymer that was prepared in the absence of GFPuv(Y39C/D103C) at the same concentration (see experimental section for additional details). Moreover, mass spectrometry studies revealed that GFPuv(Y39C/D103C) was coupled to methyl methacrylate under the polymerization conditions, which confirmed that the cysteine residues were solvent exposed. Taken together, these results were consistent with the covalent attachment of GFPuv(Y39C/D103C) to the growing polymer chains during the preparation of the aforementioned composites. As shown in **Figure 5.2**, the solid state fluorescence of the composites (λ_{em} = 507 nm; λ_{ex} = 420 nm) was in agreement with the successful incorporation of GFPuv(Y39C/D103C) into the polymeric matrix. Compression of these materials in a hydraulic press (41 MPa; 45 s) resulted in a significant reduction in fluorescence intensity (**Figure 5.2A**). Moreover, the fluorescence intensity was found to decrease monotonically with increasing pressure (0 – 41 MPa), which is a hallmark of mechanical phenomena (**Figure 5.2A**).[1-4] Compression of the materials for 1 h did not cause their fluorescence intensity to change beyond what was measured after 45 s at the same pressure. As such, these results suggested to us that mechanical forces were indeed denaturing the GFPuv(Y39C/D103C) upon compression of the composites (**Scheme 5.1**).

To further demonstrate the mechanical origin of the observed activity, we also explored the mechanophoricity of the Y39C and D103C single mutants, wherein only one attachment site was present on the surface of the protein. Presumably, these mutants would function in a manner analogous to semi-telechelic derivatives of chemical mechanophores (*i.e.*, the mutants would not exhibit the same modulation in fluorescence as the Y39C/D103C double mutant).[1–4] Overexpression in *E. coli* and subsequent purification *via* nickel affinity chromatography afforded the desired semi-telechelic biomechanophores, which were subsequently added to the polymerization of poly(methyl methacrylate) as described above. As shown in **Figure 5.2**, compression of the resulting composites did not significantly alter their photophysical properties. Similarly, compression of composites containing unmodified GFPuv did not result in modulation of their associated photophysical properties, and mass spectrometry confirmed that GFPuv was not coupled to methyl methacrylate under the polymerization conditions (see Appendix D for additional details). Collectively, these data supported the conclusion that mechanical forces were responsible for the observed fluorescence modulation in the GFPuv(Y39C/D103C)-containing composites.

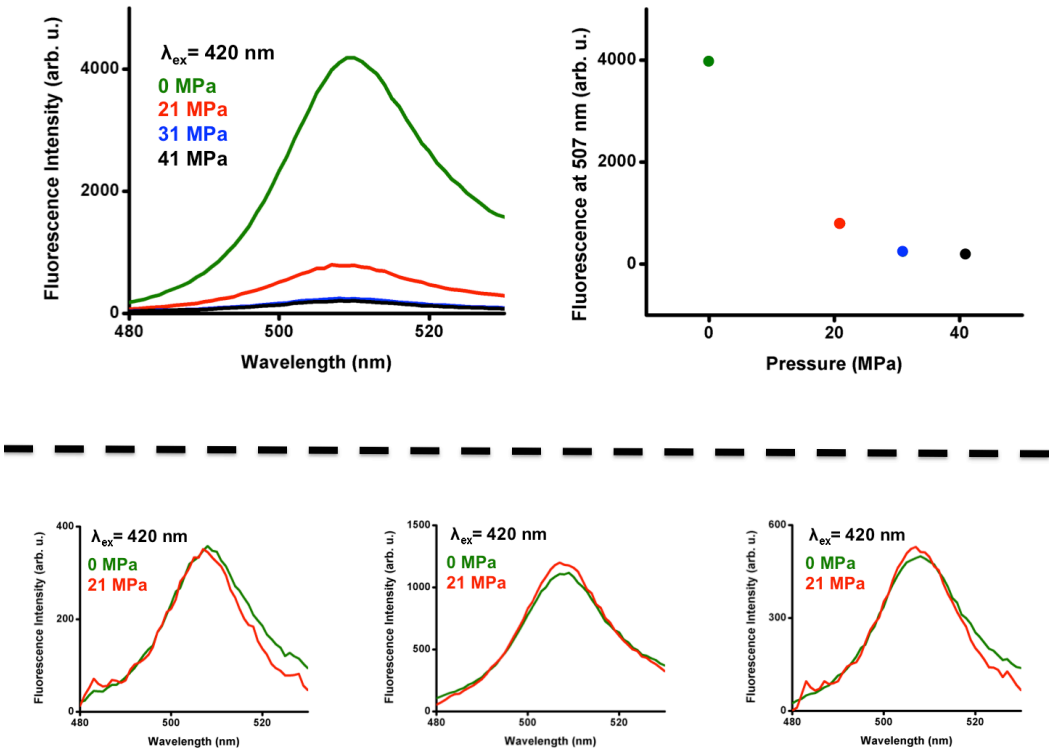


Figure 5.2 (Top) Compressing PMMA composites (0 – 41 MPa) containing double mutant GFPuv(Y39C/D103C) resulted in a monotonic decrease in the fluorescence intensity of the material. The fluorescence intensities at λ_{em} = 507 nm are plotted for clarity. (Bottom) Compression of PMMA composites containing GFPuv (left), GFPuv(Y39C) (center), or GFPuv(D103C) (right) did not significantly alter the fluorescence intensities of the materials.

CONCLUSIONS

In summary, we have demonstrated that mechanical forces may be used to modulate the photophysical properties of fluorescent proteins embedded within polymeric matrices. Our work constitutes the first example of modulating the λ_{em} of eYFP-containing polymer composite materials through the action of mechanical forces, as well as the first example of mechanically modulating the photophysical properties of GFPuv in a bulk material. Moreover, we have shown that mechanical perturbation of various fluorescent proteins alters their photophysical properties in distinct and tunable

manners (*i.e.*, the reported systems exhibited either ratiometric or intensiometric responses to mechanical stress). The ability to precisely manipulate the optical output of composite materials containing biomechanophores through targeted mutations could afford new opportunities for the facile development of stress-responsive materials with tailored sensitivities. Moreover, we have demonstrated that single-molecule experiments can guide the rational design of biomechanophores, and insight garnered from mechanochemical studies involving chemical systems can be translated to those involving force-sensitive biomolecules. Indeed, the technical simplicity associated with preparing the mechanically responsive biocomposites described herein holds promise for their development and utility as new classes of force responsive materials.

ACKNOWLEDGEMENTS

We thank professors Eric Gouaux and Arthur Glasfeld for generously providing the pNGFP-BC plasmid, and professor Andrew Ellington for providing the EYFP-pET21a expression plasmid.

REFERENCES

[1] J.N.Brantley, K.M. Wiggins, C.W. Bielawski. *Polym. Int.* **2013**, *62*, 2.

[2] K.M. Wiggins, J.N. Brantley, C.W. Bielawski. *Chem. Soc. Rev.* **2013**, *42*, 7130.

[3] M.M. Caruso, D.A. Davis, Q. Shen, S.A. Odom, N.R. Sottos, S.R. White, J.S. Moore. *Chem. Rev.* **2009**, *109*, 5755.

[4] A.L. Black, J.M. Lenhardt, S.L. Craig, *J. Mater. Chem.* **2011**, *21*,1655.

[5] C.R.Hickenboth, J.S. Moore, S.R. White, N,R, Sottos, J. Baudry, S.R. Wilson. *Nature* **2007**, *446*, 423.

[6] J.N. Lenhardt, M.T. Ong, R. Choe, C.R. Evenhuis, T.J. Martinez, S.L. Craig. *Science* **2010**, *329*, 1057.

[7] H.T. Baytekin, B. Baytekin, B. Grzybowski. *Angew. Chem.* **2012**, *124*, 3656.

[8] C.E. Diesendruck, B.D. Steinberg, N. Sugai, M.N. Silberstein, N.R. Sottos, S.R. White, P.V. Braun, J.S. Moore, *J. Am. Chem. Soc.* **2012**, *134*, 12446.

[9] R. Groote, L. van Haandel, R.P. Sijbesma. *J. Polym. Sci. Part A Polym. Chem.* **2012**, *50*, 4929.

[10] Z.S. Kean, A.L.Black, S.L. Ramirez S.L. Craig. *J. Polym. Sci. B. Polym. Phys.* **2012**, *50*, 3481.

[11] D.A. Davis, A. Hamilton, J. Yang, L.D. Cremar, D. Van Gough, S. Potisek, M.T. Ong, P.V. Braun, T.J. Martínez, S.R. White, J.S. Moore, N.R. Sottos. *Nature* **2009**, *459*, 68.

[12] C.M. Kingsbury, P.A. May, D.A. Davis, S.R. White J.S. Moore, N.R. Sottos. *J. Mater. Chem.* **2011**, *21*, 8381.

[13] B. Beiermann, D.A. Davis, S.L.B. Kramer, J.S. Moore, N.R. Sottos, S,R, White. *J. Mater. Chem.* **2011**, *21*, 8443.

[14] Y.K. Song, K.H. Lee, W.S. Hong, S.Y. Cho, H.C. Yu, C.M. Chung. *J. Mater. Chem.* **2012**, *22*, 1380.

[15] Chen, Y.; Spiering, A. J. H.; Karthikeyan, S.; Peters, G. W. M.; Meijer, E. W.; Sijbesma, R. P. *Nat. Chem.* **2012**, *4*, 559.

[16] J.N. Brantley, C.B. Bailey, K.M. Wiggins, A.T. Keatinge-Clay, C.W. Bielawski. *Polym. Chem.* **2013**, *4*, 3916.

[17] C. Bustamante, Y.R. Chemla, N.R. Forde, D. Izhaky, D. *Annu. Rev. Biochem.* **2004**, *73*, 705.

[18] A. Klibanov, G. Samokhin, K. Martinek, I.V. Berezin. *Biochim. Biophys. Acta* **1976**, *438*, 1.

[19] I.V. Berezin, A.M. Klibanov, K. Martinek. **1974**, *364*, 193.

[20] B. Choi, G. Zocchi, Y. Wu, S. Chan, L. Jeanne Perry. *Phys. Rev. Lett.* **2005**, *95*, 078102.

[21] A. Saghatelian, K.M. Guckian, D.A. Thayer, M.R. Ghadiri. *J. Am. Chem. Soc.* **2003**, *125*, 344.

[22] R.Y. Tsien. *Annu. Rev. Biochem.* **1998**, *67*, 509.

[23] A. Sawano. *Nucleic Acids Res.* **2000**, *28*, 78e.

[24] H. Niwa.; S. Inouye. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 13617.

[25] W.W. Ward, C.W. Cody, R.C. Hart, M.J. Cormier. *Photochem. Photobiol.* **1980**, *31*, 611.

[26] H. Dietz, M. Rief, G.H. Lorimer. **2004**, *101*, 16192.

[27] R. Perez-Jimenez, S. Garcia-Manyes, S.R.K. Ainavarapu, J.M. Fernandez. *J. Biol. Chem.* **2006**, *281*, 40010.

[28] V. Helms, T.P. Straatsma, J.A. McCammon. *J. Phys. Chem. B* **1999**, *103*, 3263.

[29] B. Barstow, N. Ando, C.U. Kim, S.M. Gruner. *Biophys. J.* **2009**, *97*, 1719.

[30] B. Barstow, N. Ando, C.U. Kim, S.M. Gruner. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 13362.

[31] N. Bruns, D. Clark. *Chimia*. **2010**, *65*, 245.

[32] N. Bruns, K. Pustelny, L.M. Bergeron, T.A. Whitehead, D.S. Clark. *Angew. Chem*. **2009**, *48*, 5666.

[33] K. Makyła, M. Müller, C. Lörcher, S. Winkler, M.G. Nussbaumer, M. Eder, N. Bruns. *Adv. Mater*. **2013**, *25*, 2701.

[34] T. Nagai, K. Ibata, E.S. Park, M. Kubota, K. Mikoshiba, A. Miyawaki. *Nat. Biotechnol*. **2002**, *20*, 87.

[35] T. Kawate, E. Gouaux. *Structure* **2006**, *14*, 673.

[36] A. Crameri, E.A. Whitehorn, E. Tate, W.P. Stemmer. *Nat. Biotechnol*. **1996**, *14*, 315.

[37] L. Zhao, Y. Li, X. Cao, J. You, W. Dong. *Nanotechnology* **2012**, *23*, 255702.

[38] N.B. Cramer, C.N. Bowman. *J. Polym. Sci. Part A Polym. Chem*. **2001**, *39*, 3311.

Experimental Section for Chapter 5

EXPERIMENTAL CONSIDERATIONS

1-Butyl-3-methylimidazolium hexafluorophosphtae (BMIM-PF_6) was prepared according to literature procedure.[1] All other reagents and materials were commercially available. Methyl methacrylate (MMA) was passed over a plug of basic alumina to remove any stabilizer prior to use. Azobisisobutyronitrile (AIBN) was recrystallized from acetone prior to use. Isopropyl β-D-1-thiogalactopyranoside (IPTG) and nickel-nitrilotriacetic acid agarose (Ni-NTA) were used without further purification. GFPuv (*A. victoria* green fluorescent protein exhibiting the following mutations: Q24H, A76S, L79V, A83S, Q91R, F99S, Y100F, M141L, M153T, P105Q, V163A, K173E, and I219V) was expressed from plasmid pNGFP-BC[2] (generous gift of Prof. Eric Gouaux, Oregon Health Science University). eYFP (*A. victoria* green fluorescent protein with the following mutations: S61G, S68A, R86Q, S98F, T154M, A164V, T204Y, and K207A)[3] was expressed from plasmid pET21a (generous gift of Prof. Andrew Ellington, University of Texas at Austin). Mechanical tests were performed using a standard benchtop Carver hydraulic press. Dynamic mechanical analysis (DMA) was performed using a TA Instruments Q800 series DMA outfitted with a compression clamp. Differential scanning calorimetry (DSC) was performed using a Metler Toledo 823^e DSC. Fluorescence spectra were acquired using a QuantaMaster Photon Technology International fluorometer. Gel permeation chromatography (GPC) was performed on a Viscotek GPCmax Solvent/Sample Module. Two fluorinated polystyrene columns (IMBHW-3078 and I-MBLMW-3078) were used in series and maintained at 24 °C. THF was used as the eluent at a flow rate of 1.0 mL min^{-1}. Detection was performed using a Viscotek VE 3580 Refractive Index Detector or a Viscotek 2600 Photodiode Array

Detector (tuned at 370 nm). Molecular weight and dispersity data are reported relative to polystyrene standards.

Site Directed Mutagenesis

Site directed mutagenesis of GFPuv was performed using pNGFP-BC as a template according to the quickchange method. To prepare GFPuv(D103C), the following primers were used (altered sequences are bold and red):

1) 5'-CGCACTATATCTTTCAAA**TG**TGACGGGAACTACAAGACG-3'

2) 5'-CGTCTTGTAGTTCCCGTCACATTT**GA**AAGATATAGTGCG-3'.

The following primers were used to prepare GFPuv(Y39C):

1) 5'-GAAGGTGATGCAACAT**G**CGGAAAACTTACCCTT-3'

2) 5'-AAGGGTAAGTTTTCCG**C**ATGTTGCATCACCTTC-3'

The GFPuv(Y39C/D103C) double mutant was generated using the GFPuv(Y39C) plasmid as a template and the above primers for GFPuv(D103C). All mutations were confirmed by DNA sequencing, and the mutants were expressed and purified using the same procedure as that described for the isolation of GFPuv (*vide infra*).

Protein Expression and Purification

Both GFPuv and eYFP were transformed into *E. coli* BL21 (DE3). Starter cultures (50 mL) were grown to inoculate pre-warmed Luria Broth supplemented with 50 μg/mL ampicillin. When OD_{600} = 0.4, the media was cooled to 15 °C and induced with 0.5 mM IPTG. After 16 h, the cells were harvested by centrifugation (3,000 relative centrifugal force for 20 minutes) and re-suspended in lysis buffer (10% glycerol v/v, 0.5 M NaCl, 100 mM HEPES pH 7.5), sonicated, and centrifuged (30,000 relative centrifugal force for 45 minutes) to remove cellular debris. The lysate was passed over a Ni-NTA agarose column equilibrated with lysis buffer. The column was washed with lysis buffer

containing 15 mM imidazole and protein was eluted with lysis buffer containing 150 mM imidazole. GFPuv was further purified using a Superdex 200 column equilibrated in 10 mM HEPES, pH 7.5, 150 mM NaCl. Final protein concentrations were determined using a Thermo Scientific Nanodrop 100 (absorbance at 280 nm).

PROCEDURE FOR THE PREPARATION OF EYFP COMPOSITES

An 8 mL Teflon capped vial was charged with a stir bar, eYFP (64 μL 9.4 mg mL^{-1} solution in lysis buffer; 1.8×10^{-8} mmol), MMA (0.5 mL; 4.7 mmol), BMIM-PF_6 (160 mg; 0.56 mmol) and AIBN (4.0 mg; 2.4×10^{-2} mmol). The vial was purged with nitrogen, sealed, and heated at 40 °C for 19 h with vigorous stirring. The resulting polymeric material was removed from the vial, washed with acetone (5 mL) and dried under reduced pressure. Cuboidal specimens for mechanical testing were prepared by cutting approximately 50 mg samples from the bulk material and polishing with ultrafine sanding paper.

REPRESENTATIVE PROCEDURE FOR THE PREPARATION OF GFP COMPOSITES: PREPARATION OF GFPUV(Y39C/D103C) COMPOSITES

An 8 mL Teflon capped vial was charged with a stir bar, GFPuv(Y39C/D103C) (0.35 mL 1.7 mg mL^{-1} solution in lysis buffer; 1.8 x 10^{-8} mmol), MMA (0.5 mL; 4.7 mmol), and AIBN (4.0 mg; 2.4 x 10^{-2} mmol). The vial was purged with nitrogen, sealed, and heated at 40 °C for 19 h with vigorous stirring. The resulting polymeric material was removed from the vial, washed with acetone (5 mL) and dried under reduce pressure. The resulting material was dissolved in tetrahydofuran (THF; 10 mg mL^{-1}) and analyzed with gel-permeation chromatography (**Figure E5.1**). Cuboidal specimens for mechanical testing were prepared by cutting approximately 50 mg samples from the bulk material

and polishing with ultrafine sanding paper. GFPuv, GFPuv(Y39C), and GFPuv(D103C) composites were prepared using an analogous procedure.

PREPARATION OF POLY(METHYL METHACRYLATE)

An 8 mL Teflon capped vial was charged with a stir bar, methyl methacrylate (MMA) (0.5 mL; 4.7 mmol), ethylene glycol dimethyacrylate (44 μL; 2.3×10^{-1} mmol) and AIBN (4.0 mg; 2.4×10^{-2} mmol). The vial was purged with nitrogen, sealed, and heated at 40 °C for 19 h with vigorous stirring. The resulting polymeric material was removed from the vial, washed with acetone (5 mL) and dried under reduced pressure. The resulting material was dissolved in tetrahydofuran (THF; 10 mg mL^{-1}) and analyzed with gel-permeation chromatography (**Figure E5.1**).

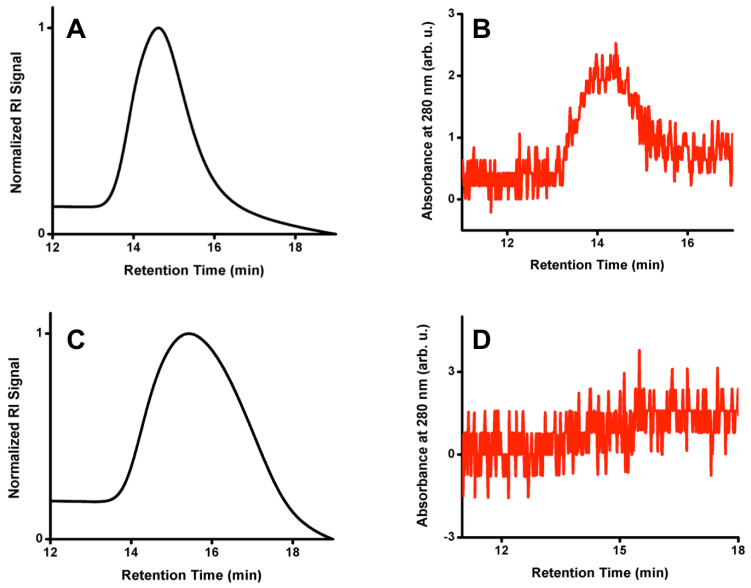


Figure E5.1 (A) Gel-permeation chromatograph of the GFPuv(Y39C/D103C) composite material (see section 1.5 for additional details). (B) Gel-permeation chromatograph of the GFPuv(Y39C/D103C) composite material visualized with ultraviolet-visible detection at 280 nm. (C) Gel-permeation chromatograph of poly(methyl methacrylate). (D) Gel-permeation chromatograph of poly(methyl methacrylate) visualized with ultraviolet-visible detection at 280 nm.

General Procedure for Mechanical Activation of eYFP Composites

A 50 mg sample of the eYFP composite was cut from the bulk material (*vide supra*), and the fluorescence of the material was recorded. The sample was then compressed (30, 110, 180, or 360 MPa) in a Carver hydraulic benchtop press for 45 s. The pressure exerted on the sample was determined using the relationship $P = FA^{-1}$, where F is the applied load, A is the cross-sectional area of the sample, and P is the applied pressure. As the Carver press employed two disk-shaped plates, the samples were found to compress into disks during the experimental studies. As such, A was approximated as the area of the disk following compression (DMA studies involving GFPuv composites validated this approximation; *vide infra*). **Figure 5.1** in **Chapter 5**

shows representative fluorescence spectra (normalized) of the material compressed at 0, 30,110, 180, and 360 MPa. Increasing the compression time to 1 h did not cause the λ_{em} of the material to shift beyond what was measured after compression for 45 s (see **Figure E5.2**). A decrease in fluorescence intensity, which was attributed to frictional denaturing (*vide infra*), was also observed (**Figure E5.3** and **Figure E5.4**).

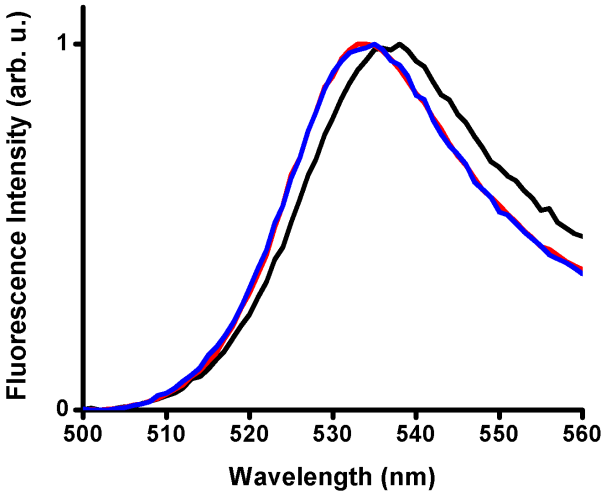


Figure E5.2 Normalized fluorescence spectra of eYFP composite (black) following compression (110 MPa) for 45 s (red) and 1 h (blue).

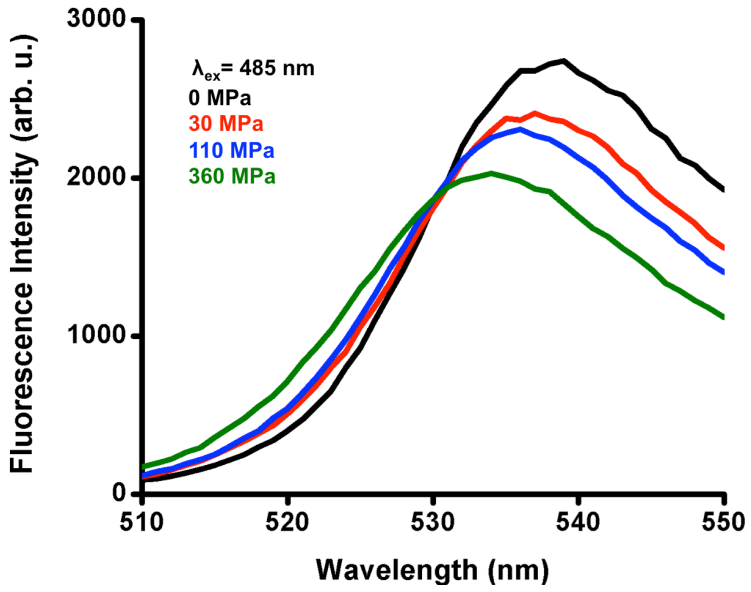


Figure E5.3 Fluorescence spectra of an eYFP composite (black) following compression at 30 MPa (red), 110 MPa (blue), and 360 MPa (green).

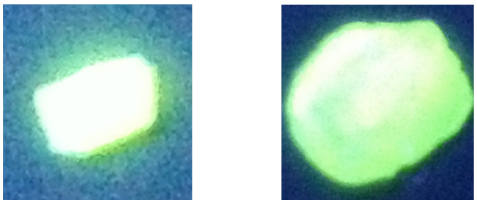


Figure E5.4 An eYFP-containing composite before (left) and after (right) compression at 110 MPa.

COMPRESSION OF LYOPHILIZED EYFP

Lyophilized eYFP (0.6 mg; 1.8×10^{-8} mmol) was loaded onto a Carver benchtop press and subjected to compression (external load of 3000 psi) for 45 s. The solid-state fluorescence of the sample was recorded (**Figure E5.5**), and the reduction in fluorescence intensity was attributed to thermal denaturation through frictional heating (a phenomenon previously observed[4] by Bruns *et al*.)

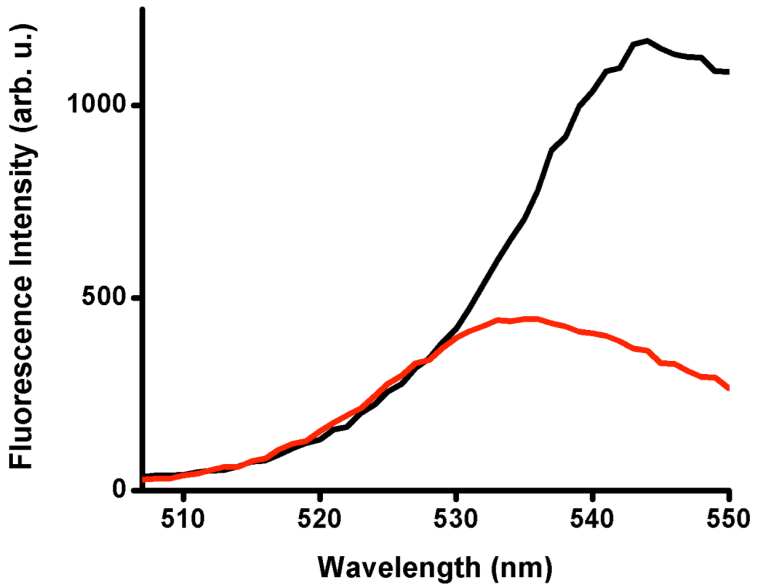


Figure E5.5 Fluorescence spectra of lyophilized eYFP (black) following compression (external load of 3000 psi; red).

PREPARATION AND MECHANICAL ACTIVATION OF MIXED eYFP/GFPUV(Y39C/D103C) COMPOSITES

A Teflon capped 8 mL vial was charged with eYFP (64 μL 9.4 mg mL^{-1} solution in lysis buffer; 1.8×10^{-5} mmol), GFPuv(Y39C/D103C) (0.24 mL 1.7 mg mL^{-1} solution in lysis buffer; 1.2×10^{-8} mmol), MMA (0.5 mL; 4.7 mmol), BMIM-PF_6 (160 mg; 0.56 mmol) and AIBN (4.0 mg; 2.4×10^{-2} mmol). The vial was purged with nitrogen, sealed, and heated at 40 °C for 19 h with vigorous stirring. The resulting polymeric material was removed from the vial, washed with acetone (5 mL) and dried under reduced pressure. A cuboidal specimen for mechanical testing was prepared by cutting an approximately 50 mg sample from the bulk material and polishing with ultrafine sanding paper. The sample was then compressed at 30 MPa for 45 s. **Figure E5.6** shows that the λ_{em} of eYFP was hypsochromically shifted following compression, whereas the λ_{em} of GFPuv(Y39C/D103C) did not shift following compression (the fluorescence intensity did, however, decrease).

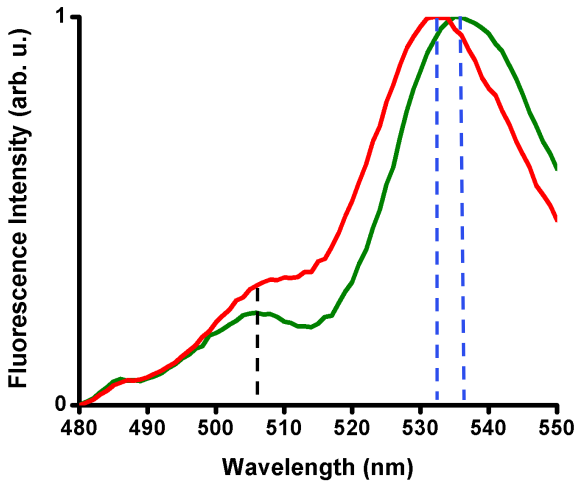


Figure E5.6 Normalized fluorescence spectra of mixed eYFP/GFPuv(Y39C/D103C) composite (green) following compression (30 MPa; 45 s). Dashed blue lines are drawn from the λ_{em} associated with eYFP. A dashed black line is drawn from the λ_{em} associated with GFPuv(Y39C/D103C).

General Procedure for Compression of GFP Composites

A 50 mg sample of the GFP composite was cut from the bulk material (*vide supra*), and the fluorescence of the material was recorded. The sample was then compressed (21, 31, or 41 MPa) in a Carver hydraulic benchtop press for 45 s. The pressure exerted on the sample was determined using the relationship $P = FA^{-1}$, where F is the applied load, A is the cross-sectional area of the sample, and P is the applied pressure. As the Carver press employed two disk-shaped plates, the samples were found to compress into disks during the experimental studies. As such, A was approximated as the area of the disk following compression. **Figure 5.2** in **Chapter 5** shows representative fluorescence spectra of the GFP materials following compression at various pressures (see also **Figure E5.9**). Increasing the compression time to 1 h did not cause the fluorescence intensity of the GFPuv(Y39C/D103C) material to decrease beyond what was measured after compression for 45 s at the same pressure (see **Figure E5.7**). Moreover, increasing the compression time to 1 h did not cause the fluorescence intensity

of the control materials (*e.g.*, the GFPuv composite) to alter significantly from that measured following compression for 45 s at the same pressure (**Figure E5.8**).

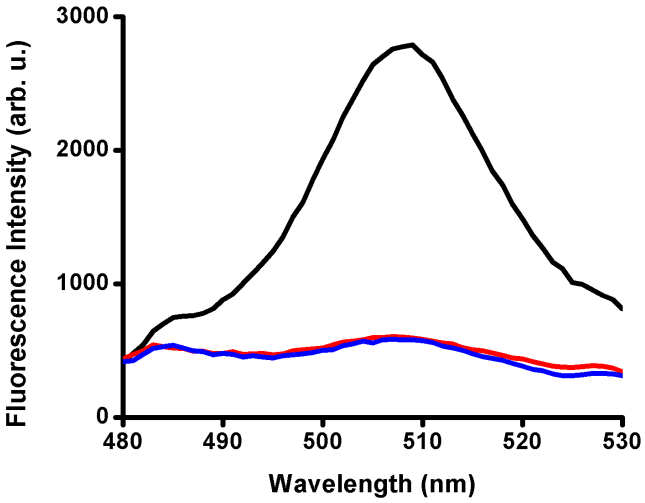


Figure E5.7 Fluorescence spectra of GFPuv(Y39C/D103C) composite (black) following compression (31 MPa) for 45 s (red) and 1 h (blue).

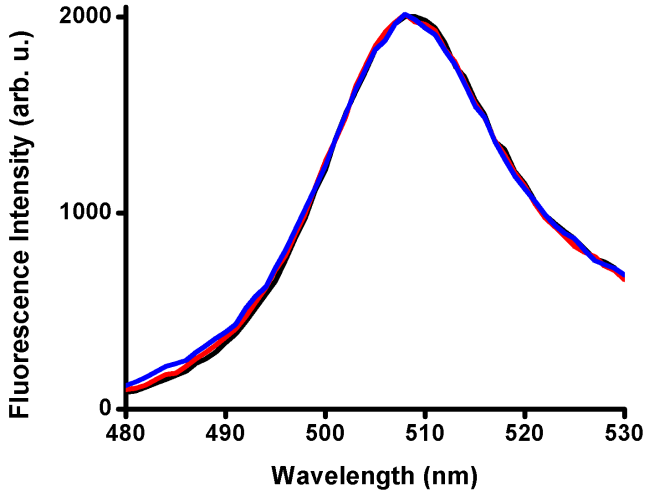


Figure E5.8 Fluorescence spectra of GFPuv composite (black) following compression (31 MPa) for 45 s (red) and 1 h (blue).

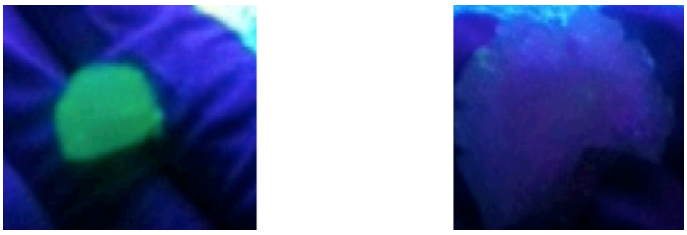


Figure E5.9 A GFPuv(Y39C/D103C)-containing composite before (left) and after (right) compression (external load of 7000 psi).

REPRESENTATIVE PROCEDURE FOR DMA ANALYSIS OF GFPUV(Y39C/D103C) COMPOSITE

A cuboidal specimen of the GFPuv(Y39C/D103C) composite (1.14 mm × 1.14 mm × 0.72 mm) was mounted in a TA Instruments Q800 series DMA outfitted with a compression clamp and subjected to controlled force compression (pre-load force of 0.001 N; force ramp rate of 1 N min^{-1}). The applied stress and resultant material strain were recorded until a maximal stress of approximately 10.4 MPa was exerted on the sample (**Figure E5.10**). The fluorescence intensity of the material following DMA analysis was found to decrease relative to the fluorescence intensity of the material prior to DMA analysis (**Figure E5.10**).

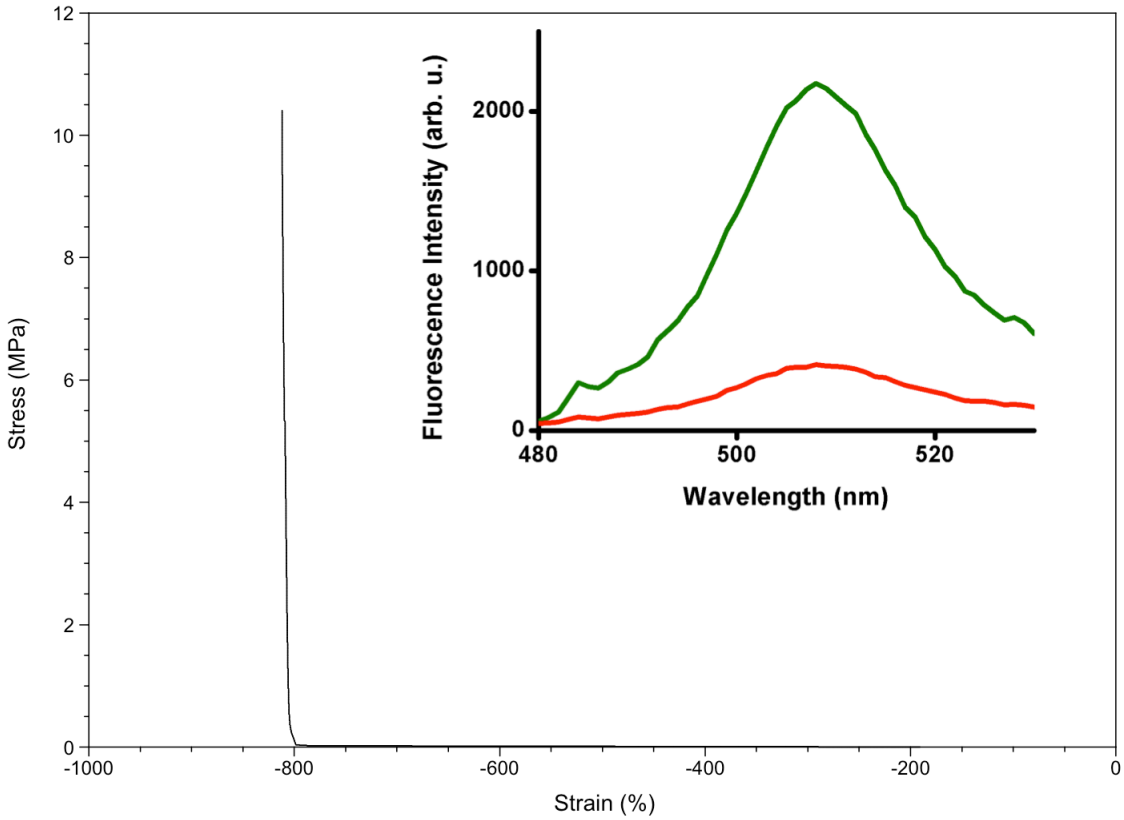


Figure E5.10 Stress/strain curve for DMA analysis of GFPuv(Y39C/D103C) composite. Inset shows the fluorescence of the composite prior to (green) and following (red) DMA analysis.

REPRESENTATIVE PROCEDURE FOR DMA ANALYSIS OF GFP CONTROL MATERIALS: DMA ANALYSIS OF GFPUV COMPOSITE.

A cuboidal specimen of GFPuv composite (1.18 mm × 1.18 mm × 0.71 mm) was mounted in a TA Instruments Q800 series DMA outfitted with a compression clamp and subjected to controlled force compression (pre-load force of 0.001 N; force ramp rate of 1 N min^{-1}). The applied stress and resultant material strain were recorded until a maximal stress of approximately 10.4 MPa was exerted on the sample (**Figure E5.11**). The fluorescence intensities of the material prior to and following DMA analysis are shown in **Figure E5.11**.

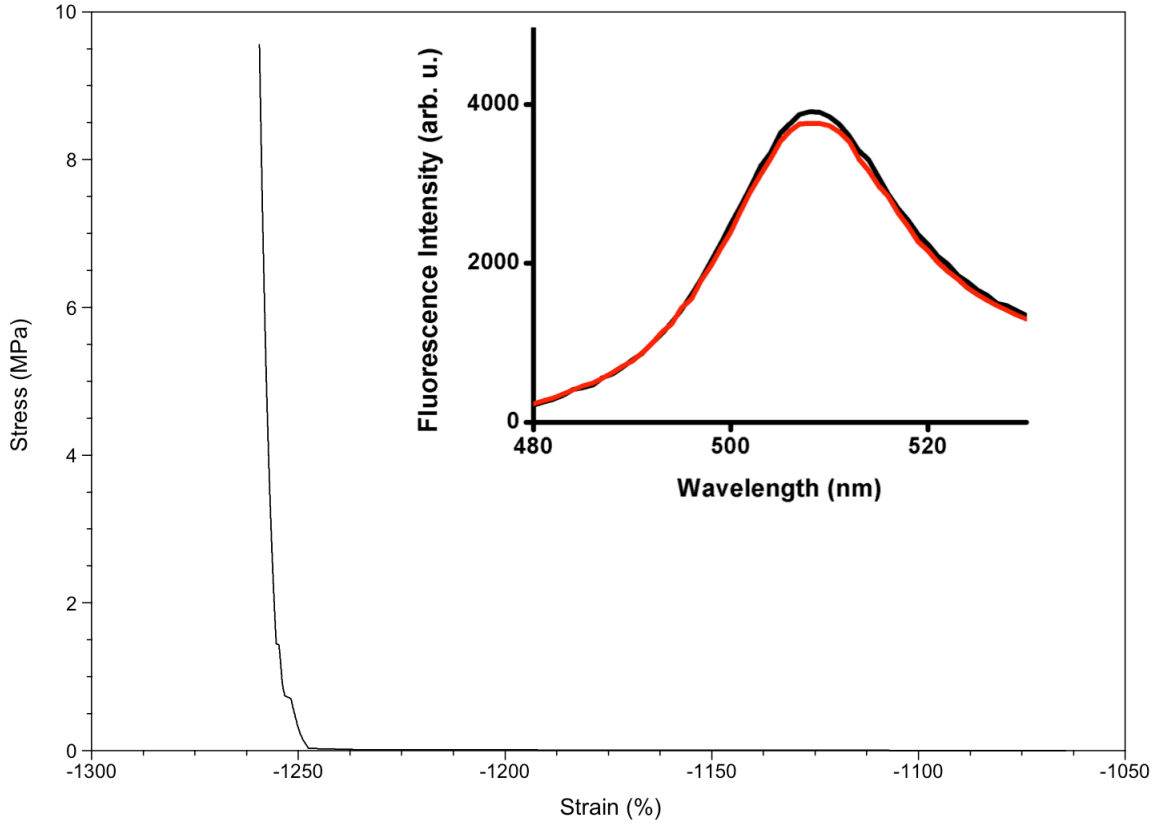


Figure E5.11 Stress/strain curve for DMA analysis of GFPuv composite. Inset shows the fluorescence of the composite prior to (black) and following (red) DMA analysis.

GFPUV MASS SPECTROMETRY STUDIES: GENERAL CONSIDERATIONS

Proteins were infused into an Orbitrap Elite mass spectrometer (ThermoFisher, San Jose, CA) at 5 μL/min. Intact protein spectra were collected in positive mode and were a composite of 50 averaged scans acquired at maximum resolution (240,000 at m/z 400). Neutral masses were then extrapolated using the Xtract algorithm (ThermoFisher, San Jose, CA) with a signal to noise ratio of 5:1.

Procedure for Evaluating the Reactivity of GFPuv(Y39C/D103C) with Methyl Methacrylate under Polymerization Conditions

An 8 mL Teflon capped vial was charged with a stir bar, GFPuv(Y39C/D103C) (0.35 mL 1.7 mg mL^{-1} solution in lysis buffer; 1.8×10^{-8} mmol), MMA (0.5 mL; 4.7 mmol), and AIBN (4.0 mg; 2.4×10^{-2} mmol). The vial was purged with nitrogen, sealed, and heated at 40 °C for 10 minutes with vigorous stirring. The vial was exposed to the ambient atmosphere, and the reaction mixture was subsequently desalted and buffer exchanged via 5 sequential exchanges using a 10 kDa molecular weight cutoff filter into LC-MS grade water. The resulting protein solution was diluted to 10 μM in 49/50/1 water/acetonitrile/formic acid, and infused into an Orbitrap Elite mass spectrometer. As shown in **Figure E5.12**, both cysteine mutations were reactive with methyl methacrylate under these conditions.

Procedure for Evaluating the Reactivity of GFPuv(Y39C/D103C) with Methyl Methacrylate in the Absence of Radical Initiators

An 8 mL Teflon capped vial was charged with a stir bar, GFPuv(Y39C/D103C) (0.35 mL 1.7 mg mL^{-1} solution in lysis buffer; 1.8×10^{-8} mmol), and MMA (0.5 mL; 4.7 mmol). The vial was purged with nitrogen, sealed, and heated at 40 °C for 10 minutes with vigorous stirring. The vial was exposed to the ambient atmosphere, and the reaction mixture was subsequently desalted and buffer exchanged via 5 sequential exchanges using a 10 kDa molecular weight cutoff filter into LC-MS grade water. The resulting protein solution was diluted to 10 μM in 49/50/1 water/acetonitrile/formic acid, and infused into an Orbitrap Elite mass spectrometer. As shown in **Figure E5.12**, both cysteine mutations were not reactive with methyl methacrylate under these conditions.

Procedure for Evaluating the Reactivity of GFPuv with Methyl Methacrylate under Polymerization Conditions

An 8 mL Teflon capped vial was charged with a stir bar, GFPuv (0.14 mL 4.2 mg mL^{-1} solution in lysis buffer; 1.8×10^{-8} mmol), MMA (0.5 mL; 4.7 mmol), and AIBN (4.0 mg; 2.4×10^{-2} mmol). The vial was purged with nitrogen, sealed, and heated at 40 °C for 10 minutes with vigorous stirring. The vial was exposed to the ambient atmosphere, and the reaction mixture was subsequently desalted and buffer exchanged via 5 sequential exchanges using a 10 kDa molecular weight cutoff filter into LC-MS grade water. The resulting protein solution was diluted to 10μM in 49/50/1 water/acetonitrile/formic acid, and infused into an Orbitrap Elite mass spectrometer. **Figure E5.12** shows that no change in the mass of the protein was observed under these conditions.

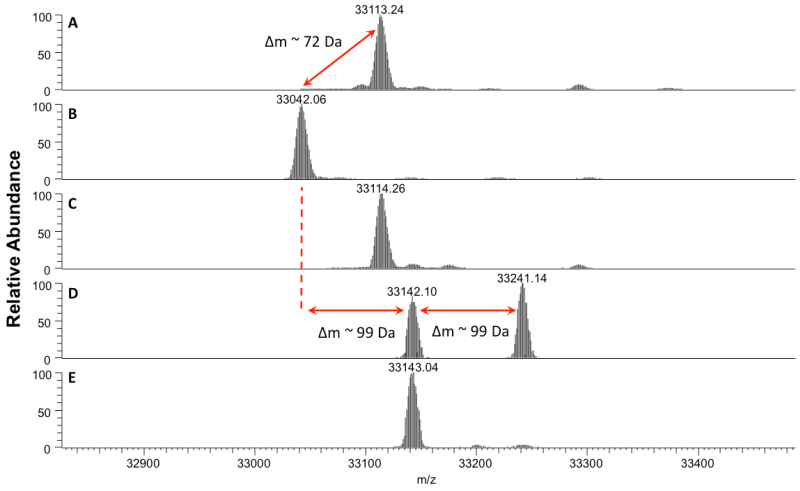


Figure E5.12 (a, b) Deconvoluted intact mass spectra of GFPuv (a) and GFPuv(Y39C/D103C) (b). The expected 72 Da mass shift associated with both mutations is observed. (c) Deconvoluted intact mass spectrum of GFPuv following reaction with methyl methacrylate in the presence of AIBN. (d) Deconvoluted intact mass spectrum of GFPuv(Y39C/D103C) following reaction with methyl methacrylate in the presence of AIBN. (e) Deconvoluted intact mass spectrum of GFPuv(Y39C/D103C) following reaction with methyl methacrylate in the absence of AIBN.

WIDE FIELD FLUORESCENCE MICROSCOPY GENERAL CONSIDERATIONS

Fluorescence images were collected at room temperature using an upright microscope (Olympus BX60) with high-pressure mercury lamp excitation and a dichroic mirror (Olympus WB filter cube) combined with a digital camera (SPOT). The microscope objective lens used was a 5 X, 0.12 N.A., dry, CP-Achromat (Zeiss, 440920). All fluorescence images are false-colored (ImageJ thallium filter).

PREPARATION OF GFPUV(Y39C/D103C) COMPOSITE FOR WIDE FIELD FLUORESCENCE MICROSCOPY ANALYSIS

An 8 mL Teflon capped vial was charged with a stir bar, GFPuv(Y39C/D103C) (0.35 mL 1.7 mg mL^{-1} solution in lysis buffer; 1.8×10^{-8} mmol), MMA (0.5 mL; 4.7 mmol), and AIBN (4.0 mg; 2.4×10^{-2} mmol). The vial was purged with nitrogen, sealed, and heated at 40 °C for 19 h with vigorous stirring. The resulting polymeric material was removed from the vial, washed with acetone (5 mL) and dried under reduced pressure. A cuboidal specimen was prepared by cutting an approximately 50 mg sample from the bulk material and polishing with ultrafine sanding paper. The specimen was mounted into a Carver benchtop hydraulic press, and a square plate (2.41 mm × 2.41 mm) was compressed into the material. The area of the plate and the applied load were used to calculate the pressure exerted on the sample (approximately 10.3 MPa). The sample was then analyzed using a wide field fluorescence microscope, which revealed decreased fluorescence intensity in the compressed area relative to the uncompressed material (**Figure 5E.13**).

PREPARATION OF GFPUV COMPOSITE FOR WIDE FIELD FLUORESCENCE MICROSCOPY ANALYSIS

An 8 mL Teflon capped vial was charged with a stir bar, GFPuv (0.14 mL 4.2 mg mL^{-1} solution in lysis buffer; 1.8×10^{-8} mmol), MMA (0.5 mL; 4.7 mmol), and AIBN

(4.0 mg; 2.4×10^{-2} mmol). The vial was purged with nitrogen, sealed, and heated at 40 °C for 19 h with vigorous stirring. The resulting polymeric material was removed from the vial, washed with acetone (5 mL) and dried under reduced pressure. A cuboidal specimen was prepared by cutting an approximately 50 mg sample from the bulk material and polishing with ultrafine sanding paper. The specimen was mounted into a Carver benchtop hydraulic press, and a square plate (2.41 mm × 2.41 mm) was compressed into the material. The area of the plate and the applied load were used to calculate the pressure exerted on the sample (approximately 10.3 MPa). The sample was then analyzed using a wide field fluorescence microscope, which revealed negligible changes in the fluorescence intensity in the compressed area relative to the uncompressed material (**Figure E5.14**).

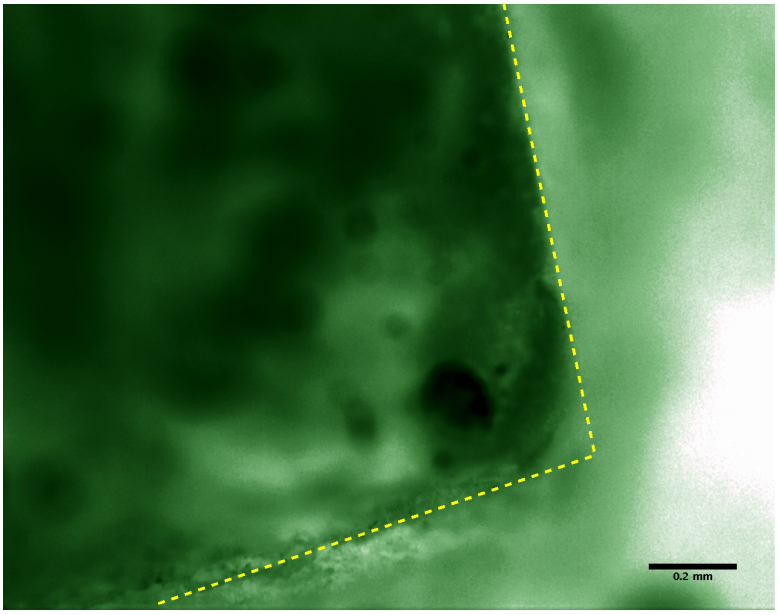


Figure E5.13 Wide field fluorescence micrograph of GFPuv(Y39C/D103C) composite following compression. The yellow dashed line outlines the edge of the square compression site (see text for additional details).

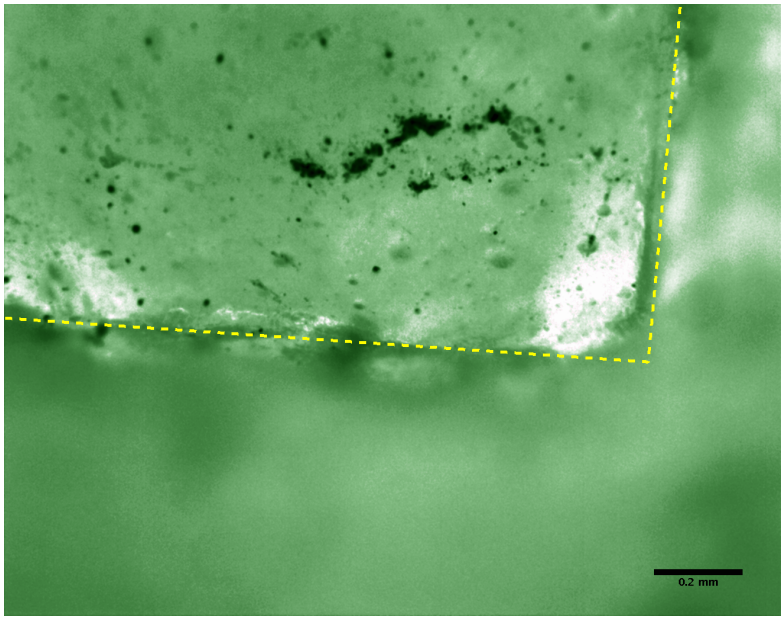


Figure E5.14 Wide field fluorescence micrograph of GFPuv composite following compression. The yellow dashed line outlines the edge of the square compression site (see text for additional details).

REFERENCES

[1] J. Dupont, C. Consorti, P. Suarez, R. de Souza, R. *Org. Synth.* **2004**, *10*, 184.

[2] T. Kawate, E. Gouaux. *Structure* **2006**, *14*, 673.

[3] T. Nagai, K. Ibata, E.S. Park, M. Kubota, K. Mikoshiba, A. Miyawaki. *Nat. Biotech.* **2002**, *20*, 87.

[4] K. Makyła, C. Müller, S. Lörcher, T. Winkler, N.G. Nussbaumer, M. Eder, N. Bruns. *Adv. Mater.* **2013**, *25*, 2701.

Chapter 6: Conclusions and Outlook

This dissertation presents a few incremental advances in deciphering enzymatic mechanisms, understanding the collaborative actions of the various domains within a module in *in vitro* environments, and structural dissection that contribute to the ultimate goal of realizing the biocatalytic potential of PKSs. For PKSs to be a useful and rationally reprogrammable platform for synthetic biology and biocatalysis, several enzymatic features must be understood. First, we must understand how the domains work at a mechanistic level, and how to reprogram their selectivities (*e.g.* changing the AT's extender unit selectivity as was described in **Chapter 1,** or changing the stereocontrol that the KR confers, as was experimentally demonstrated in **Chapter 3**). Second, we must understand the downstream effects of such structural changes in terms of the interacting kinetics and selectivities of the various domains. In **Chapter 2,** a platform that examines a miniaturized system—a ModTE construct—evaluates how the various domains within a module interact in unnatural environments. Finally, we must understand bridging protein-protein interactions that facilitate chain transfer and afford different gene products to operate in concert with one another. **Chapter 4** illuminates one of these potential protein-protein interactions in a relatively uncharacterized class of polyketide synthases, the *trans*-AT PKSs.

Polyketide synthases have potential for developing of facile strategies to generate chiral building blocks. ModTE systems like the one described in **Chapter 2** have the potential to generate libraries of stereoisomeric products such as the triketide lactones highlighted in **Chapter 1**. **Chapter 2** explores a robust platform to generate one of these triketide lactones. In the future, steps will be taken to optimize the biocatalytic scheme including exploring the use of promiscuous acyl-CoA ligases as well as directed

evolution approaches to attenuate the thioesterase activity towards diketide substrates to improve yields. Furthermore, other ModTE systems that have 1) different selectivities for the stereoisomeric configurations of the starter units and 2) different KR types harbored within the module will be explored to build a library of stereoisomeric chiral building blocks (as was described in **Chapter 1**). Additionally, the substrate scope of ModTE systems needs to be further elucidated, including exploring the incorporation of synthetic handles such as terminal olefins, terminal alkynes, halides, or azides. This has been explored somewhat by the Keatinge-Clay laboratory to develop of chemical biology tools. One such study explored the use of triflouromethyl groups as a ^{19}F probe to detect reactions performed in cell lysate.[1] A second explored the use of terminal alkynes for use in attaching a fluorescent tag via CuAAC, which resulted in the detection of minor PKS products from a ModTE platform.[2] In the future, however, we hope to use this demonstrated substrate promiscuity to incorporate synthetic handles to construct complex, polyketide scaffolds. Future studies will include collaborations with total synthesis groups to demonstrate that concise, chemoenzymatic syntheses can be realized with these chiral building blocks.

In addition to their role in generating chiral building blocks, although this was not our intention when we embarked upon this project, the ModTE system described in **Chapter 2** has received some notice from the "cell free" biocatalysis/synthetic biology community.[3,4] The triketide lactones described in **Chapter 2** generated through the multi-enzyme cascades (utilizing MatB, the KR, the GDH-based cofactor-recycling scheme in addition to the multi-domain Mod6TE protein) presents one of the early examples forming complex small molecules from cell lysate. The goals of the "cell free" biocatalysis/synthetic biology movement are to generate the most minimal systems required to generate molecules of interest through biological methods. Thus, this work

presents incremental advancements in terms of the molecular complexity that can be generated outside of the context of a cell. Perhaps this work will assist in the development of cell free "chassis" platforms to enable powerful transformations in synthetic biology in the future.

In **Chapter 3**, a number of insights regarding the intrinsic reactivities of polyketide synthase ketoreductases were obtained. Most notably, that the energetic default pathway for PKS KRs is the one that forms the A2 stereoisomer, which is partially explained by the Felkin-Anh model to rationalize the intrinsic preference for *anti* diastereoselectivity. Interestingly, we demonstrate that polar "orienting" interactions from hydrogen bonding residues play a small role in enforcing KR stereocontrol. Indeed, with truncated mimics, the KRs are very much acting as loosely bound catalysts that aid asymmetric induction. With that insight, then, there are a lot of potential future directions to improve the utility of KRs as biocatalysts, as well as engineer specific enzymatic outcomes. Because the energetic differences are small, it appears that drastic differences in stereochemical outcome can be achieved by a very constrained set of residues that surround the active site and comprise the active site stereoelectronic environment.

These weak binding interactions in KRs may seem like a shortcoming in terms of their power as biocatalysts, but in reality there is an element of engineerability in weak non-covalent interactions, as minor structural changes can be harnessed to effect large changes. A saturation mutagenesis approach would likely be fruitful in terms of designing specific mutants that are active and stereocontrolled when reducing specific substrates. Moreover, the quantitative structure-activity relationships and "Big Data" approaches harnessed effectively by the group of Matthew Sigman at University of Utah could be a powerful way to explore engineering KRs. Indeed, the robust chemical models of carbonyl nucleophilic attack, the small substrates (resulting in the ability to use

computational chemistry to determine chemical parameters), and the stereochemical distribution of products (which correlates to $\Delta\Delta G^{\ddagger}$) taken together render PKS KRs an ideal system to translate this approach to a biological system. While the "big data" approach has been successful in providing useful predictive information that affords the rational design of catalysts for assymetric transformations in chemical systems, it has yet to be applied to an enzymatic system.[5,6] Perhaps these approaches in the future will afford the ability to design mutant enzymes to selectively reduce a broad range of desired substrates in a stereocontrolled fashion.

In **Chapter 4**, a domain from a *trans*-acyltransferase PKS is investigated structurally, functionally, and bioinformatically. *Trans*-AT PKSs have remained elusive due to their more recent identification and biochemical investigation as well as their truly bizarre modular organizations. One of the organizations, termed "split bimodules" by Piel, is a paradigm of the unusual biosynthetic strategies found in *trans*-AT PKSs.[7] Our structural characterization has revealed some potential clues regarding how these unusual modules assemble. Notably, new helical regions and sequence truncations are revealed through structural characterization and examination of sequence trends. Future functional experiments need to be explored to examine the fundamental reactivity of some of the strange biosynthetic strategies described in **Chapter 4**. To further explore the mechanistic details of split bimodules, one approach might be performing alanine scanning along the new helical region of the A' KRs, or deleting this helical region entirely and then determining if there is a buildup of intermediates on one or both ACPs *via* mass spectrometry. Also, deuterium-labeling experiments, elucidating the exact nature of the double dehydration, might be informative.

In **Chapter 5**, an additional project unrelated to the main work of this dissertation is described. This project explores the use of mechanical force in the context of biological

systems through using fluorescent proteins as reporter molecules. In the case of a green fluorescent protein, a ratiometric response was observed, whereas in the case of yellow fluorescent protein an intensiometric response was observed. Work that has built on our findings from Jierry and coworkers has established that similar responses can be observed with much weaker mechanical compressions than those reported by us.[8] Additionally, Jierry and coworkers developed a reversible system building on the work presented in our irreversible system. Thus, the work described in this dissertation demonstrates a key example of harnessing alternate reactivity pathways in biological systems.

In summary, this dissertation demonstrates that PKS domains can be harnessed to effect chemical complex transformations. As we continue to decipher the mechanistic details of these systems, polyketide synthases will certainly become a more powerful platform for applications synthetic biology and biocatalysis.

REFERENCES

[1] S. Piasecki, A. Keatinge-Clay, *Synlett* **2012**, *23*, 1840.

[2] A. J. Hughes, M. R. Tibby, D. T. Wagner, J. N. Brantley, A. T. Keatinge-Clay, *Chem. Commun.* **2014**, *50*, 5276.

[3] Q. M. Dudley, A. S. Karim, M. C. Jewett, *Biotechnol. J.* **2015**, *10*, 69–82.

[4] S. Billerbeck, J. Härle, S. Panke, *Curr. Opin. Biotechnol.* **2013**, *24*, 1037–43.

[5] K. C. Harper, E. N. Bess, M. S. Sigman, *Nat. Chem.* **2012**, *4*, 366–74.

[6] A. Milo, A. J. Neel, F. D. Toste, M. S. Sigman, *Science* **2015**, *347*, 737–43.

[7] J. Piel, *Nat. Prod. Rep.* **2010**, *27*, 996–1047.

[8] J. Longo, C. Yao, C. Rios, N. T. T. Chau, F. Boulmedais, J. Hemmerlé, P. Lavalle, S. M. Schiller, P. Schaaf, L. Jierry, *Chem. Commun.* **2015**, *51*, 232–5.

References

D. L. Akey, J. R. Razelun, J. Tehranisa, D. H. Sherman, W. H. Gerwick, J. L. Smith, *Structure*. **2010**, *18*, 94.

D. L. Alkon, M.-K. Sun, T. J. Nelson, *Trends Pharmacol. Sci*. **2007**, *28*, 51.

T. Annaval, C. Paris, P. F. Leadlay, C. Jacob, K. J. Weissman, *Chembiochem* **2015**, *16*, 1357.

C. B. Bailey, M. E. Pasman and A. T. Keatinge-Clay, *ChemComm*, 2015, DOI: 10.1039/C5CC07315D.

S. Bali, H. M. O'Hare, K. J. Weissman, *Chembiochem* **2006**, *7*, 478.

S. Bali, K. J. Weissman, *Chembiochem* **2006**, *7*, 1935.

A. Baerga-Ortiz, B. Popovic, A. P. Siskos, H. M. O'Hare, D. Spiteller, M. G. Williams, N. Campillo, J. B. Spencer, P. F. Leadlay, *Chem. Biol*. **2006**, *13*, 277.

B. Barstow, N. Ando, C.U. Kim, S.M. Gruner. *Biophys. J*. **2009**, *97*, 1719.

B. Barstow, N. Ando, C.U. Kim, S.M. Gruner. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 13362.

H.T. Baytekin, B. Baytekin, B. Grzybowski. *Angew. Chem*. **2012**, *124*, 3656.

B. Beiermann, D.A. Davis, S.L.B. Kramer, J.S. Moore, N.R. Sottos, S,R, White. *J. Mater. Chem*. **2011**, *21*, 8443.

I.V. Berezin, A.M. Klibanov, K. Martinek. **1974**, *364*, 193.

S. Billerbeck, J. Härle, S. Panke, *Curr. Opin. Biotechnol*. **2013**, *24*, 1037.

A.L. Black, J.M. Lenhardt, S.L. Craig, *J. Mater. Chem*. **2011**, *21*,1655.

K. Blin, M. H. Medema, D. Kazempour, M. A. Fischbach, R. Breitling, E. Takano, T. Weber, *Nucl. Acids Res*. **2013**, *41*, W204.

S. A. Bonnett, J. R. Whicher, K. Papireddy, G. Florova, J. L. Smith, K. A. Reynolds, *Chem. Biol.* **2013**, *20*, 772.

J.W. Bosco, B.R. Rama Raju, A.K. Saikia. *Synth. Comm.* **2004**. *34*, 2849.

S. F. Brady, L. Simmons, J. H. Kim, E. W. Schmidt, *Nat. Prod. Rep.* **2009**, *11*, 1488-1503.

J. N. Brantley, C. B. Bailey, J. R. Cannon, K. a. Clark, D. a. Vandenbout, J. S. Brodbelt, A. T. Keatinge-Clay, C. W. Bielawski, *Angew. Chem.* **2014**, *53*, 5088.

J.N. Brantley, C.B. Bailey, K.M. Wiggins, A.T. Keatinge-Clay, C.W. Bielawski. *Polym. Chem.* **2013**, *4*, 3916.

J.N.Brantley, K.M. Wiggins, C.W. Bielawski. *Polym. Int.* **2013**, *62*, 2.

T. Bretschneider, J. B. Heim, D. Heine, R. Winkler, B. Busch, B. Kusebauch, T. Stehle, G. Zocher, C. Hertweck, *Nature*. **2013**, *502*, 124.

G. Berkhan, F. Hahn, *Angew. Chem.* **2014**, *53*, 14240.

N. Bruns, D. Clark. *Chimia*. **2010**, *65*, 245.

N. Bruns, K. Pustelny, L.M. Bergeron, T.A. Whitehead, D.S. Clark. *Angew. Chem. Int. Ed.* **2009**, *48*, 5666.

B. Busch, N. Ueberschaar, S. Behnken, Y. Sugimoto, M. Werneburg, N. Traitcheva, J. He, C. Hertweck, *Angew. Chem. Int.* **2013**, *52*, 5285.

R. A. Butcher, F. C. Schroeder, M. A. Fischbach, P. D. Straight, R. Kolter, C. T. Walsh, J. Clardy, *Proc. Natl. Acad. Sci., U. S. A.* **2007**, *104*, 1506.

C. Bustamante, Y.R. Chemla, N.R. Forde, D. Izhaky, D. *Annu. Rev. Biochem.* **2004**, *73*, 705.

E. Busto, V. Gotor-Fernández, V. Gotor, *J. Org. Chem.* **2012**, *77*, 4842.

P. Caffrey, *ChemBioChem* **2003**, *4*, 654.

P. Caffrey, *Chem. Biol.* **2005**, *12*, 1060.

C. T. Calderone, D. F. Iwig, P. C. Dorrestein, N. L. Kelleher, C. T. Walsh, *Chem. Biol.* **2007**, *14*, 835.

C. T. Calderone, W. E. Kowtoniuk, N. L. Kelleher, C. T. Walsh, P. C. Dorrestein, *Proc. Natl. Acad. Sci., U. S. A.* **2006**, *103*, 8977.

M.M. Caruso, D.A. Davis, Q. Shen, S.A. Odom, N.R. Sottos, S.R. White, J.S. Moore. *Chem. Rev.* **2009**, *109*, 5755.

R. Castonguay, W. He, A. Y. Chen, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.* **2007**, *129*, 13758.

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

A. Y. Chen, D. E. Cane, C. Khosla, *Chem. Biol.* **2007**, *14*, 784.

A. Y. Chen, N. A. Schnarr, C.-Y. Kim, D. E. Cane, C. Khosla, *J. Am. Chem. Soc.* **2006**, *128*, 3067.

Y. Chen, A.J.H. Spiering, S. Karthikeyan, G.W.M. Peters, E.W. Meijer, R.P. Sijbesma, *Nat. Chem.* **2012**, *4*, 559.

X.-H. Chen, J. Vater, J. Piel, P. Franke, R. Scholz, K. Schneider, A. Koumoutsi, G. Hitzeroth, N. Grammel, A. W. Strittmatter, et al., *J. Bacteriol.* **2006**, *188*, 4024.

Y.-Q. Cheng, J. M. Coughlin, S.-K. Lim, B. Shen, *Methods Enzymol.* **2009**, *459*, 165.

B. Choi, G. Zocchi, Y. Wu, S. Chan, L. Jeanne Perry. *Phys. Rev. Lett.* **2005**, *95*, 078102.

J. Cortes, K. E. Wiesmann, G. A. Roberts, M. J. Brown, J. Staunton, P. F. Leadlay, *Science* **1995**, *268*, 1487.

N.B. Cramer, C.N. Bowman. *J. Polym. Sci. Part A Polym. Chem.* **2001**, *39*, 3311.

A. Crameri, E.A. Whitehorn, E. Tate, W.P. Stemmer. *Nat. Biotechnol.* **1996**, *14*, 315.

D.A. Davis, A. Hamilton, J. Yang, L.D. Cremar, D. Van Gough, S. Potisek, M.T. Ong, P.V. Braun, T.J. Martínez, S.R. White, J.S. Moore, N.R. Sottos. *Nature* **2009**, *459*, 68.

A.-M. R. Dechert-Schmitt, D. C. Schmitt, X. Gao, T. Itoh, M. J. Krische, *Nat. Prod. Rep.* **2014**, *31*, 504.

B. A. DeChristopher, B. A. Loy, M. D. Marsden, A. J. Schrier, J. A. Zack, P. A. Wender. *Nat. Chem.* **2012**.

F. Del Vecchio, H. Petkovic, S. G. Kendrew, L. Low, B. Wilkinson, R. Lill, J. Cortés, B. A. M. Rudd, J. Staunton, P. F. Leadlay, *J. Ind. Microbiol. Biotechnol.* **2003**, *30*, 489.

DENZO-SMN. (1997). Otwinowski, Z. and Minor, W., Methods in Enzymology, **276**: Macromolecular Crystallography, part A, 307 – 326, C. W. Carter, Jr. and R. M. Sweets, Editors, Academic Press.

C.E. Diesendruck, B.D. Steinberg, N. Sugai, M.N. Silberstein, N.R. Sottos, S.R. White, P.V. Braun, J.S. Moore, *J. Am. Chem. Soc.* **2012**, *134*, 12446.

H. Dietz, M. Rief, G.H. Lorimer. **2004**, *101*, 16192.

Q. M. Dudley, A. S. Karim, M. C. Jewett, *Biotechnol. J.* **2015**, *10*, 69.

B. J. Dunn, D. E. Cane, C. Khosla, *Biochemistry* **2013**, *52*, 1839.

J. Dupont, C. Consorti, P. Suarez, R. de Souza, R. *Org. Synth.* **2004**, *10*, 184

A. K. El-Sayed, J. Hothersall, S. M. Cooper, E. Stephens, T. J. Simpson, C. M. Thomas, *Chem. Biol.* **2003**, *10*, 419.

P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr. D. Biol. Crystallogr.* **2010**, *66*, 486–501.

A. S. Eustáquio, R. P. McGlinchey, Y. Liu, C. Hazzard, L. L. Beer, G. Florova, M. M. Alhamadsheh, A. Lechner, A. J. Kale, Y. Kobayashi, et al., *Proc. Natl. Acad. Sci., U. S. A.* **2009**, *106*, 12295.

A. S. Eustáquio, S.-J. Nam, K. Penn, A. Lechner, M. C. Wilson, W. Fenical, P. R. Jensen, B. S. Moore, *ChemBioChem.* **2011**, *12*, 61.

A. S. Eustáquio, D. O'Hagan, B. S. Moore, *J. Nat. Prod.* **2010**, *73*, 378.

A.K. Evans, J.T. Batroli, L. Shih, L. *J. Am. Chem. Soc.* **1981**. *103*, 2127.

D. A. Evans, P. H. Carter, E. M. Carreira, A. B. Charette, J. A. Prunet, M. Lautens, *J. Am. Chem. Soc.* **1999**, *121*, 7540.

J.R. Gage, D.A. Evans. *Org. Syn.* **1993**, *68*, 83.

A. Garg, X. Xie, A. Keatinge-Clay, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.* **2014**, *136*, 10190.

D. C. Gay, G. Gay, A. J. Axelrod, M. Jenner, C. Kohlhaas, A. Kampa, N. J. Oldham, J. Piel, A. T. Keatinge-Clay, *Structure* **2014**, *22*, 444.

D. C. Gay, P. J. Spear, A. T. Keatinge-Clay, *ACS Chem. Biol.* **2014**, *9*, 2374.

G. Gay, D. T. Wagner, A. T. Keatinge-Clay, D. C. Gay, *Plasmid* **2014**, *76C*, 66.

D. Gay, Y.-O. You, A. Keatinge-Clay, D. E. Cane, *Biochemistry* **2013**, *52*, 8916.

T. W. Geders, L. Gu, J. C. Mowers, H. Liu, W. H. Gerwick, K. Håkansson, D. H. Sherman, J. L. Smith, *J. Biol. Chem.* **2007**, *282*, 35954.

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

R. Groote, L. van Haandel, R.P. Sijbesma. *J. Polym. Sci. Part A Polym. Chem.* **2012**, *50*, 4929.

X. Guo, T. Liu, C. R. Valenzano, Z. Deng, D. E. Cane, *J. Am. Chem. Soc.* **2010**, *132*, 14694.

A. M. Haapalainen, G. Meriläinen, R. K. Wierenga, *Trends Biochem. Sci.* **2006**, *31*, 64.

M. Häckh, M. Müller, S. Lüdeke, *Chemistry* **2013**, *19*, 8922.

S. Hanessian, *Total Synthesis of Natural Products: The "Chiron" Approach*, Pergamon Press, Oxford, **1883**.

S. Hanessian, *Pure Appl. Chem.* **1993**, *65*, 1189.

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

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

T. P. A. Hari, P. Labana, M. Boileau, C. N. Boddy, *ChemBioChem* **2014**, *15*, 2656.

A. D. Harper, C. B. Bailey, A. D. Edwards, J. F. Detelich, A. T. Keatinge-Clay, *ChemBioChem.* **2012**, *13*, 2200.

K. C. Harper, E. N. Bess, M. S. Sigman, *Nat. Chem.* **2012**, *4*, 366.

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

C. Hertweck, *Angew. Chem. Int.* **2009**, *48*, 4688.

C. Hertweck, A. Luzhetskyy, Y. Rebets, A. Bechthold, *Nat. Prod. Rep.* **2007**, *24*, 162.

V. Helms, T.P. Straatsma, J.A. McCammon. *J. Phys. Chem. B* **1999**, *103*.

C.R.Hickenboth, J.S. Moore, S.R. White, N,R, Sottos, J. Baudry, S.R. Wilson. *Nature*. **2007**, *446*.

L. Hilterhaus, A. Liese, *Adv. Biochem. Eng. Biotechnol.* **2007**, *105*, 133.

M. E. Horsman, T. P. A. Hari, C. N. Boddy, *Nat. Prod. Rep.* **2015**.

A. J. Hughes, A. T. Keatinge-Clay, *Chem. Biol.* **2011**, *18*, 165.

A. J. Hughes, J. F. Detelich, A. T. Keatinge-Clay, *MedChemComm.* **2012**, *3*, 956.

A. J. Hughes, M. R. Tibby, D. T. Wagner, J. N. Brantley, A. T. Keatinge-Clay, *ChemCom.* **2014**, *50*, 5276.

G. W. Huisman, J. Liang, A. Krebber, *Curr. Opin. Chem. Biol.* **2010**, *14*, 122.

G. W. Huisman, S. J. Collier, *Curr. Opin. Chem. Biol.* **2013**, *17*, 284.

International Tables for X-ray Crystallography (1992). Vol. C, Tables 4.2.6.8 and 6.1.1.4, A. J. C. Wilson, editor, Boston: Kluwer Academic Press.

K. Ishida, T. Lincke, C. Hertweck, *Angew. Chem.* **2012**, *51*, 5470.

J. R. Jacobsen, C. R. Hutchinson, D. E. Cane, C. Khosla, *Science* **2009**, *277*, 367.

D. Janssen, D. Albert, R. Jansen, R. Müller, M. Kalesse, *Angew. Chem.* **2007**, *46*, 4898.

A. Kirschning, F. Hahn, *Angew. Chem.* **2012**, *51*, 4012.

Z. A. Kasun, X. Gao, R. M. Lipinski, M. J. Krische, *J. Am. Chem. Soc.* **2015**, *137*, 8900.

A. Keatinge-Clay, *J. Mol. Biol.* **2008**, *384*, 941.

A. T. Keatinge-Clay, *Chem. Biol.* **2007**, *14*, 898.

A. T. Keatinge-Clay, *Nat. Prod. Rep.* **2012**, *29*, 1050.

A. T. Keatinge-Clay, R. M. Stroud, *Structure* **2006**, *14*, 737.

C. M. Kao, G. Luo, L. Katz, D. E. Cane, C. Khosla, *J. Am. Chem. Soc.* **1994**, *116*, 11612.

Z.S. Kean, A.L.Black, S.L. Ramirez S.L. Craig. *J. Polym. Sci. B. Polym. Phys.* **2012**, *50*, 3481.

G. E. Keck, Y. B. Poudel, T. J. Cummins, A. Rudra, J. A. Covel, *J. Am. Chem. Soc.* **2011**, *133*, 744.

L. Kellenberger, I. S. Galloway, G. Sauter, G. Böhm, U. Hanefeld, J. Cortés, J. Staunton, P. F. Leadlay, *ChemBioChem* **2008**, *9*, 2740.

C. Khosla, *J. Org. Chem.* **2009**, *74*, 6416.

C. Khosla, D. Herschlag, D. E. Cane, C. T. Walsh, *Biochemistry* **2014**, *53*, 2875.

C. Khosla, Y. Tang, A. Y. Chen, N. a Schnarr, D. E. Cane, *Annu. Rev. Biochem.* **2007**, *76*, 195.

C.M. Kingsbury, P.A. May, D.A. Davis, S.R. White J.S. Moore, N.R. Sottos. *J. Mater. Chem.* **2011**, *21*, 8381.

J. Kim, J. H. Chang, E.-J. Kim, K.-J. Kim, *Biochem. Biophys. Res. Commun.* **2014**, *443*, 783.

A. Klibanov, G. Samokhin, K. Martinek, I.V. Berezin. *Biochim. Biophys. Acta* **1976**, *438*, 1.

W. S. Knowles, *Acc. Chem. Res.* **1983**, *16*, 106.

I. Koryakina, J. B. McArthur, M. M. Draelos, G. J. Williams, *Org. Biomol. Chem.* **2013**, *11*, 4449.

S. J. Kodumal, K. G. Patel, R. Reid, H. G. Menzella, M. Welch, D. V Santi, *Proc. Natl. Acad. Sci., U. S. A.* **2004**, *101*, 15573.

B. Kusebauch, B. Busch, K. Scherlach, M. Roth, C. Hertweck, *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 1460.

S. Kapur, B. Lowry, S. Yuzawa, S. Kenthirapalan, A. Y. Chen, D. E. Cane, C. Khosla, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 4110.

D. H. Kwan, P. F. Leadlay, *ACS Chem. Biol.* **2010**, *5*, 829.

D. H. Kwan, Y. Sun, F. Schulz, H. Hong, B. Popovic, J. C. C. Sim-Stark, S. F. Haydock, P. F. Leadlay, *Chem. Biol.* **2008**, *15*, 1231.

D. H. Kwan, M. Tosin, N. Schläger, F. Schulz, P. F. Leadlay, *Org. Biomol. Chem.* **2011**, *9*, 2053.

T. Kawate, E. Gouaux. *Structure* **2006**, *14*, 673.

J.N. Lenhardt, M.T. Ong, R. Choe, C.R. Evenhuis, T.J. Martinez, S.L. Craig. *Science* **2010**, *329*, 1057.

D. Leonori, V. K. Aggarwal, *Acc. Chem. Res.* **2014**, *47*, 3174.

J. W.-H. Li, J. C. Vederas, *Science* **2009**, *325*, 161.

J. Li, D. Menche, *Synthesis*. **2009**, *2009*, 2293.

F. Lohr, I. Jenniches, M. Frizler, M. J. Meehan, M. Sylvester, A. Schmitz, M. Gütschow, P. C. Dorrestein, G. M. König, T. F. Schäberle, *Chem. Sci.* **2013**, *4*, 4175.

J. Longo, C. Yao, C. Rios, N. T. T. Chau, F. Boulmedais, J. Hemmerlé, P. Lavalle, S. M. Schiller, P. Schaaf, L. Jierry, *Chem. Comm.* **2015**, *51*, 232.

Y. Lu, S. K. Woo, M. J. Krische, *J. Am. Chem. Soc.* **2011**, *133*, 13876.

K. Makyła, M. Müller, C. Lörcher, S. Winkler, M.G. Nussbaumer, M. Eder, N. Bruns. *Adv. Mater.* **2013**, *25*, 2701.

A. F. Marsden, P. Caffrey, J. F. Aparicio, M. S. Loughran, J. Staunton, P. F. Leadlay, *Science*. **1994**, *263*, 378.

A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Crystallogr.* **2007**, *40*, 658.

M. H. Medema, K. Blin, P. Cimermancic, V. de Jager, P. Zakrzewski, M. A. Fischbach, T. Weber, E. Takano, R. Breitling, *Nucleic Acids Res.* **2011**, *39*, W339.

D. Menche, F. Arikan, O. Perlova, N. Horstmann, W. Ahlbrecht, S. C. Wenzel, R. Jansen, H. Irschik, R. Müller, *J. Am. Chem. Soc.* **2008**, *130*, 14234.

A. Mengel, O. Reiser, *Chem. Rev.* **1999**, *99*, 1191.

H. G. Menzella, J. R. Carney, Y. Li, D. V. Santi, *Appl. Environ. Microbiol.* **2010**, *76*, 5221.

H. G. Menzella, J. R. Carney, D. V Santi, *Chem. Biol.* **2007**, *14*, 143.

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

S. J. Mickel, G. H. Sedelmeier, D. Niederer, R. Daeffler, A. Osmani, K. Schreiner, M. Seeger-Weibel, B. Bérod, K. Schaer, R. Gamboni, *Org. Process Res. Dev.* **2004**, *8*, 92.

A. Milo, A. J. Neel, F. D. Toste, M. S. Sigman, *Science* **2015**, *347*, 737.

N. Moebius, C. Ross, K. Scherlach, B. Rohm, M. Roth, C. Hertweck, *Chem. Biol.* **2012**, *19*, 1164.

J. Moldenhauer, X.-H. Chen, R. Borriss, J. Piel, *Angew. Chem.* **2007**, *46*, 8195.

J. Moldenhauer, D. C. G. Götz, C. R. Albert, S. K. Bischof, K. Schneider, R. D. Süssmuth, M. Engeser, H. Gross, G. Bringmann, J. Piel, *Angew. Chem.* **2010**, *49*, 1465.

R. Montaser, H. Luesch. *Future Med. Chem.* **2011**, *12*, 1475.

A.G.G. Moliterni, G. Polidori, and R.J. *Spagna Appl. Cryst.* 32, 115-119. A.G.G. Moliterni, G. Polidori, and R.J. *Spagna Appl. Cryst.* 32, 115.

J. D. Mortison, J.D. Kittendorf, D.H. Sherman, *J. Am. Chem. Soc.* **2009**, *131*, 15784.

M. L. Mugnai, Y. Shi, A. T. Keatinge-Clay, R. Elber, *Biochemistry* **2015**, *54*, 2346.

S. Murli, J. Kennedy, L. C. Dayem, J. R. Carney, J. T. Kealey, *J. Ind. Microbiol. Biotechnol.* **2003**, *30*, 500.

G. N. Murshudov, A. A. Vagin, E. J. Dodson, *Acta Crystallogr. D. Biol. Crystallogr.* **1997**, *53*, 240.

T. Nagai, K. Ibata, E.S. Park, M. Kubota, K. Mikoshiba, A. Miyawaki. *Nat. Biotechnol.* **2002**, *20*, 87.

T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck, J. Piel, *Nat. Biotechnol.* **2008**, *26*, 225.

H. Niwa.; S. Inouye. *Proc. Natl. Acad. Sci., USA* **1996**, *93*, 13617.

M. Oliynyk, M. J. B. Brown, J. Cortés, J. Staunton, P. F. Leadlay, *Chem. Biol.* **1996**, *10*, 833.

L. H. Østergaard, L. Kellenberger, J. Cortés, M. P. Roddis, M. Deacon, J. Staunton, P. F. Leadlay, *Biochemistry* **2002**, *41*, 2719.

Otwinowski, W. Minor, *Methods Enzymol.* **1997**, *276*, 307.

H. M. O'Hare, A. Baerga-Ortiz, B. Popovic, J. B. Spencer, P. F. Leadlay, *Chem. Biol.* **2006**, *13*, 287.

R. Perez-Jimenez, S. Garcia-Manyes, S.R.K. Ainavarapu, J.M. Fernandez. *J. Biol. Chem.* **2006**, *281*, 40010.

O. Perlova, K. Gerth, O. Kaiser, A. Hans, R. Müller, *J. Biotechnol.* **2006**, *121*, 174–191.

G. R. Pettit, C. L. Herald, D. L. Doubek, D. L. Herald, E. Arnold, J. Clardy, *J. Am. Chem. Soc.* **1982**, *104*, 6846.

S. Piasecki, A. Keatinge-Clay, *Synlett* **2012**, *23*, 1840.

S. K. Piasecki, C. A. Taylor, J. F. Detelich, J. Liu, J. Zheng, A. Komsoukaniants, D. R. Siegel, A. T. Keatinge-Clay, *Chem. Biol.* **2011**, *18*, 1331.

S. K. Piasecki, J. Zheng, A. J. Axelrod, M. E Detelich, A. T. Keatinge-Clay, *Proteins* **2014**, *82*, 2067.

J. Piel, *Nat. Prod. Rep.* **2010**, *27*, 996.

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

N. L. Pohl, M. Hans, H. Y. Lee, Y. S. Kim, D. E. Cane, C. Khosla, *J. Am. Chem. Soc.* **2001**, *123*, 5822.

P. Pöplau, S. Frank, B. I. Morinaka, J. Piel, *Angew. Chem.* **2013**, *52*, 13215.

S. Poust, A. Hagen, L. Katz, J. D. Keasling, *Curr. Opin. Biotechnol.* **2014**, *30C*, 32.

B.C. Raimundo, C.H. Heathcock. *Synlett*, 12, 1213.

R. Regentin, J. Kennedy, N. Wu, J. R. Carney, P. Licari, J. Galazzo, R. Desai, *Biotechnol. Prog.* **2004**, *20*, 122.

R. Reid, M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D. V Santi, C. R. Hutchinson, R. McDaniel, *Biochemistry* **2003**, *42*, 72.

A. Rivkin, T-. C. Chou, S. J. Danishefsky. *Angew. Chem.* **2005**, *44*, 2838.

A. Saghatelian, K.M. Guckian, D.A. Thayer, M.R. Ghadiri. *J. Am. Chem. Soc.* **2003**, *125*, 344.

D. E. Schaufelberger, M. P. Koleck, J. A. Beutler, A. M. Vatakis, A. B. Alvarado, P. Andrews, L. V. Marzo, G. M. Muschik, J. Roach, J. T. Ross, et al., *J. Nat. Prod.* **1991**, *54*, 1265.

K. Schneider, X.-H. Chen, J. Vater, P. Franke, G. Nicholson, R. Borriss, R. D. Süssmuth, *J. Nat. Prod.* **2007**, *70*, 1417.

B. Schetter, R. Mahrwald, *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 7506.

K. K. Sharma, C. N. Boddy, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 3034.

L. S. Sheehan, R. E. Lill, B. Wilkinson, R. M. Sheridan, W. A. Vousden, A. L. Kaja, G. D. Crouse, J. Gifford, P. R. Graupner, L. Karr, et al., *J. Nat. Prod.* **2006**, *69*, 1702.

G.M. Sheldric. (2008). SHELXL97. Program for the Refinement of Crystal Structures. *Acta Cryst,* A64, 112.

S. Sheng, S. Zhao, *Sheng Wu Gong Cheng Xue Bao* **2009**, *25*, 1601.

SIR97. (1999). A program for crystal structure solution. A. Altomare, M.C. Burla, M. Camalli, G.L. Cascarano, C. Giacovazzo, A. Guagliardi

V. Simunovic, R. Müller, *ChemBioChem* **2007**, *8*, 497.

A. P. Siskos, A. Baerga-Ortiz, S. Bali, V. Stein, H. Mamdani, D. Spiteller, B. Popovic, J. B. Spencer, J. Staunton, K. J. Weissman, et al., *Chem. Biol.* **2005**, *12*, 1145.

Y.K. Song, K.H. Lee, W.S. Hong, S.Y. Cho, H.C. Yu, C.M. Chung. *J. Mater. Chem.* **2012**, *22*, 1380.

A.L. Spek, (1998). PLATON, A Multipurpose Crystallographic Tool. Utrecht University, The Netherlands.

U. Sundermann, K. Bravo-Rodriguez, S. Klopries, S. Kushnir, H. Gomez, E. Sanchez-Garcia, F. Schulz, *ACS Chem. Biol.* **2013**, *8*, 443.

K. Sung, S.Y. Wu. *Synth. Comm.* **2001**, 31, 3069.

A. Sawano. *Nucl, Acids Res.* **2000**, *28*, 78e.

P. D. Straight, M. A. Fischbach, C. T. Walsh, D. Z. Rudner, R. Kolter, *Proc. Natl. Acad. Sci., U. S. A.* **2007**, *104*, 305.

C. M. Theodore, B. W. Stamps, J. B. King, L. S. L. Price, D. R. Powell, B. S. Stevenson, R. H. Cichewicz, *PLoS One*. **2014**, *9*, e90124.

B. M. Trost, G. Dong, *Nature*. **2008**, *456*, 485.

S. C. Tsai, L. J. Miercke, J. Krucinski, R. Gokhale, J. C. Chen, P. G. Foster, D. E. Cane, C. Khosla, R. M. Stroud, *Proc. Natl. Acad. Sci., U. S. A.* **2001**, *98*, 14808.

R.Y. Tsien. *Annu. Rev. Biochem.* **1998**, *67*, 509.

R. Ueoka, A. R. Uria, S. Reiter, T. Mori, P. Karbaum, E. E. Peters, E. J. N. Helfrich, B. I. Morinaka, M. Gugger, H. Takeyama, et al., *Nat. Chem. Biol.* **2015**, *11*, 705.

C. R. Valenzano, Y.-O. You, A. Garg, A. Keatinge-Clay, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.* **2010**, *132*, 14697.

M. C. Walker, B. W. Thuronyi, L. K. Charkoudian, B. Lowry, C. Khosla, M. C. Y. Chang, *Science* **2013**, *341*, 1089.

W.W. Ward, C.W. Cody, R.C. Hart, M.J. Cormier. *Photochem. Photobiol.* **1980**, *31*, 611.

T. Weber, K. Blin, S. Duddela, D. Krug, H. U. Kim, R. Bruccoleri, S. Y. Lee, M. A. Fischbach, R. Muller, W. Wohlleben, et al., *Nucl. Acids Res.* **2015**, *43*, W237.

T. Weber, K. J. Laiple, E. K. Pross, A. Textor, S. Grond, K. Welzel, S. Pelzer, A. Vente, W. Wohlleben, *Chem. Biol.* **2008**, *15*, 175.

K. E. H. Weismann, J. Cortes, M. J. B. Brown, A. L. Cutter, J. Staunton, P. F. Leadlay, *Chem. Biol.* **1995**, *2*, 583.

K. J. Weissman, *ChemBioChem* **2006**, *7*, 485.

K. J. Weissman, *ChemBioChem* **2006**, *7*, 1334.

K. J. Weissman, *Nat Chem Biol* **2014**, *11*, 660.

K. J. Weissman, M. Timoney, M. Bycroft, P. Grice, U. Hanefeld, J. Staunton, P. F. Leadlay, *Biochemistry*. **1997**, *36*, 13849.

M. Werneburg, B. Busch, J. He, M. E. A. Richter, L. Xiang, B. S. Moore, M. Roth, H.-M. Dahse, C. Hertweck, *J. Am. Chem. Soc.* **2010**, *132*, 10407.

K.M. Wiggins, J.N. Brantley, C.W. Bielawski. *Chem. Soc. Rev.* **2013**, *42*, 7130.

J. Willwatcher, N. Kausch-Busies, A. Fürstner. *Angew. Chem.* **2012**, *48*, 12041.

K. E. Wilson, J. E. Flor, R. E. Schwartz, H. Joshua, J. L. Smith, B. A. Pelak, J. M. Liesch, O. D. Hensens, *J. Antibiot. (Tokyo)*. **1987**, *40*, 1682.

WinGX 1.64. (1999). An Integrated System of Windows Programs for the Solution, Refinement and Analysis of Single Crystal X-ray Diffraction Data. Farrugia, L. J. *J. Appl. Cryst.* 32. 837.

M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, et al., *Acta Crystallogr. D. Biol. Crystallogr.* **2011**, *67*, 235.

R. B. Woodward, *Perspectives in Organic Chemistry*, Wile-Interscience, New York, **1956**.

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

F. T. Wong, X. Jin, I. I. Mathews, D. E. Cane, C. Khosla, *Biochemistry* **2011**, *50*, 6539.

J. Wu, K. Kinoshita, C. Khosla, D. E. Cane, *Biochemistry* **2004**, *43*, 16301.

J. Wu, T. J. Zaleski, C. Valenzano, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.* **2005**, *127*, 17393.

J. Yan, S. Gupta, D. H. Sherman, K. A. Reynolds, *Chembiochem* **2009**, *10*, 1537.

Y. Yin, R. Gokhale, C. Khosla, D. E. Cane, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1477.

M. J. Yu, W. Zheng, B. M. Seletsky, *Nat. Prod. Rep.* **2013**, *30*, 1158.

J. Zhan, *Curr. Top. Med. Chem.* **2009**, *9*, 1958.

L. Zhang, T. Mori, Q. Zheng, T. Awakawa, Y. Yan, W. Liu, I. Abe, *Angew. Chem.* **2015.**

J. Zheng, D. C. Gay, B. Demeler, M. A. White, A. T. Keatinge-Clay, *Nat. Chem. Biol.* **2012**, *8*, 615.

J. Zheng, A. T. Keatinge-Clay, *J. Mol. Biol.* **2011**, *410*, 105.

J. Zheng, A. T. Keatinge-Clay, *Medchemcomm* **2012**. *4*, 34.

J. Zheng, S. K. Piasecki, A. T. Keatinge-Clay, *ACS Chem. Biol.* **2013**, *8*, 1964.

J. Zheng, C. A. Taylor, S. K. Piasecki, A. T. Keatinge-Clay, *Structure*. **2010**, *18*, 913.

C. Zhao, J. M. Coughlin, J. Ju, D. Zhu, E. Wendt-Pienkowski, X. Zhou, Z. Wang, B. Shen, Z. Deng, *J. Biol. Chem.* **2010**, *285*, 20097.

L. Zhao, Y. Li, X. Cao, J. You, W. Dong. *Nanotechnology* **2012**, *23*, 255702.

H. Zhao, W. A. van der Donk, *Curr. Opin. Biotechnol.* **2003**, *14*, 583-589.

Y. Zhu, K. Burgess, *J. Am. Chem. Soc.* **2008**, *130*, 8894.

S. B. Zimmerman, C. D. Schwartz, R. L. Monaghan, B. A. Pelak, B. Weissberger, E. C. Gilfillan, S. Mochales, S. Hernandez, S. A. Currie, E. Tejera, *J. Antibiot. (Tokyo).* **1987**, *40*, 1677.

Vita

Constance Beryl Bailey was born in San Francisco, CA and raised in Washington, D.C. and Wilmette, IL (a suburb of Chicago). She attended Reed College in Portland, OR from 2006-2010 receiving a B.A. in biochemistry. As an undergraduate, she pursued a variety of research directions including: at the University of Illinois-Chicago Medical Center in the division of surgery (under the mentorship of Prof. Jose Oberholzer, M.D.) working on translational research related to pancreatic islet cells, at Reed College (under the mentorship of Prof. Patrick McDougal) pursuing organic synthesis, and at the Hans Knoell Institute in Jena, Germany (under the mentorship of Prof. Christian Hertweck) investigating natural products biosynthesis (specifically, polyketide biosynthesis). Inspired by her work at the Hans Knoell Institute, she instigated a collaboration between Prof. Arthur Glasfeld at Reed College and Prof. Kevin Reynolds at Portland State University for her undergraduate thesis in the field of natural products biosynthesis. For graduate school, a natural direction was to continue in the area of polyketide biosynthesis in the laboratory of Prof. Adrian Keatinge-Clay. In addition to her research pursuits, she has a commitment to teaching and scientific outreach. At Reed College, she was involved in the Reed Biology Outreach program, working with elementary school teachers to augment the curriculum. At the University of Texas, she received a GK-12 fellowship that collaborated with two teachers in the Austin Independent School district to enrich middle and high school curriculums.

Permanent email: ConstanceBBailey@gmail.com

This dissertation was typed by Constance B. Bailey.