ENGINEERING OF FUNGAL AROMATIC POLYKETIDE BIOSYNTHESIS
AND
ELUCIDATION OF CERCOSPORIN BIOSYNTHESIS IN THE PLANT PATHOGEN *CERCOSPORA NICOTIANAE*

by
Adam G. Newman

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy in Chemical Biology

Baltimore, Maryland
January 2016

Abstract

In fungi, the iterative, non-reducing polyketide synthases (NR-PKSs) are responsible for the biosynthesis of aromatic polyketide products. While the modular type I PKSs have been extensively studied for 20 years, the biochemistry of NR-PKS is only now beginning to be elucidated. The NR-PKSs share a common domain architecture that is intrinsically linked to their function. The mode of biosynthesis is analogous to that of fatty acids by animal fatty acid synthases (FAS), but simplified. The three N-terminal domains, the starter unit:acyl-carrier protein transacylase (SAT), ketosynthase (KS) and malonyl acyl transferase (MAT) domains are responsible for the initiation and polyketide elongation phases. The SAT domain selects a precursor or starter unit substrate as an acyl thioester while the MAT domain introduces ketide extender units from malonyl-CoA. The KS works in collaboration with the acyl-carrier protein (ACP) to catalyze the decarboxylative Claisen condensation of these substrates generating a linear, ACP-bound β-ketone intermediate. The C-terminal domains of NR-PKSs control the final stage of biosynthesis, which includes regiospecific aldol cyclizations/aromatizations by the product template (PT) domain and product release by the thioesterase (TE) domain. In this way, the four factors governing chemical diversity in aromatic polyketides are entirely controlled by the enzyme, with the N-terminal half determining starter unit selection and chain length, and the C-terminal half controlling the cyclization pattern and mode of product evolution. Our lab has innovated the use of enzyme deconstruction in NR-PKSs to determine the catalytic program of these enzymes. The enzyme deconstruction approach requires the dissection of an individual NR-PKS into its constituent domains or multidomain fragments. Mono- and multidomain fragments are

expressed and purified individually and then recombined *in vitro*, reconstituting wild-type activity.

Using the enzyme deconstruction approach, the catalytic activity of CTB1—the NR-PKS of cercosporin biosynthesis in the fungal plant pathogen *Cercospora nicotianae*—was determined. The CTB1 TE domain was demonstrated to catalyze an unprecedented enol-lactonization to form the naphthopyrone *nor*-toralactone—representing an expansion of known TE chemistry. The formation of *nor*-toralactone was unexpected and in conflict with the accepted cercosporin biosynthetic pathway. Using a combination of gene knockout strains and *in vitro* enzymology, a new cercosporin biosynthetic pathway was identified. Of particular interest was the activity of an unusual didomain enzyme CTB3. The flavin-dependent monooxygenase domain of CTB3 was identified to catalyze a unique oxidative aromatic ring cleavage—expanding the chemical repertoire of these ubiquitous proteins. Enzyme deconstruction served as the basis of an NR-PKS engineering project in which homologous domains from non-cognate parent NR-PKSs were swapped *in vitro* to produce non-native polyketide products. Using a systematic approach to domain-swapping, several rules governing rational engineering of NR-PKS were codified. Domain-swapping of deconstructed NR-PKSs facilitated rational engineering of intact, chimeric NR-PKSs. A library of chimeric NR-PKSs was prepared and selected members were assayed for activity. Combining the rules of engineering gleaned from deconstructed NR-PKS domain-swapping and the construction of the chimeric NR-PKS library, an attempt to rationally design a topopyrone synthase was unsuccessfully made. By examining the activities of individual domains and perturbing the composition of NR-PKSs, we have made invaluable insights into the NR-PKS

catalytic program—understanding that will underpin future efforts towards rational NR-PKS engineering.

Thesis Advisor:

    Dr. Craig A. Townsend, Department of Chemistry, Johns Hopkins University

Additional Readers:

    Dr. Caren L. Freel Meyers, Department of Pharmacology and Molecular Sciences,

        Johns Hopkins University

    Dr. Rebekka Klausen, Department of Chemistry, Johns Hopkins University

Dedicated to my parents

Jonathan and Sherry Newman

Dwarves' tongues run on when speaking of their handiwork, they say.

—J.R.R. Tolkien, *The Lord of the Rings*—

Acknowledgments

No man is an island. The completion of any Ph.D. requires the mentorship, support, assistance, commiseration, motivation, demotivation, and friendship of a small army. Although not an exhaustive list, I would like to take a moment here to acknowledge the contributions of some very important people without whom I would not have made it so far.

Words are poor substitutes for expressing my deepest thanks and admiration for Prof. Craig A. Townsend. Prof. Townsend provided a scientific home for me at Johns Hopkins. His steadfast mentorship and wise leadership helped to shape me into the scientist I am today. There are few people with as wide a breadth of knowledge, as disciplined a work ethic, and as faithful a commitment to his students. I will be forever in his debt and can only hope to be half the scientist he is.

I have had many colleagues in the Townsend laboratory, but none as helpful and dedicated as Team PKS. In particular I must thank Dr. Anna L. Vagstad who was my student mentor when I first joined the group. Anna's diligence, optimism, and intelligence made working on Team PKS a pleasure. Both Mr. Philip A. Storm and Ms. Callie R. Huitt-Roehl have been talented and thoughtful collaborators. I look forward to following their careers and know that great things lie ahead for both of them. Dr. Katherine Belecki was always a helpful sounding board and generous colleague. I thank her for several trips to the Medical Campus to conduct LCMS runs for me. Both Prof. Jonathan R. Scheerer and Dr. Eric A. Hill provided me with synthetic standards, for which I am very grateful. I learned much from both Dr. Martina M. Adams and Dr. Jason W. Labonte, who were instrumental to my growth as a scientist. In addition to these

Team PKS member, I must thank my lab mates in Remsen 250—Dr. Rongfeng Li, Dr. Andrew R. Buller, Dr. Jeanne M. Davidsen, Dr. Michael F. Freeman, Ms. Michel Lau, and Mr. Douglas Cohen—for being such a great group of scientists with which to share a laboratory.

I have had the opportunity to collaborate with some truly fascinating scientists in my time at Johns Hopkins. Chief among them is Prof. Sheryl Tsai for UC Irvine (Irvine, CA) with whom we have investigated the structural biology of NR-PKSs. Prof. Kuang-Ren Chung of the National Hsing Chung University (Taichung City, Taiwan) provided me with *Cercospora nicotianae* CTB cluster knockout strains without which the cercosporin project would have gone nowhere. The late Prof. Robert Cotter and his student Dr. Katherine Feilder of Johns Hopkins University (Baltimore, MD) allowed us access to their MS facilities which greatly accelerated progress on our engineering efforts.

I wish to thank the members of my committee including Prof. Caren L. Freel Meyers, Prof. Rebekka S. Klausen, and Prof. R. Blake Hill.

When I first joined the CBI Program at Johns Hopkins, it was still in its infancy. I thank Prof. Marc Greenberg and his successor Prof. Steve Rokita for their service as CBI Program Director. Through their leadership, the CBI Program has flourished and I look forward to following the progress of its many members. Instrumental to the success of the CBI Program is the tireless efforts of Ms. Lauren McGhee, our administrator. Lauren is the unseen force keeping the entire circus together. I thank her for this and for letting me complain to her about the minor injustices of the world.

The support staff of the Chemistry Department at Johns Hopkins often goes overlooked despite its indispensable role. Thank you to Ms. Jean Goodwin, Ms. Rosalie Elder, Mr. John Kidwell, Mr. Joseph Russell, Mr. Scott McGhee, Mr. Jordan Patterson, Mr. Boris Steinberg, and Mr. David Brewster. I also thank Dr. Phil Mortimer for keeping the MS facility running—a service without which I would have never completed my Ph.D.—and Dr. Cathy Moore and Dr. Joel Tang for maintaining our NMR facility.

Two former teachers were instrumental in my decision to pursue a career in science. Prof. Julie T. Millard of Colby College (Waterville, ME) was my first true scientific advisor. She taught me the value of discipline and careful experimental design. I thank her for her loyal support and instruction. Dr. Fred Lamb of Chariho Regional High School (Wood River Junction, RI) was my first chemistry teacher. He set me on the path that eventually brought me to Johns Hopkins and the completion of this Ph.D. I thank him.

Finally, I thank my two sisters Rachel and Rebecca and the two most important people in my life: mom and dad. They have always been my biggest fans, even when I ramble on about science that they do not understand. Believe it when I say that—without you—not a single word of this dissertation would be possible. I love you.

Publications

Newman A.G. and Townsend, C.A. (2015). Molecular Characterization of the Cercosporin Biosynthetic Pathway in the Fungal Plant Pathogen *Cercospora nicotianae*. *Manuscript in Preparation*.

J.F. Barajas, K. Finzel, T.R. Valentic, G. Shakya, N. Gamarra, D. Martinez, J.L. Meier, A.L. Vagstad, **A.G. Newman**, C.A Townsend, M.D. Burkart, S.C. Tsai (2015) Structural and biochemical analysis of protein-protein interactions between the acyl carrier protein and the product template domains. *Under Review*.

Newman, A.G., Vagstad, A.L., Storm, P.A., and Townsend, C.A. (2014). Systematic Domain Swaps of Iterative, Nonreducing Polyketide Synthases Provide a Mechanistic Understanding and Rationale For Catalytic Reprogramming. *J. Am. Chem. Soc. 136*, 7348-7362.

Vagstad, A.L., **Newman, A.G.,** Storm, P.A., Belecki, K., Crawford, J.M., and Townsend, C.A. (2013). Combinatorial domain swaps provide insights into the rules of fungal polyketide synthase programming and the rational synthesis of non-native aromatic products. *Angew. Chem. Int. Ed. Engl. 52*, 1718-1721.

Newman, A.G., Vagstad, A.L., Belecki, K., Scheerer, J.R., and Townsend, C.A. (2012). Analysis of the cercosporin polyketide synthase CTB1 reveals a new fungal thioesterase function. *Chem. Commun. 48*, 11772-11774.

Korman, T.P., Crawford, J.M., Labonte, J.W., **Newman, A.G.,** Wong, J., Townsend, C.A., and Tsai, S.C. (2010). Structure and function of an iterative polyketide synthase thioesterase domain catalyzing Claisen cyclization in aflatoxin biosynthesis. *Proc. Natl. Acad. Sci. USA 107*, 6246-6251.

LaRiviere, F.J., **Newman, A.G.,** Watts, M.L., Bradley, S.Q., Juskewitch, J.E., Greenwood, P.G., and Millard, J.T. (2009). Quantitative PCR analysis of diepoxybutane and epihalohydrin damage to nuclear versus mitochondrial DNA. *Mutat. Res., Fundam. Mol. Mech. Mutagen. 664*, 48-54.

Romano, K.P., **Newman, A.G.,** Zahran, R.W., and Millard, J.T. (2007). DNA interstrand cross-linking by epichlorohydrin. *Chem. Res. Toxicol. 20*, 832-838.

Table of Contents

List of Tables

List of Figures

Chapter 1: An Introduction to Polyketide Biosynthesis

1.1. Introduction
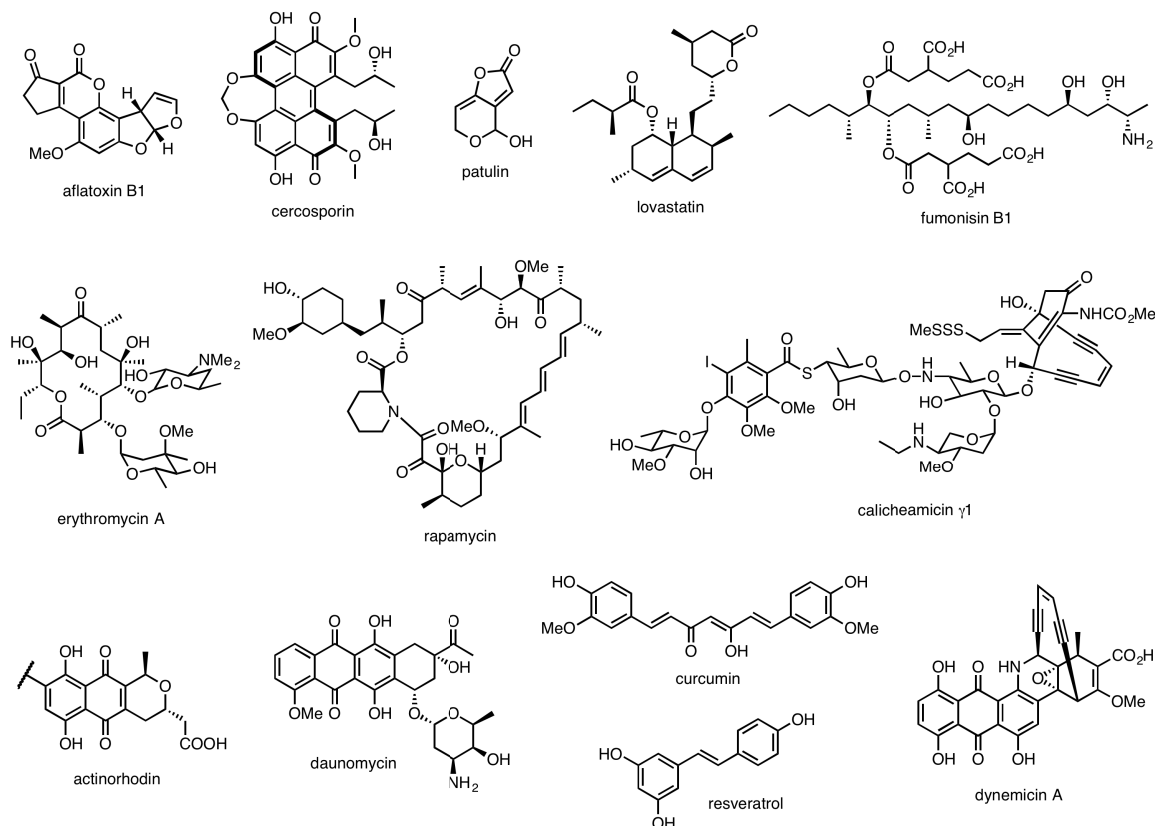
Figure 1.1 | Representative sample of polyketide natural products.

The polyketides are an astounding class of natural products encompassing a structurally and functionally diverse set of metabolites (Figure 1.1). Polyketides are produced by an immense array of organisms including bacteria, fungi, and plants occupying an incalculable number of ecological niche environments. Because of their structural variability, it is not surprising that they display a huge range of biological activities and thus represent some of the most commercially important natural products.[1] They serve as potent pharmaceuticals including the hugely successful cholesterol-lowering agent lovastatin, the immunosuppressant rapamycin, the anti-cancer agent
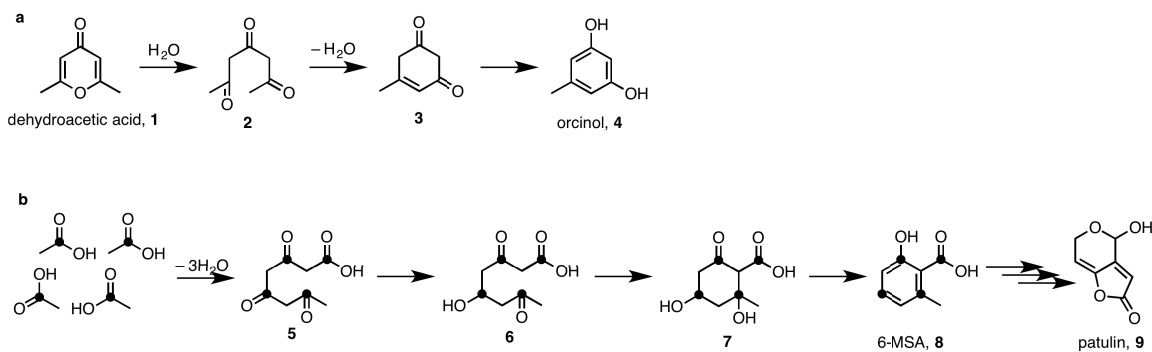
Figure 1.2 │ Earliest model systems for polyketide biosynthesis. (a) Collie's synthesis of orcinol (**4**) from dehydroacetic acid (**1**) implied a triketone intermediate (**2**). (b) Using a radiolabel tracer assay, Birch demonstrated that 6-MSA (**8**) was the product of head-to-tail Claisen condensation of four acetate units. Dots indicate positions of radioisotope ^{14}C incorporation. 6-MSA is converted to the phytotoxin patulin (**9**).

daunomycin, and the antibiotic erythromycin. Many agriculturally destructive fungal mycotoxins are of polyketide origin including aflatoxin B1, cercosporin, patulin, and fumonisin B1. Still others can act as food coloring additives (*e.g.* curcumin) while others show great promise as potential therapeutic agents (*e.g.* dynemicin A). Understanding the biosynthetic origin of polyketide structural variation is of critical importance for new product discovery and rational engineering of polyketide pathways, especially considering the rapid growth of fully sequenced microbial genomes. The biosynthesis of polyketides has been reviewed elsewhere.[2] Here I present an abbreviated primer for understanding polyketide biosynthesis.

Despite their structural diversity and widespread source organisms, the polyketides share a common biosynthetic origin.[3] The carbon skeleton of all polyketide products derives from simple carboxylic acids and their derivatives with acetate serving as the standard precursor. In this way, polyketide biosynthesis is analogous to fatty acid biosynthesis from primary metabolism. In fact, the enzymes of polyketide biosynthesis are often homologous to their counterparts from fatty acid biosynthesis. As early as 1955,

Arthur Birch recognized the connection between polyketide biosynthesis and fatty acid biosynthesis despite the limited knowledge of either pathway at the time.[4]

As with fatty acid biosynthesis, the central process of polyketide biosynthesis is carbon–carbon bond formation between acetate units. Birch hypothesized that—like fatty acids—repeated condensation of acetate units could generate a "poly-β-ketone," which—in turn—could serve as the reactive intermediate for aromatic natural products (Figure 1.2b). Birch's hypothesis was indeed a rediscovery of an earlier idea proposed by James Collie in 1893.[5] Collie was attempting to prove the structure of dehydroacetic acid (**1**) through degradation. When boiling the acid in barium hydroxide produced the aromatic orcinol (**4**), Collie happened upon a crucial finding. He proposed that orcinol production from dehydroacetic acid proceeded through a polyketone intermediate (**2**) and that such transformations could account for similar biological aromatic compounds—a bold claim for the time (Figure 1.2a). Although his contemporaries rejected this hypothesis, we now know it to be true and the foundation of polyketide biosynthesis.

Confirmation of the polyketone hypothesis was provided by Birch in an elegant radioisotope-labeling experiment.[4] Birch chose 6-methyl-salicylic acid (6-MSA, **8**) as his test subject—a polyketide produced by the fungus *Penicillium patulum* as an intermediate in the biosynthesis of patulin (**9**).[6] By feeding the fungus acetate—the proposed precursor metabolite—labeled with ^{14}C at the C1 position, the pattern of radioactive isotopes incorporation in 6-MSA could be used to determine the metabolite's biosynthetic origin. The observed labeling pattern in 6-MSA suggested that the metabolite derived from the "head-to-tail" linkage of four acetate units producing a triketo acid (**5**, Figure 1.2b). Reduction of the C5 ketone of intermediate **5** followed by an
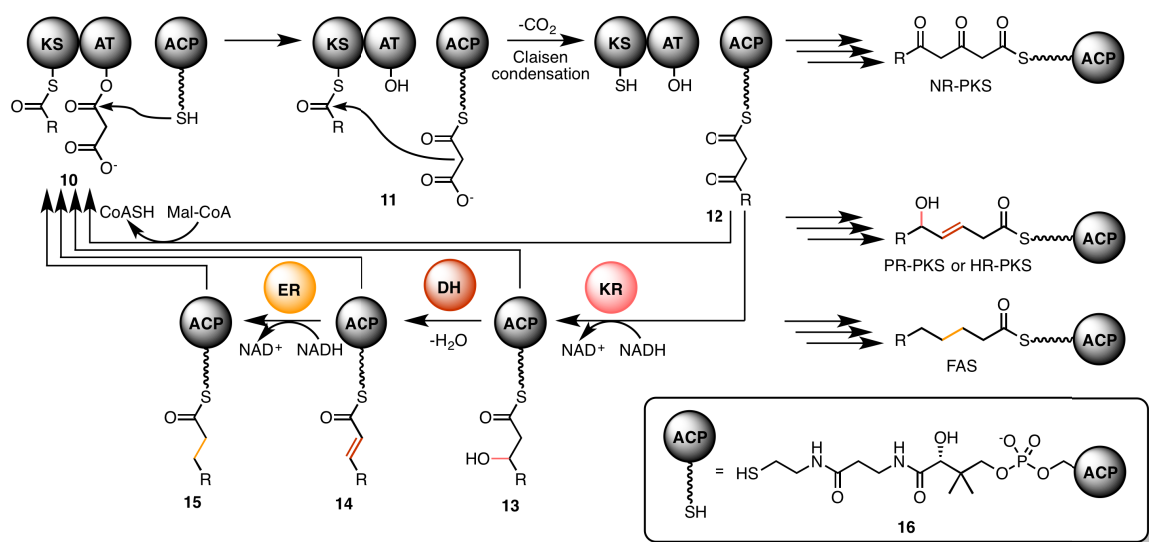
Figure 1.3 │ General mechanism for polyketide biosynthesis by PKSs. The central mechanism of polyketide biosynthesis is repeated Claisen condensations of acetate units catalyzed by the KS. The AT domain introduces malonyl extender units from malonyl-CoA. The ACP tethers the growing poly-β-ketone intermediate via a thioester linkage to a phosphopantetheine modification (**16**). Following each round of extension, the PKS can process the β-position through the action of KR, DH, and ER domains. β-position processing determines the catalytic program.

intramolecular aldol condensation and two dehydrations would result in 6-MSA with the observed labeling pattern.

Birch's findings were groundbreaking and radioisotope-labeled acetate feeding assays became the norm in the natural products community. Labeling assays became even more powerful with the advent of NMR and MS[7]—both techniques that accelerated the rate of discovery in the field. At long last, the Collie-Birch polyketide hypothesis gained general acceptance and not just for aromatic natural products. It was soon realized that the repeated condensation of acetate units accounted for the biosynthesis of a huge class of natural products—the polyketides.

1.2. *Polyketide Synthases*

Polyketide synthases (PKSs) are the enzymes responsible for the initial phases of polyketide natural product formation by setting the carbon skeleton of the eventual

metabolite (Figure 1.3).[2a, 8] The PKSs are homologous in both form and function to the fatty acid synthases (FASs) of primary metabolism. Repeated acetate homologation to form the polyketone intermediate is the central reaction of polyketide biosynthesis and is the primary function of the PKSs. This reaction is accomplished through decarboxylative Claisen condensation between an acyl thioester and a malonyl thioester (**11**)—usually derived from malonyl-CoA. The reaction is universally catalyzed by a β-ketoacyl synthase (KS)—the indispensible feature of all FASs and PKSs. As with FASs, the malonyl extender unit and growing polyketone intermediate are covalently tethered to an acyl carrier protein (ACP) via a thioester linkage (**12**). The thioester tether is provided by the terminal thiol of a phosphopantetheine (PPT) post-translational modification to a critical Ser residue on the ACP (**16**). The final essential feature of most PKSs is the acyl transferase (AT)—an enzyme responsible for the transfer of an acyl extender unit from CoA to the ACP via a transient AT-linked acyl ester (**10**).

As in fatty acid biosynthesis, the PKSs can further process the nascent β-keto-ACP thioester (**12**) through serial reduction at the β-ketone. Initial reduction at this position is catalyzed by a ketoreductase (KR), resulting in a secondary alcohol (**13**). Dehydration of this alcohol to form an α,β unsaturated thioester (**14**) is catalyzed by the dehydratase (DH). Finally, reduction of the unsaturated intermediate **14** by the enoylreductase (ER) generates a fully saturated thioester (**15**). Following reductive processing, transthioesterification from the ACP to the active site Cys of the KS initiates another round of acetate extension. Product release in the FASs is catalyzed by a thioesterase (TE), which catalyzes the hydrolysis of the acyl-ACP thioester to generate the free acid. TEs exist in PKSs where they serve the same function of product release,

5

but often with expanded chemistry. Unlike FASs, PKSs can employ any number of

reductive steps or forego any reductive transformations to generate intermediates of great

chemical complexity. Additionally, they can incorporate the activity of *C*-

methyltransferases (CMeTs) to methylate the α-position of the β-keto-ACP thioester (**12**).

This unique feature of PKSs is termed programing. Programing controls not only the

degree of reduction, but also the chain length and mechanism of product release. The

structural basis for programing is dependent upon the type of PKS—detailed below—and

in some cases is still unknown.

The FASs fall into two classes named type I and type II. The type I FASs are

large, multifunctional enzymes in which a single protein—or pair of proteins that work in

complex—contains all of the necessary catalytic functionalities as individual domains

(KS-AT-DH-ER-KR-ACP-TE).[9] Type I FASs are found in both fungi and animals—

although their architectures differ. Crystal structures of both fungal[10] and animal[11]

FASs have been solved revealing significant structural differences commensurate with

their divergent domain architecture. The type II FASs are complexes of many single

proteins each with its particular function.[12] The individual proteins of type II FASs work

in trans to accomplish fatty acid biosynthesis. Type II FASs are found in bacteria and

plants. Generally, the individual members of type II FASs are structurally similar to their

homologous domains in type I FASs. Both type I and type II FASs function iteratively—

meaning active sites are reused during the catalytic cycle.

PKSs are also classified into types corresponding to the homologous FAS types.

Type I PKSs are large, multidomain proteins where each domain catalyzes a specific

transformation in polyketide formation.[2a, 3a] Type I PKSs can be further classified as

modular[2d] or iterative.[13] The modular PKSs function in a fashion analogous to an assembly line. Unlike type I FASs where active sites are reused during the catalytic cycle, in the modular PKSs a dedicated module containing the necessary functional domains catalyzes each individual acetate extension and subsequent β-ketone processing. The growing polyketide intermediate is then passed to the next module in the assembly line, which catalyzes its prescribed extension and β-ketone processing. In this way, each domain of these gigantic megasynthases is responsible for only a single reaction. The iterative type I PKSs are structurally simpler, but retain the more complex iterative functionality observed in their type I FAS homologs. Like their FAS counterparts, type II PKSs contain many individual proteins that work in complex to catalyze iterative polyketide formation.[2c, 14] A third type—type III PKSs—has no FAS homolog.[15] Instead, these enzymes function as stand-alone KSs without the typical AT, ACP, or TE functionalities. Type I PKSs are found in bacteria and fungi. Type II PKSs are only found in bacterial. Type III PKSs are found in bacteria, fungi, and plants.

David Hopwood—a venerated natural products researcher—conducted much of
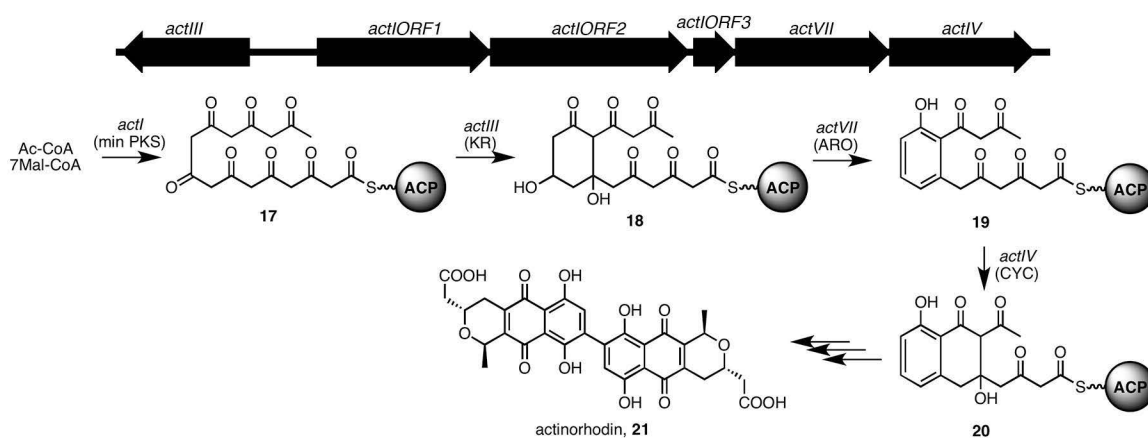
Figure 1.4 | Actinorhodin is the primary model system for type II PKS biosynthesis. The central actinorhodin gene cluster from *S. coelicolor* is shown. Each member of the PKS exists as its own, stand-alone protein. These proteins work in complex to carry out the biosynthesis of actinorhodin.

7

the early work in identifying the PKSs as the central enzymes of polyketide biosynthesis in the 1980s.[16] His model system was actinorhodin biosynthesis in *Streptomyces coelicolor* (Figure 1.4). Actinorhodin (**21**)—an antibiotic of the isochromanequinone class—was chosen due to its distinctive blue pigmentation. Through careful inactivation of specific *S. coelicolor* genes, Hopwood was able identify those responsible for the cessation of pigmentation and—therefore—actinorhodin biosynthesis. Gene sequencing—a challenging task at the time—identified several type II FAS homologs establishing the functional link between fatty acid biosynthesis and polyketide biosynthesis as well as identifying the first type II PKS. Three open reading frames were eventually identified as encoding the type II minimal PKS.[17] The protein products of these genes comprised a KS heterodimer and ACP—the only domains necessary for successful acetate extension. Downstream enzymes included a KR, aromatase (ARO), and cyclase (CYC), which work in tandem to catalyze regiospecific reduction and cyclization of the poly-β-ketone directing the intermediate toward actinorhodin.[18] It was later discovered that the minimal PKS is a universal feature of type II PKSs and not only conducts acetate extension but also controls the final chain length of the polyketide product. As seen with actinorhodin, downstream "tailoring" proteins—*e.g.* ARO, CYC, KR, CMeT—exert control over polyketide programming through regiospecific catalysis.
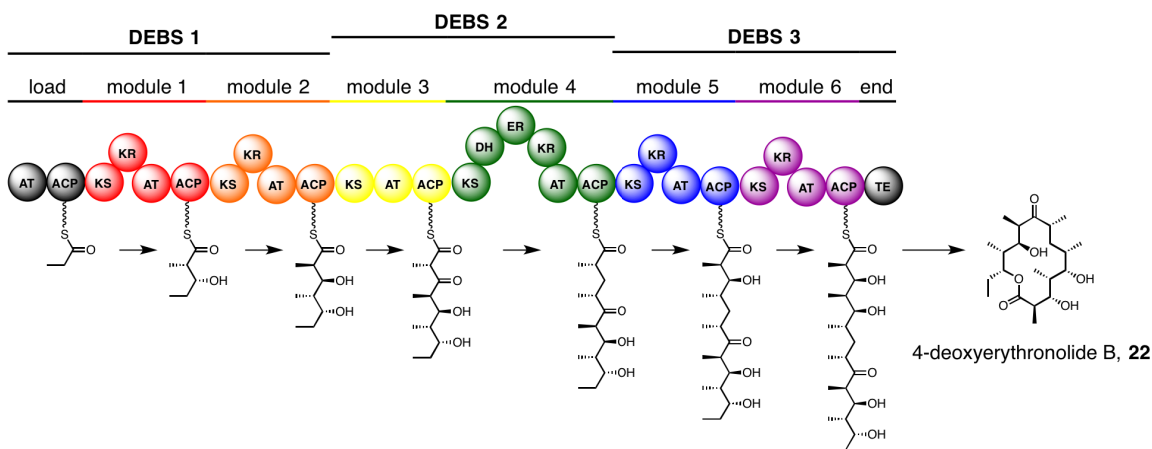
Figure 1.5│The 6-deoxyerythronolide B synthase is the model system for modular type I PKS. The modular type I PKSs are huge megasynthases where each domain is used only once in the production of the polyketide product. Sets of domains are organized into modules responsible for a single acetate extension and subsequent β-position processing. Biosynthesis is linear with each module passing the growing intermediate to the next making biosynthesis analogous to an assembly line.

Following Hopwood's success with actinorhodin, other laboratories entered the fray searching for still more PKSs. The groups of Peter Leadlay[19] and Leonard Katz[20] independently—and using complementary methods—discovered the genes responsible for the formation of 6-deoxyerythronolide B (6-dEB, **22**)—the polyketide precursor to the antibiotic erythromycin A from *Saccharopolyspora erythraea* (Figure 1.5). To their surprise, the PKS was encoded in three enormous open reading frames—*eryAI, eryAII, and eryAIII*—each producing a large, multidomain protein—DEBS 1, 2, and 3, respectively—together called the 6-deoxyerythronolide B synthase. Individual domains of the DEBS proteins bore homology to FAS and putative PKS domains. Of even greater interest was the domain architecture of the DEBS complex. The repetitive nature of the domains suggested a modular, assembly-line style of biosynthesis in which each module provided for a single acetate extension and subsequent β-ketone processing. Indeed, it is now known that modular PKSs are collinear with their eventual products, making product prediction a simple task of deciphering the polypeptide sequence.[2d] The modular nature

of these PKSs also account for the programing inherent in these enzymes. DEBS remains the model system for type I modular PKSs and has served as the basis of significant engineering and structural biology efforts.

1.3. *Fungal, Iterative PKSs*

The subset of fungal polyketides is incredibly diverse.[13] Fungal polyketides include some of the simplest natural products—*e.g.* orsellinic acid or 6-MSA—as well as some of the most complex metabolites—*e.g.* fumonisin B1 or lovastatin. Despite this apparent structural variety, iterative type I PKSs are responsible for the biosynthesis of nearly all fungal polyketides.[21] Whereas the programing inherent to the modular PKSs is explicit, the enzymatic control over polyketide programing is cryptic in the iterative PKSs. It is postulated that the catalytic program is encoded within the PKS itself; however, little is understood about the elements governing programing. Indeed, understanding catalytic programing in iterative PKSs remains one of the greatest unsolved challenges in natural products biosynthesis.

Innovative research conducted by Lazarus and Simpson helped to clarify the programming conundrum in fungal, iterative PKSs.[22] They categorized the fungal polyketides into two groups based upon their structures: nonreduced metabolites—largely encompassing the aromatic polyketides—and partially reduced metabolites—a group containing 6-MSA. Later a third category of highly reduced metabolites—an incredibly diverse class including metabolites such as fumonisin B1 and lovastatin—were included in the analysis.[23] Believing that subtle differences in protein sequence accounted for the structural differences in the eventual polyketide product, they compared the KS domains of PKS genes known at the time by preparing product class-specific degenerate KS
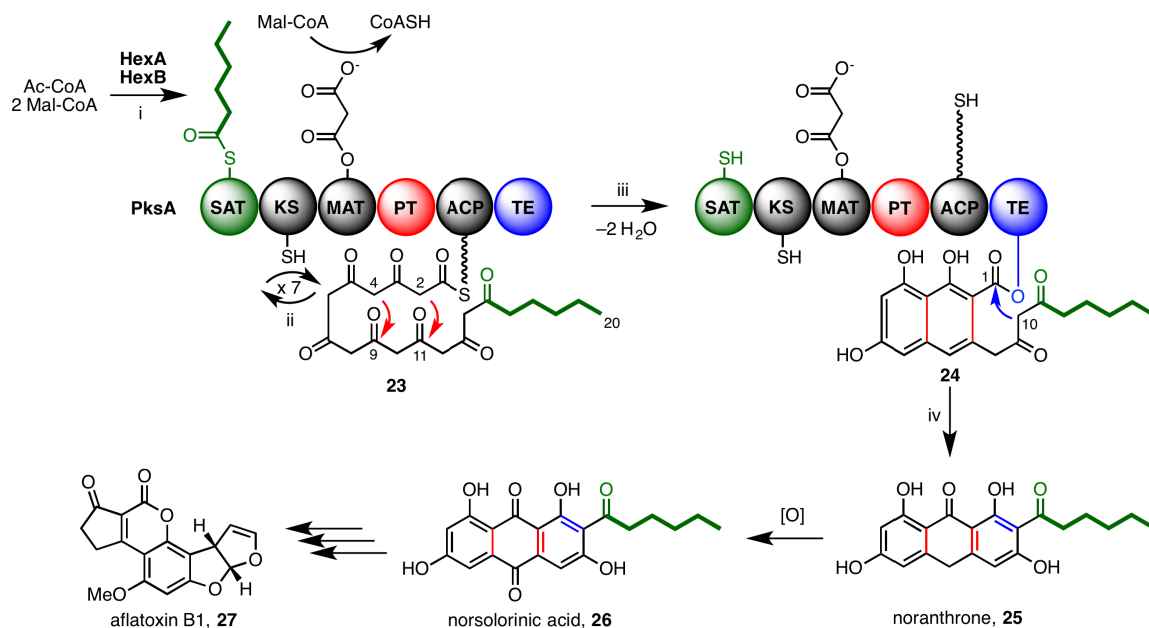
Figure 1.6│The norsolorinic acid synthase PksA is the model system for iterative type I NR-PKSs. Biosynthesis occurs in three stages: (i) initiation, (ii) elongation, (iii and iv) cyclization and release. The canonical NR-PKS domain architecture is indicated. Unique domains to NR-PKSs include the starter unit:ACP transacylase (SAT) and product template (PT).

primers. Using these class-specific degenerate primers, they amplified many fungal PKS genes from genomic DNA. After carefully analyzing the resulting sequences, they discovered that indeed each class of polyketide corresponded to a family of PKSs with specific and distinct domain architectures. These enzymes were classified as the nonreducing PKSs (NR-PKSs), the partially reducing PKSs (PR-PKSs), and the highly reducing PKSs (HR-PKSs), respectively.[13, 21] In recent years, widespread fungal genome sequencing projects have revealed multiple PKSs exist in each organism. Nevertheless, each of these genes fall within one of the three originally identified fungal, iterative PKS classes.

1.4. *Non-reducing PKSs*

The NR-PKSs have become the best-understood family of fungal, iterative PKSs. Much of this knowledge derives from pioneering work conducted within the Townsend

group to elucidate the enzymatic control of polyketide programing. This work forms the foundation of much of the research presented in this dissertation. NR-PKSs catalyze the synthesis of aromatic polyketide metabolites. Because they lack any of the reductive domains found in either FASs or other PKSs, no change of oxidation state occurs at the β-position during acetate extension. In this way, the NR-PKSs form the critical poly-β-ketone intermediate, which undergoes regiospecific cyclization/aromatization and release, producing aromatic metabolites (Figure 1.6). NR-PKSs share a common domain architecture that includes the canonical KS, AT—here more specifically named the malonyl-CoA:ACP transacylase (MAT) domain—and ACP domains. Unique to the NR-PKSs are the starter-unit:ACP transacylase (SAT)[24] and product template (PT)[25] domains. Additionally, many NR-PKSs contain a TE domain, although these domains often conduct more complicated transformations than simple hydrolysis.

The primary model system for studying NR-PKS biochemistry is PksA from the destructive agricultural pathogen *Aspergillus parasiticus* (Figure 1.6).[26] PksA is the central enzyme implicated in the biosynthesis of aflatoxin B1 (**27**)—a potent mycotoxin which undergoes epoxidation in the liver activating the DNA intercalator for crosslinking.[27] PksA catalyzes the formation of noranthrone (**25**) from a hexanoyl fatty acid precursor[28] produced by a pair of dedicated fungal FASs—HexA and HexB[29]— and seven malonyl-CoA extender units. Noranthrone is converted to norsolorinic acid (**26**)—the first isolable intermediate of the aflatoxin pathway—through the activity of an anthrone oxidase,[30] although this oxidation can occur spontaneously in aqueous solution. The biosynthetic pathway proceeds through a complicated series of oxidative rearrangements, eventually generating aflatoxin. Prolonged ingestion of aflatoxin-

contaminated food is linked to an increased risk for hepatocellular and renal carcinomas.[31]

1.4.1. Enzyme Deconstruction

Traditionally, limited proteolysis has been a powerful method for studying multidomain proteins.[32] By subjecting a multidomain protein to partial proteolysis, individual functional domains can be excised and assayed. Such approaches were invaluable in elucidating the biosynthesis of mammalian FASs.[33] The dissection-reconstitution approach became very attractive for those wishing to understand NR-PKS enzymology. Townsend and coworkers sought to deconstruct the model system PksA in order to understand its biochemistry. Despite a herculean effort, PksA remained an elusive target for heterologous expression. Indeed, the NR-PKSs have been notoriously difficult to proteins to express and purify. Without native soluble protein, limited proteolysis was impossible. A new approach had to be considered.

That new approach came in the form of bioinformatics and the Udwary-Merski algorithm (UMA).[34] UMA provided a reliable and general method for determining the domain boundaries and linker regions of multidomain proteins. The algorithm required two inputs: a multiple sequence alignment of the multidomain protein of interest and secondary structure prediction for each member of the multiple sequence alignment. From this input, an UMA score would be calculated for each position in the alignment that reflected the degree of primary sequence homology, the degree of predicted secondary sequence homology, and the hydrophobicity for a given position. It was demonstrated that linker domains corresponded to low UMA scores due to their inherent primary sequence variability, low secondary structure, and low hydrophobicity. UMA

analysis of PksA indicated the presence of six defined domains including two previously unappreciated domains: an N-terminal domain bearing limited predicted secondary sequence homology to the AT domains of FASs and PKSs and a central domain of unknown function and no predicted homology. These domains were eventually classified as the SAT and PT domains, respectively. Indeed, it was found that these domains were not exclusive to PksA, but highly conserved among the NR-PKSs contributing to the canonical domain architecture of the subclass.

UMA not only helped in determining the domain architecture of the NR-PKSs, but also aided artificial proteolysis of PksA for mechanistic determination *in vitro*. Using contemporary cloning strategies, Townsend and coworkers were able to heterologously express the PksA enzyme as mono- to multidomain fragments in *E. coli*. Their results were reported in a landmark paper where—remarkably—these protein fragments retained their wild type activity when recombined *in vitro*.[26] The dissection-reconstitution approach has been widely adapted for a variety of NR-PKSs and has formed the foundation for both biochemical and structural assays. At last, a reliable method for studying the *in vitro* biochemistry of the NR-PKSs was achieved and these enigmatic enzymes began to reveal their secrets.

1.4.2. *The Starter Unit Effect*

For over 60 years, a rich history of radioisotopic and later stable isotopic tracer assays established that—in most fungal polyketides—the fungal PKS selects a dedicated starter unit precursor to initiate biosynthesis.[7, 24, 35] The starter unit—usually acetyl-CoA—is separate and distinct from the Mal-CoA extender units that are used to generate the crucial poly-β-ketone intermediate. This observation has become known as the

"starter unit effect". In the majority of known fungal polyketides the starter unit is acetyl-CoA; however, other products utilize far more complicated precursors. In these cases, dedicated upstream biosynthetic machinery—including other PKSs or FASs as is the case for PksA[29b, 35-36]—generates the critical starter unit. It was not until Townsend and coworkers utilized the *in silico* directed deconstruction approach that the biochemical basis of the starter unit effect was understood.

In the initial *in silico* analysis of PksA by UMA,[24] sequence analysis of the extreme N-terminal domain of unknown function led to the hypothesis that it might be responsible for selecting the HexA/HexB generated hexanoyl fatty acid starter unit of norsolorinic acid biosynthesis. In fact, the sequence analysis showed that not only was this domain conserved among the NR-PKSs, but it also contained a conserved GXCXG motif, presumptively for the tethering of the precursor metabolite as an acyl thioester. This observation was in contrast to the MAT domains, which contain a GHSXG motif where the Ser serves as the active site acyl ester tether for the malonyl extender units. On the basis of this analysis and subsequent biochemical assays, the mysterious domain was given the name starter unit:ACP transacylase (SAT).

Using protein deconstruction of PksA, the enigmatic SAT domain and the ACP domain were heterologously expressed as monodomain fragments and used in a radioisotope labeling transacylase assay to measure if the PksA SAT domain could catalyze the transfer of [1-^{14}C]hexanoyl-CoA to the ACP domain.[24] The results of this assay showed that the SAT domain could not only catalyze transfer of the hexanoyl moiety to the ACP, but that this transfer proceeded through a hexanoyl-SAT acyl intermediate. The PksA SAT was also shown to be reasonably selective for hexanoyl-

CoA.[24, 37] It is now known that in NR-PKSs utilizing complicated starter units, the SAT

domain is selective for these unusual precursors serving as the link between the starter

unit generator and the NR-PKSs (Figure 1.6, reaction i). In fact, this crucial link is

believed to account for much of the incredible structural diversity of fungal polyketides.

In the absence of a dedicated starter unit biosynthetic module, the SAT domain has been

shown to be selective for acetyl-CoA thereby accounting for the prevalence of the acetyl-

derived starter units in NR-PKS biosynthesis. It has also been demonstrated that the PksA

SAT domain can transfer non-native precursors so long as they are reasonably similar to

the native hexanoyl starter unit.[37]

Recently, the crystal structure of the SAT from CazM was solved (Figure 1.7).[38]

CazM is a NR-PKS from *Chaetomium globosum* responsible for the biosynthesis of

chaetoviridin A and chaetomugilin A.[39] Like many fungal NR-PKSs, CazM works in

concert with a dedicated HR-PKS—CazF—that generates a highly reduced polyketide
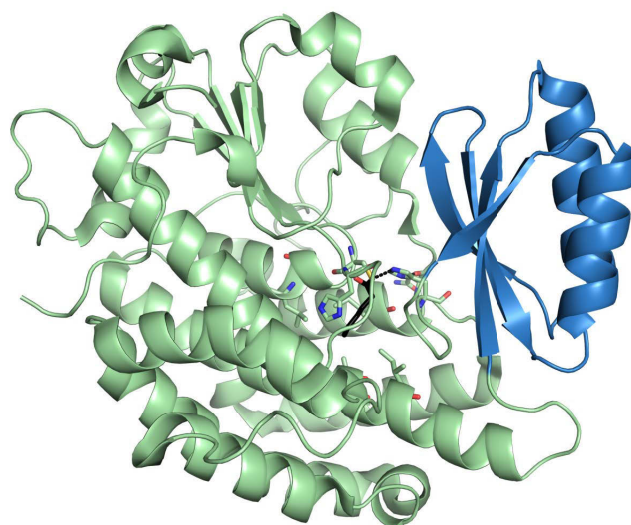
Figure 1.7│Crystal structure of the CazM SAT domain.[38] The α/β hydrolase fold is displayed in green. The ferrodoxin-like subdomain is shown in blue. Hexanoyl bound to the active site Cys is displayed in black. Residues constituting the active site are shown as sticks.

starter unit. The modular separation of precursor formation from aromatic moiety construction is a common biosynthetic strategy among NR-PKSs observed for many classes including the azaphilone synthases[39-40]—the family to which CazM belongs—and the resorcylic acid lactone synthases.[41] In all of these cases, the SAT domain serves as the bridge connecting the two separate modes of biosynthesis. The CazM SAT domain structure comprises an α/β hydrolase core and a small ferrodoxin-like subdomain with the active site situated at the interface of these two subdomains. The active site is hydrophobic presumably to stabilize the highly reduced polyketide starter unit. Additionally, small conformational changes to the SAT active site are observed upon binding of a hexanoyl thioester—a mimic of the native triketide CazM starter unit. The conformational change positions residues such that they can stabilize the oxyanion tetrahedral intermediate generated during the acyl transfer.

1.4.3. Iterative Claisen Condensation and the Poly-β-Ketone

As previously stated, the central chemistry of any PKS is repeated decarboxylative Claisen condensation of acetate units. In the case of the bacterial type II PKSs, this activity is catalyzed by the so-called minimal PKS.[2c, 14b] In these systems, the minimal PKS is defined as a heterodimeric KS complex (KSα and KSβ) and the ACP—the only proteins necessary to catalyze poly-β-ketone formation. A similar term for the minimal NR-PKSs is defined as the requisite domains for efficient acetate homologation. From a biological point of view, the minimal NR-PKS is an absurd classification, as NR-PKSs by their very nature comprise all domains of the canonical domain architecture. The term is exclusively used to describe individual steps of NR-PKS biosynthesis and is

useful when discussing deconstructed NR-PKSs, where canonical domains can be excised.

The minimal NR-PKS comprises the three N-terminal domains (SAT-KS-MAT) and the ACP.[14a, 26] The SAT—as described—selects the starter unit and initiates biosynthesis whereas the MAT and KS collaborate to introduce malonyl extender units from malonyl-CoA and catalyze decarboxylative Claisen condensation generating the ACP-bound poly-β-ketone, respectively (Figure 1.6, reaction ii). Although this division of biosynthesis was long hypothesized, it was first biochemically demonstrated in a pioneering effort by the Townsend group in collaboration with the Kelleher group using the PksA model system.[26, 42] For the first time, the fully mature ACP-bound linear poly-β-ketone was observed by MS/MS fragmentation analysis. Astonishingly, extension in all cases was complete and no intermediate poly-β-ketones could be detected. Furthermore, chain length was completely controlled—a feature previously observed in reactions of the KS, MAT, and ACP domains of Pks4, the NR-PKS from *Gibberella fujikuroi* implicated in bikaverin biosynthesis.[14a, 43] It is believed that the KS domain exerts stringent chain length control in the NR-PKSs as in the case of the type II PKSs. In those systems, the active site volume of the KS heterodimer limits chain length to the prescribed number of extensions.[44] Furthermore, in reactions of PksA in which non-native starter units were used, the number of malonyl extensions accounted for the difference favoring linear intermediates of the correct C_{20} chain length.[14a, 37]

In the absence of downstream cyclizing domains, the highly reactive linear poly-β-ketone of the minimal NR-PKS undergoes spontaneous cyclization and release generating a thermodynamic distribution of so-called derailment products.[14a, 26] In

general, derailment products derive from simple intramolecular aldol and pyrone cyclizations, dehydrations, aromatizations, and—in some cases—decarboxylations. Derailment of this nature is greatly reduced when downstream domains of the NR-PKS are introduced and when the level of deconstruction is reduced—meaning the enzyme is cut in fewer places.[26] Both observations point towards a significant role for substrate channeling in the native enzymes as well as an inherent protective functionality in which the enzyme stabilizes the linear poly-β-ketone reactive intermediate. The nature of this stabilization is unknown, but likely involves domain-mediated restriction of conformational sampling by the reactive intermediate. In the bacterial type II PKSs and FASs, the ACP plays a protective role by sequestering bound intermediates within the hydrophobic core of the ACP in a mechanism similar to that of a switchblade.[45] The switchblade mechanism is unlikely in the type I PKSs—including the NR-PKSs. In fact, a solution structure of the PksA ACP showed no evidence for intermediate sequestration.[46]
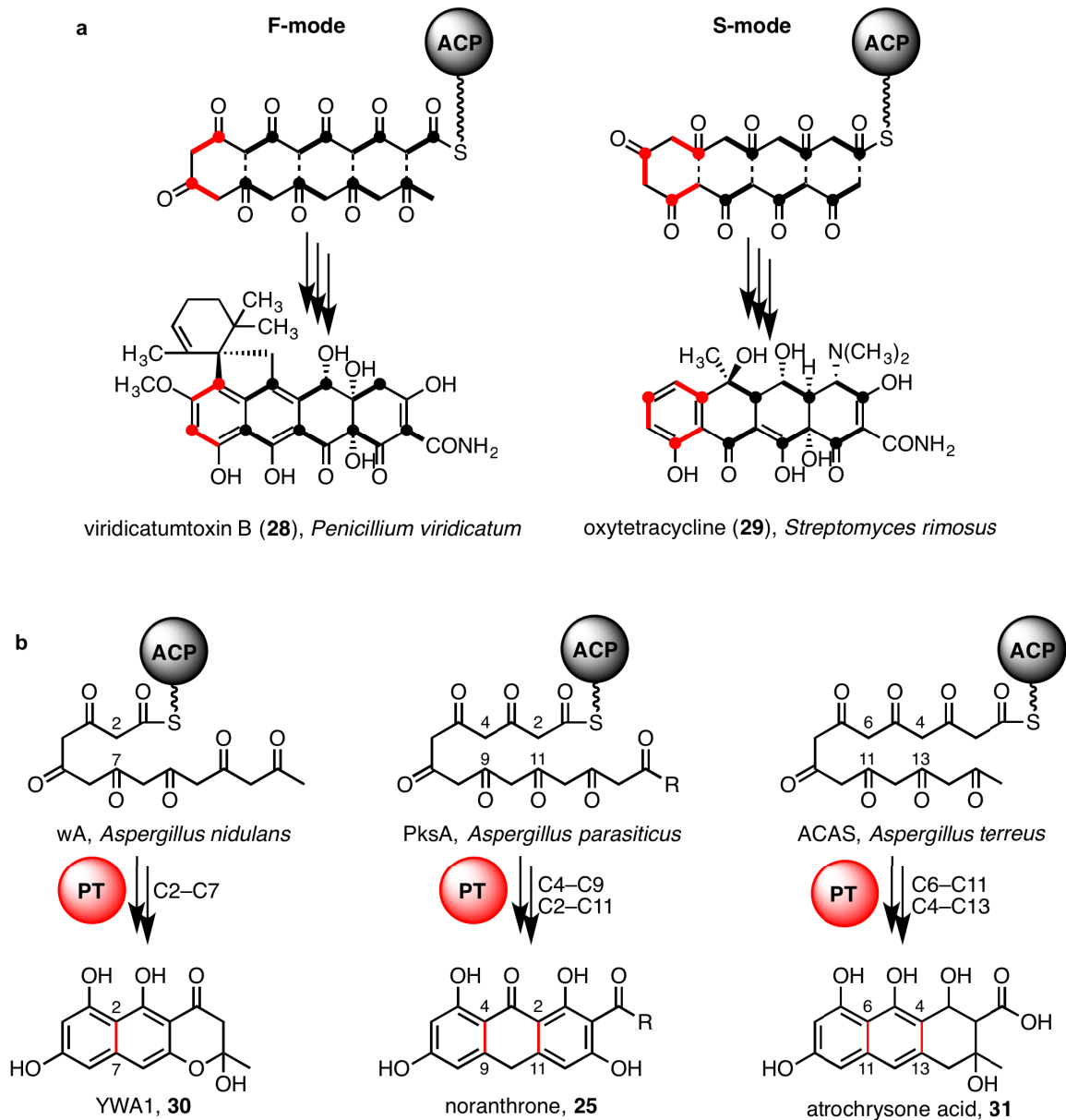
Figure 1.8│The cyclization modes of aromatic polyketides. (a) The F-mode and S-mode of cyclization are characteristic of fungal and bacterial polyketides, respectively. The origins of the carbons in the similar products viridicatumtoxin B (**28**) and oxytetracycline (**29**) are indicated to illustrate the two folding modes. (b) The PT domains of NR-PKS adopt three distinct cyclization regiospecificities: C2–C7 (as in YWA1, **30**), C4–C9 (as in noranthrone, **25**), and C6–C11 (as in atrochrysone acid, **31**).

1.4.4. Regiospecific Cyclization by the PT Domain

Both bacterial type II PKSs and fungal type I NR-PKSs catalyze the formation of

aromatic products through a linear poly-β-ketone intermediate. In either case, the PKS

20

system controls the regiospecificity of the cyclization/aromatization.[47] Interestingly, the bacterial and fungal aromatic polyketides adopt distinctive cyclization folding registers— the S-mode and F-mode, respectively—as determined by stable isotope labeling patterns (Figure 1.8a).[47a, 48] The origins of the S-mode and F-mode regiospecificities lie in the distinct mechanisms of aromatization catalyzed by the divergent PKS systems. In the bacterial type II PKSs, aldol cyclization and aromatization is catalyzed by dedicated aromatases (AROs).[49] The AROs bind the ACP-bound linear poly-β-ketone intermediate in a pre-folded conformation that brings the regiospecific electrophile and nucleophile pair in proximity to one another, favoring the S-mode. Cyclization in the fungal type I NR-PKSs occurs in the product template (PT) domain by a distinct mechanism.[25-26] The PT domain was the final canonical NR-PKS domain to be recognized and characterized. It serves the crucial role of setting the aromatic scaffold of the final polyketide product.

The role of the PT domain in cyclization was first demonstrated in the model system PksA (Figure 1.6, reaction iii).[25-26, 42] When a deconstructed PT domain was added to reactions of the minimal PksA (SAT-KS-MAT + ACP), biosynthesis was redirected away from spontaneous derailment products and towards norpyrone—a naphthopyrone. Norpyrone adopted the predicted F-mode of folding with apparent aldol cyclization/aromatization occurring between C4–C9 and C2–C11 followed by O14–C1 bond closure—a low barrier spontaneous transformation. In fact, ions corresponding to the ACP-bound fully elongated C_{20} poly-β-ketone with either one or two dehydrations were observed by MS/MS fragmentation in reconstitution reactions containing the PT domain.[26, 42] The presence of both the singly and doubly dehydrated intermediates suggested sequential aldol cyclization catalyzed by the PT domain. Structural analysis of

NR-PKS products indicated that three so-called cyclization registers exist for fungal aromatic polyketides: C2–C7, C4–C9, and C6–C11 (Figure 1.8b).[2b, 50] Phylogenetic analysis of a curated list of NR-PKSs for which products were known similarly showed that the PT domains separate into five distinct clades, each with a characteristic cyclization register.[50] Furthermore, the active site sequences of known members of each cyclization register are distinct, suggesting a structural basis for cyclization regiospecificity.[25]
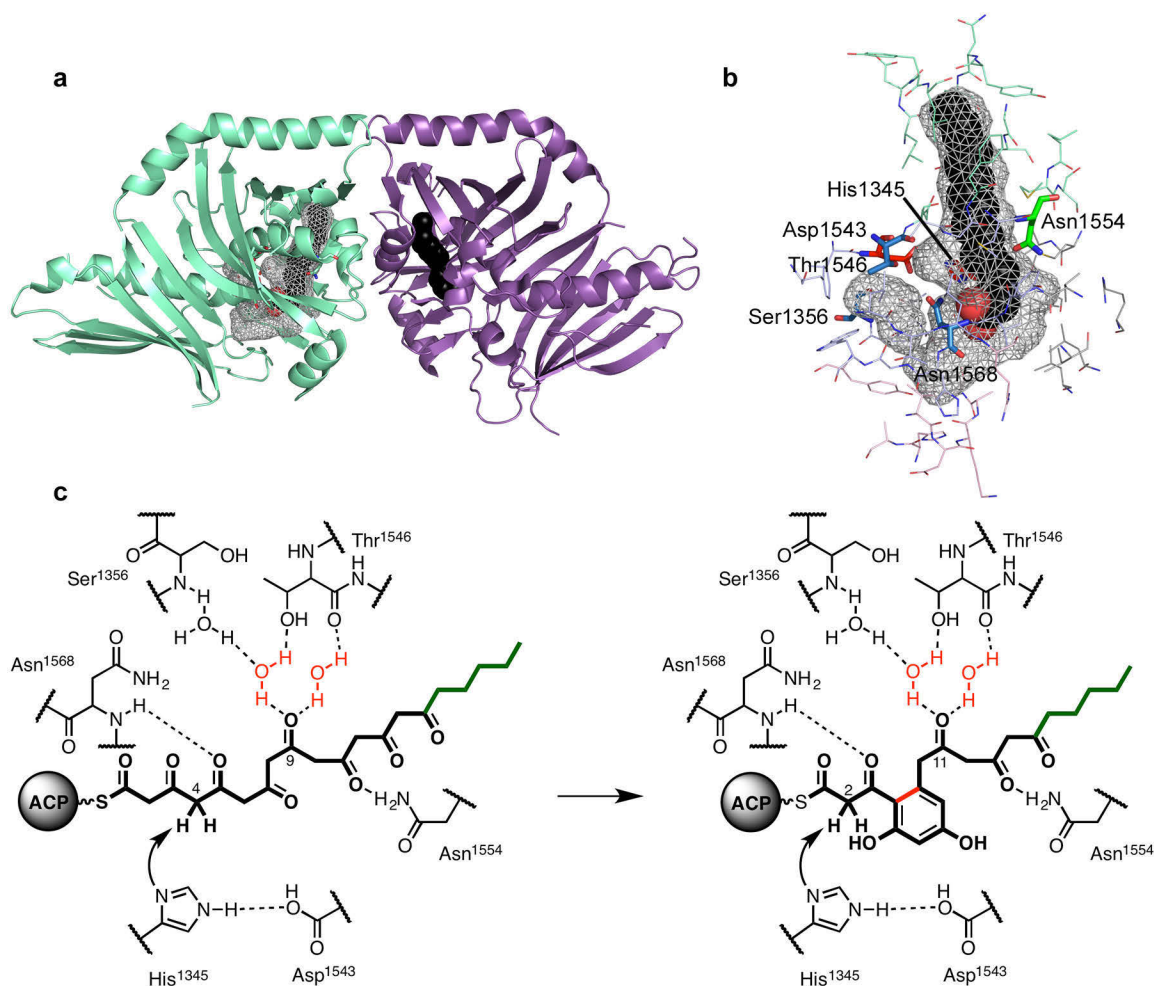
Figure 1.9│The PksA PT domain structure serves as the basis for mechanistic understanding of regiospecific cyclizations/aromatizations in NR-PKSs. (a) The PksA PT crystal structure. The PT domain is a homodimer with each monomeric unit adopting the double hot dog fold. The monomers are displayed in green and purple, respectively. The active site cavity is displayed for one monomer with palmitate bound. Palmitate is displayed as black spheres. (b) The active site of the PksA PT contains three distinct regions: phosphopantetheine binding region (pink residues), cyclization chamber (gray and blue residues for the dry and wet sides, respectively), and hexanoyl binding region (green residues). The residues important for regiospecific aldol cyclization/aromatization are presented as sticks. (c) The mechanism for regiospecific aldol cyclization/aromatization requires activation of the electrophilic ketone through dual hydrogen bonding.

The current understanding of the PT cyclization mechanism is largely based upon a pair of crystal structures of the PksA PT domain with either palmitate (a mimic of the linear poly-β-ketone) or HC8 (a bicyclic mimic of the cyclized PT product) bound in the active site (Figure 1.9).[25] The PksA PT monodomain exists as a homodimer—both in

the crystal structure and in solution—and adopts a modified double hot dog fold.[51] The double hot dog fold is a common structural fold observed in the DH domains of mammalian FAS[11], modular PKSs,[52] and bacterial type II FAS.[53] Despite having little sequence homology, the PksA PT structure is similar to that of the mammalian FAS DH domain further underlining the similarities between the NR-PKSs and mammalian FASs.

The PksA PT active site is defined by the double hot dog fold and comprises three distinct regions: the phosphopantetheine binding region, the cyclization chamber, and the hexanoyl-binding region (Figure 1.9b).[25] All told, the three regions of the active site define a deep and narrow pocket reaching about 30 Å from the surface into the hydrophobic core of the protein. Electron densities for either palmitate or HC8 bound in the active site are well defined. There is virtually no conformational change in the protein structure for either the palmitate or HC8 structures. Furthermore, the palmitate binds in its fully extended conformation, indicating that—unlike the bacterial type II AROs—the linear poly-β-ketone intermediate binds in its fully extended conformation. The tail of the palmitate binds the narrow and hydrophobic hexanoyl-binding region, stabilizing palmitate's extended conformation and suggesting a structural rationale for hexanoyl starter unit accommodation. Interestingly, the cyclization chamber contains clear orthogonal "wet" and "dry" sides. A network of crystallographic water molecules organized through residue-directed hydrogen bonding defines the wet side while hydrophobic residues compose the dry side. The HC8 compound occupies the dry side of the cyclization chamber. Situated at the center of the cyclization chamber is the catalytic His1345, activated through hydrogen bonding by its catalytic dyad partner, Asp1543.

The mechanistic origin of regiospecificity is based upon *in silico* docking studies of the linear poly-β-ketone and monocyclic intermediates (Figure 1.9c).[25] It is believed that the linear substrate binds in an elongated conformation with a slight kink in the cyclization chamber, allowing for regiospecific C4–C9 and C2–C11 cyclization. This slightly kinked structure is stabilized through dual hydrogen bonding of the C9 ketone to two crystallographic waters, activating it as an electrophile.[54] The positioning of these waters is determined by the hydrogen bonding network of the wet side of the cyclization chamber. Furthermore, the polarity of the cyclization chamber likely contributes to the extent of enol-keto tautomerization in the poly-β-ketone. Solvent polarity has a determining effect on the extent of enolization of 1,3-diketones, with a more polar solvent favoring the diketo tautomer.[55] By projecting the C9 carbonyl into the wet side of the active site, the enzyme stabilizes the electrophilic keto tautomer. The catalytic His1345—with participation of Asp1543—removes a methylene proton at the C4 position, generating an enzyme-stabilized enolate intermediate. Collapse of the enolate together with C9 electrophilic activation directs regiospecific aldol cyclization and aromatization. Following first-ring cyclization, translocation of the distal end of the monocyclic intermediate occurs, allowing for the second C2–C11 aldol cyclization to occur in an analogous mechanism.

1.4.5. Product Release by the TE Domain
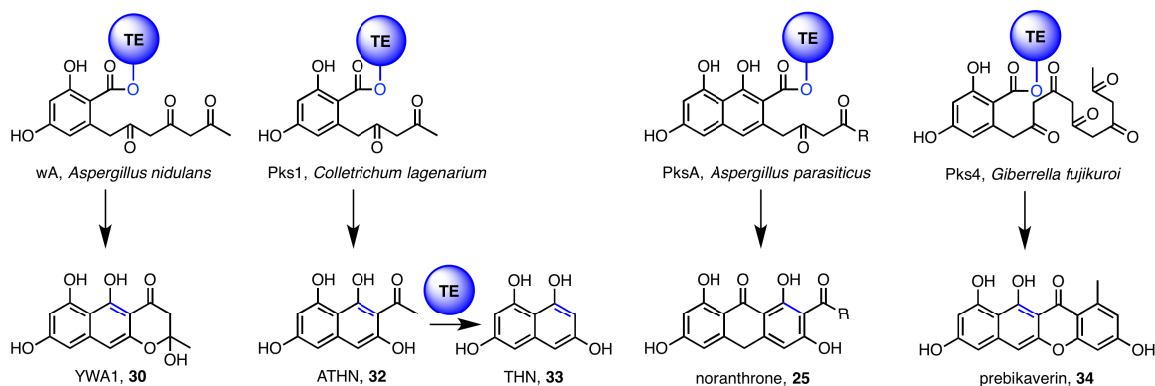
Figure 1.10 | Mode of release by NR-PKS TE domains. NR-PKS TE domains often catalyze regiospecific Claisen cyclizations. They can catalyze additional reactions including deacetylation as is the case for Pks1, which transforms ATHN (**32**) to THN (**33**).

PKSs utilize a variety of methods to catalyze product release.[56] In canonical NR-PKSs, the thioesterase (TE) domain catalyzes this activity. TE biochemistry in FASs and modular type I PKSs is varied and includes macrolactonization, macrolactamization, and hydrolysis. The TEs of NR-PKSs expand upon these functions and can also catalyze Claisen cyclization/Dieckmann C–C bond formation (Figure 1.6, reaction iv; Figure 1.10).[26, 57] Expression of the wA gene from *Aspergillus nidulans* in the heterologous host *Aspergillus oryzae*, Ebizuka and coworkers established the first NR-PKS harboring a Claisen cyclase type TE (TE-CLC).[57a] The wA protein produces the spore pigment YWA1 (**30**), the product of the TE-CLC.[27a, 58] In earlier studies where the TE domain was mistakenly truncated, wA generated citreoisocoumarins—the products of spontaneous pyrone formation by C–O bond closure, a lower energy transformation.[58a] Since these experiments, TE-CLCs have been identified in many NR-PKSs including PksA[26, 57b] and Pks4.[43]

The crystal structure of the PksA TE monodomain[57b] reveals that it adopts the characteristic α/β-hydrolase fold of acyltransferases and hydrolases (Figure 1.11a).[59] The α/β-hydrolase fold is well characterized and found in the structures of modular type I

PKSs and NRPSs, where they usually catalyze macrolactonization and macrolactamization, respectively. These TEs rely on a classic catalytic triad (Asp, His, Ser) which functions in a fashion analogous to the serine protease catalytic triad.[60] Unique to the α/β-hydrolase fold is a distinctive lid subdomain, which forms a portion of the active site and is key to substrate binding.[57b] Delivery of the penultimate intermediate to the TE active site occurs upon ACP docking, inducing a lid opening conformational change. A similar lid opening interaction is observed for the TE-peptidyl carrier protein (PCP)—an analog of the FAS and PKS ACP—pair of enterobactin biosynthesis.[61] The ACP-bound acyl intermediate is transferred to the TE active site Ser, where it is tethered as an enzyme acyl ester (Figure 1.11b). As the ACP disassociates from the acylated TE, the lid is proposed to close upon the active site sequestering the substrate within a hydrophobic binding pocket. The shape of this pocket imposes a strict conformational bias on the substrate, bringing the side chain within deprotonation distance to the catalytic His thereby promoting regiospecific C–C bond, Dieckmann, closure.
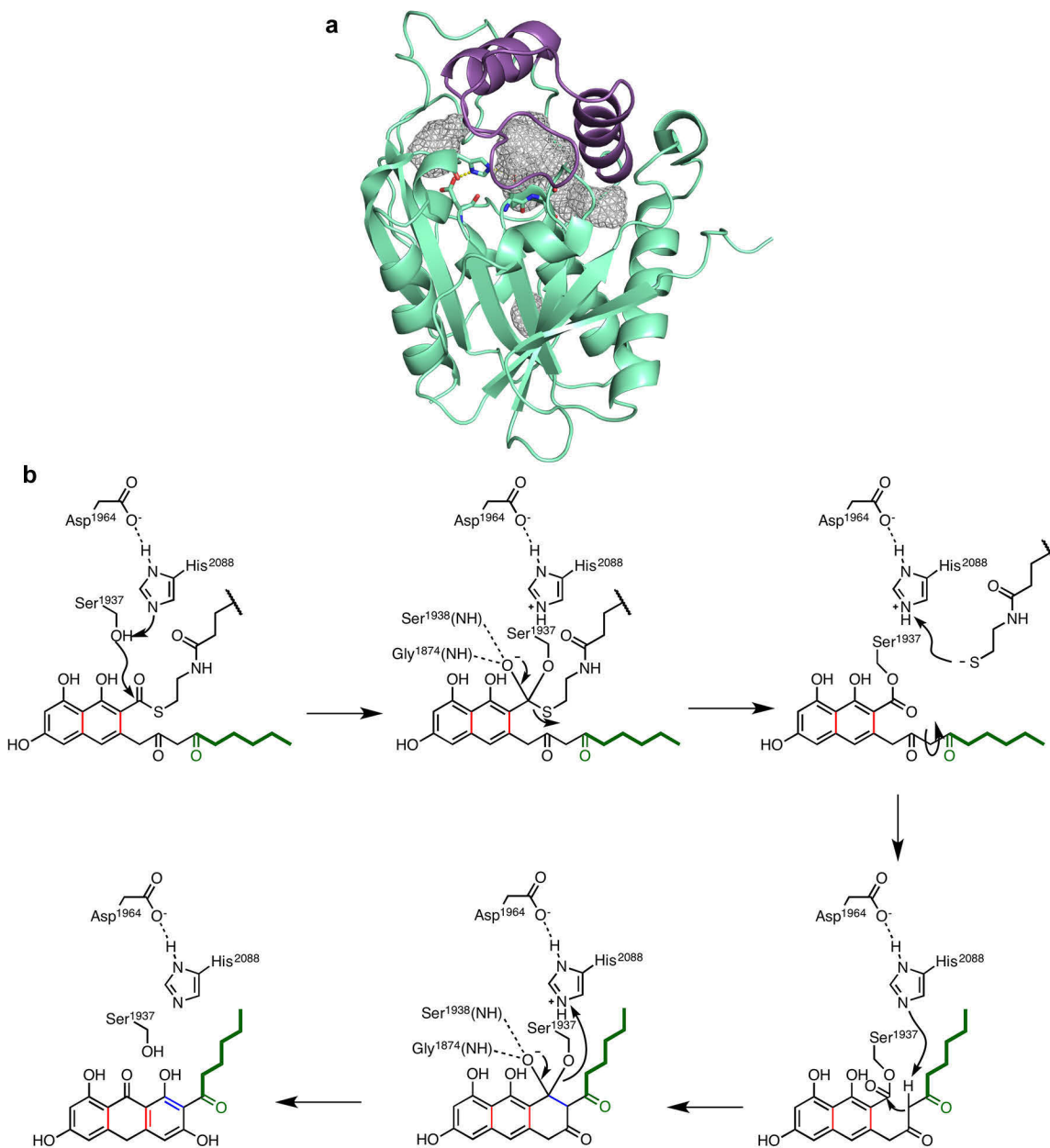
Figure 1.11 │ The PksA TE structure and mechanism. (a) The PksA TE adopts an α/β-hydrolase fold (green) with a flexible lid region (purple). The lid helps define the active site cavity (cage) into which the final polyketide intermediate is delivered. The classic Ser-His-Asp catalytic triad is shown as sticks. (b) The PksA TE mechanism first requires transfer of the acyl intermediate from the ACP to the TE. As the phosphopantetheine of the ACP leaves the active site, it is protonated and the hexanoyl moiety flips into the site that it vacates. Flipping of the hexanoyl group brings the methylene position into proximity of the catalytic His, which facilitates Claisen cyclization. The oxyanion tetrahedral intermediate is stabilized through backbone interactions.

28

As stated, the lid region is critical in forming the hydrophobic binding pocket and therefore catalytic specificity. The NR-PKS Pks1 from *Colletotrichum lagenarium* is implicated in melanin biosynthesis.[62] Like wA and PksA, the Pks1 TE domain catalyzes Claisen cyclization to generate the hexaketide product, 2-acetyl-1,3,6,8-tetrahydroxynaphthalene (ATHN, **32**, Figure 1.10). In addition to the CLC activity, the Pks1 TE domain also harbors a deacetylase activity and converts ATHN to 1,3,6,8-tetrahydroxynaphthalene (THN, **33**). Like Pks1, the TE-CLC domain of WdPks1 from *Wangiella dermatitidis* catalyzes the formation of ATHN from the hexaketide product.[63] However—unlike Pks1—WdPks1 does not contain an additional deacetylase activity. When the lid subdomains of Pks1 and WdPks1 were swapped, the product distribution between ATHN and THN was predictably altered with WdPks1 harboring the Pks1 lid capable of performing the deacetylation.[62]

In addition to its biosynthetic role, the TE domain of NR-PKSs also performs a critical editing function.[42] TE-mediated editing occurs when a nonproductive metabolite accumulates on the ACP. Because the ACP thiol tether serves as the locus of all biochemistry in the NR-PKSs, an improperly loaded ACP renders the entire enzyme inactive. In these cases, the TE domain can intercede and hydrolyze the improper acyl-ACP species, resetting the enzyme. This effect was first observed by MS/MS fragmentation of PksA domain bound intermediates.[42] In reconstitution reactions where the TE domain was withheld, both acetyl and hexanoyl intermediates would accumulate on the ACP. Neither of these intermediates was observed when the TE domain was included. The TE domain was shown to directly hydrolyze acetyl-ACP and hexanoyl-ACP species. Furthermore, derailment and truncation products of a minimal PksA were

reduced during titration of increasing amounts of PksA TE. Together these data suggest the critical editing role of the TE domain. The interplay between the biosynthetic and editing roles of the TE domain have been described as decision gating, and are predicted to be of crucial importance for NR-PKS engineering.[64]

1.5. *Outlook*
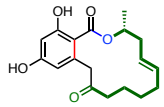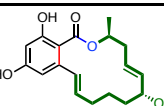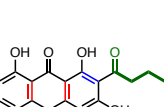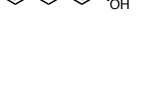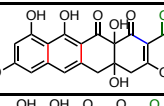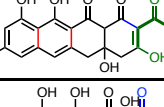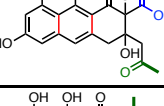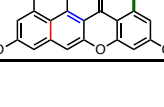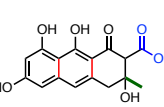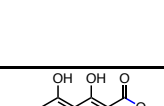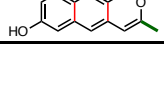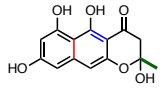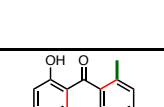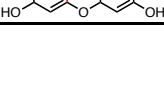
Our understanding of polyketide biosynthesis has rapidly expanded in the last decade due in no small part to experimental advances in protein deconstruction. The catalytic program of the NR-PKSs—once considered an intractable black box—is now significantly better understood on the biosynthetic, structural, and mechanistic levels. Four chemical features govern the diversity of fungal aromatic polyketides: starter unit, chain length, cyclization register, and product release. As described above, each of these features is controlled by a separate domain of the NR-PKSs—SAT, KS, PT, and TE, respectively—allowing us to reliably predict the products of canonical NR-PKSs simply by investigating their primary sequences. Indeed, the individual activity of NR-PKSs linked to known products invariably follow the inherent, or programed, biosynthetic logic of their SAT, KS, PT, and TE domains (Table 1.1).
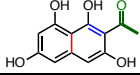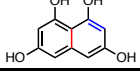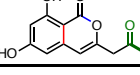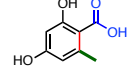
Despite these dramatic leaps, much is still left to learn. Although the foundations of catalytic programing in the NR-PKSs is well understood, it is unclear how they can be used to aid in NR-PKS engineering. Addressing this challenge is a central feature of this dissertation. While the canonical NR-PKSs are now well understood, catalytic programing in common NR-PKS variants is still unclear. Many NR-PKSs contain *C*-methylation domains which function during elongation to methylate the α-position at programed stages of extension. The structural basis of the mechanism governing this

level of programing is completely unknown. Furthermore, little is known of the global structure of these enzymes or how interdomain linker regions influence catalysis. In addition to global structure, the nature of domain-domain interactions in NR-PKSs is a central question. With so many unanswered questions, it is clear that NR-PKS biosynthesis will remain an active area of research in the decades to come.

Table 1.1 │ Known products of canonical NR-PKSs.

Species	PKS	Accession	NR-PKS Product	Final	SAT	KS	PT	TE	Ref.
Pochonia chlamydosporia	Rdc1	ACD39770		Radicinol	C10 Rdc5	4	C2–C7	Macro O10′–C1	[65]
Chaetoium chiversii	RadS2	ACM42403		Radicinol	C10 RadS1	4	C2–C7	Macro O10′–C1	[66]
Gibberella zeae	GzPks13	ABB90282		Zearalenone	C12 Pks4	3	C2–C7	Macro O10′–C1	[67]
Hypomyces subiculosus	Hpm3	ACD39762		Hypothemycin	C12 Hpm8	3	C2–C7	Macro O10′–C1	[65a]
Aspergillus parasiticus	PksA	AAS66004		Aflatoxin	C6 HexA/B	7	C4–C9 C2–C11	CLC C14–C1	[27a]
Aspergillus oryzea	PksA	BAE71314		Aflatoxin	C6	7	C4–C9 C2–C11	CLC C14–C1	[68]
Aspergillus ochraceoroseus	Aflc	ACH72912		Aflatoxin	C6	7	C4–C9 C2–C11	CLC C14–C1	[69]
Aspergillus flavus	PksA	AAS90093		Aflatoxin	C6	7	C4–C9 C2–C11	CLC C14–C1	[70]
Aspergillus nidulans	pksST	AAA81586		Sterigmatocystin	C6	7	C4–C9 C2–C11	CLC C14–C1	[71]
Dothistroma septosporum	PksA	AAZ95017		Dothistromin	C6	7	C4–C9 C2–C11	CLC C14–C1	[72]
Aspergillus niger	AdaA	AEN83889		BMS-192548	C2	9	C6–C11 C4–C13	CLC C18–C1 AdaB	[73]
Penicillium aethiopicum	VrtA	ADI24926		Viridicatumtoxin	Mal-ona-moyl	8	C6–C11 C4–C13	CLC C18–C1 VrtG	[74]
Aspergillus nidulans	AptA	XP_663604		Asperthecin	C2	8	C6–C11 C4–C13	Hyd AptB	[75]
Gibberella fujikuroi	Pks4	CAB92399		Bikaverin	C2	8	C2–C7	CLC C10–C1	[43]
Aspergillus terreus	ACAS	XP_001217072		Emodin Endocrocin	C2	7	C6–C11 C4–C13	Hyd ACTE	[76]
Aspergillus nidulans	mpdG	XP_657754		Prenyl Xyanthones	C2	7	C6–C11 C4–C13	Hyd MpdF	[77]
Aspergillus fumigatus	EncA	XP_746435		Endocrocin	C2	7	C6–C11 C4–C13	Hyd EncB	[78]
Aspergillus fumigatus	TpcC	XP_751377		Tripacidin	C2	7	C6–C11 C4–C13	Hyd TpcB	[79]
Cercospora nicotianae	CTB1	AAT69682		Cercosporin	C2	6	C4–C9 C2–C11	Pyrone O13–C1	[80]
Aspergillus nidulans	wA	Q03149		Conidal Pigment	C2	6	C2–C7	CLC C10–C1	[57a]
Aspergillus fumigatus	Alb1	XP_756095		Melanin	C2	6	C2–C7	CLC C10–C1	[81]
Aspergillus niger	FwnA	CAL00851		Melanin	C2	6	C2–C7	CLC C10–C1	[82]
Gibberella zeae	Aur1	AAU10633		Aurofusarin	C2	6	C2–C7	CLC C10–C1	[83]
Penicillium aethiopicum	GsfA	ADI24953		Griseofulvin	C2	6	S-type C6–C1 C8–C13	n/a	[74]

Exophiala dermatitidis	WdPks1	EHY55015		Melanin	C2	5	C2–C7	CLC C10–C1	[63]
Colletotrichum lagenarium	Pks1	BAA18956		Melanin	C2	5	C2–C7	CLC C10–C1	[62]
Aspergillus oryzae	AoiG	XP_001827098		Orthosporin Diaporthin	C2	5	C2–C7	?	[84]
Aspergillus nidulans	OrsA	XP_681178		Lecanoric acid Violaceol F-9775A, F-9775B	C2	3	C2–C7	Hyd	[85]

1.6. References

[1]   D. J. Newman, G. M. Cragg, *J Nat Prod* **2012**, *75*, 311-335.
[2]   (a) J. Staunton, K. J. Weissman, *Nat Prod Rep* **2001**, *18*, 380-416; (b) J. M. Crawford, C. A. Townsend, *Nature Rev Microbiol* **2010**, *8*, 879-889; (c) C. Hertweck, A. Luzhetskyy, Y. Rebets, A. Bechthold, *Nat Prod Rep* **2007**, *24*, 162-190; (d) C. Hertweck, *Trends Biochem Sci* **2015**, *40*, 189-199.
[3]   (a) S. Smith, S. C. Tsai, *Nat Prod Rep* **2007**, *24*, 1041-1072; (b) J. E. Cronan, J. Thomas, *Methods Enzymol* **2009**, *459*, 395-433.
[4]   A. Birch, R. Massy-Westropp, C. Moye, *Aust J Chem* **1955**, *8*, 539-544.
[5]   N. Collie, W. S. Myers, *Journal of the Chemical Society, Transactions* **1893**, *63*, 122-128.
[6]   O. Puel, P. Galtier, I. P. Oswald, *Toxins (Basel)* **2010**, *2*, 613-631.
[7]   T. J. Simpson, *Chem Soc Rev* **1987**, *16*, 123-160.
[8]   (a) I. Fujii, *Nat Prod Rep* **2009**, *26*, 155-169; (b) K. J. Weissman, *Methods Enzymol* **2009**, *459*, 3-16; (c) K. J. Weissman, *Nat Chem Biol* **2015**, *11*, 660-670.
[9]   (a) M. Leibundgut, T. Maier, S. Jenni, N. Ban, *Curr Opin Struct Biol* **2008**, *18*, 714-725; (b) T. Maier, M. Leibundgut, D. Boehringer, N. Ban, *Q Rev Biophys* **2010**, *43*, 373-422.
[10]   S. Jenni, M. Leibundgut, D. Boehringer, C. Frick, B. Mikolasek, N. Ban, *Science* **2007**, *316*, 254-261.
[11]   T. Maier, M. Leibundgut, N. Ban, *Science* **2008**, *321*, 1315-1322.
[12]   C. O. Rock, J. E. Cronan, *Biochim Biophys Acta* **1996**, *1302*, 1-16.
[13]   R. J. Cox, T. J. Simpson, in *Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates, Vol. 459* (Ed.: D. A. Hopwood), Elsevier Academic Press Inc, San Diego, CA, **2009**, pp. 49-78.
[14]   (a) W. J. Zhang, Y. R. Li, Y. Tang, *Proc Natl Acad Sci USA* **2008**, *105*, 20683-20688; (b) A. Das, C. Khosla, *Acc Chem Res* **2009**, *42*, 631-639.
[15]   (a) M. B. Austin, J. P. Noel, *Nat Prod Rep* **2003**, *20*, 79-110; (b) M. B. Austin, M. Izumikawa, M. E. Bowman, D. W. Udwary, J. L. Ferrer, B. S. Moore, J. P. Noel, *J Biol Chem* **2004**, *279*, 45162-45174; (c) M. Hashimoto, T. Nonaka, I. Fujii, *Nat Prod Rep* **2014**, *31*, 1306-1317.
[16]   F. Malpartida, D. A. Hopwood, *Nature* **1984**, *309*, 462-464.

[17] R. McDaniel, S. Ebertkhosla, H. Fu, D. A. Hopwood, C. Khosla, *Proc Natl Acad Sci USA* **1994**, *91*, 11542-11546.

[18] (a) R. McDaniel, C. R. Hutchinson, C. Khosla, *J Am Chem Soc* **1995**, *117*, 6805-6810; (b) P. J. Kramer, R. J. X. Zawada, R. McDaniel, C. R. Hutchinson, D. A. Hopwood, C. Khosla, *J Am Chem Soc* **1997**, *119*, 635-639.

[19] (a) J. Cortes, S. F. Haydock, G. A. Roberts, D. J. Bevitt, P. F. Leadlay, *Nature* **1990**, *348*, 176-178; (b) D. J. Bevitt, J. Cortes, S. F. Haydock, P. F. Leadlay, *Eur J Biochem* **1992**, *204*, 39-49.

[20] (a) J. S. Tuan, J. M. Weber, M. J. Staver, J. O. Leung, S. Donadio, L. Katz, *Gene* **1990**, *90*, 21-29; (b) S. Donadio, M. J. Staver, J. B. McAlpine, S. J. Swanson, L. Katz, *Science* **1991**, *252*, 675-679.

[21] R. J. Cox, *Org Biomol Chem* **2007**, *5*, 2010-2026.

[22] L. E. Bingle, T. J. Simpson, C. M. Lazarus, *Fungal Genet Biol* **1999**, *26*, 209-223.

[23] T. P. Nicholson, B. A. Rudd, M. Dawson, C. M. Lazarus, T. J. Simpson, R. J. Cox, *Chem Biol* **2001**, *8*, 157-178.

[24] J. M. Crawford, B. C. R. Dancy, E. A. Hill, D. W. Udwary, C. A. Townsend, *Proc Natl Acad Sci USA* **2006**, *103*, 16728-16733.

[25] J. M. Crawford, T. P. Korman, J. W. Labonte, A. L. Vagstad, E. A. Hill, O. Kamari-Bidkorpeh, S. C. Tsai, C. A. Townsend, *Nature* **2009**, *461*, 1139-U1243.

[26] J. M. Crawford, P. M. Thomas, J. R. Scheerer, A. L. Vagstad, N. L. Kelleher, C. A. Townsend, *Science* **2008**, *320*, 243-246.

[27] (a) P. K. Chang, J. W. Cary, J. Yu, D. Bhatnagar, T. E. Cleveland, *Mol Gen Genet* **1995**, *248*, 270-277; (b) C. M. Watanabe, D. Wilson, J. E. Linz, C. A. Townsend, *Chem Biol* **1996**, *3*, 463-469; (c) C. M. Watanabe, C. A. Townsend, *Chem Biol* **2002**, *9*, 981-988; (d) J. Yu, P. K. Chang, K. C. Ehrlich, J. W. Cary, D. Bhatnagar, T. E. Cleveland, G. A. Payne, J. E. Linz, C. P. Woloshuk, J. W. Bennett, *Appl Environ Microbiol* **2004**, *70*, 1253-1262; (e) K. Yabe, H. Nakajima, *Appl Microbiol Biotechnol* **2004**, *64*, 745-755.

[28] C. A. Townsend, S. B. Christensen, K. Trautwein, *J Am Chem Soc* **1984**, *106*, 3868-3869.

[29] (a) S. W. Brobst, C. A. Townsend, *Can J Chem* **1994**, *72*, 200-207; (b) T. S. Hitchman, E. W. Schmidt, F. Trail, M. D. Rarick, J. E. Linz, C. A. Townsend, *Bioorg Chem* **2001**, *29*, 293-307.

[30] K. C. Ehrlich, P. Li, L. Scharfenstein, P. K. Chang, *Appl Environ Microbiol* **2010**, *76*, 3374-3377.

[31] S. P. Hussain, J. Schwank, F. Staib, X. W. Wang, C. C. Harris, *Oncogene* **2007**, *26*, 2166-2176.

[32] H. Klenow, I. Henningsen, *Proc Natl Acad Sci U S A* **1970**, *65*, 168-175.

[33] Y. Tsukamoto, H. Wong, J. S. Mattick, S. J. Wakil, *J Biol Chem* **1983**, *258*, 15312-15322.

[34] D. W. Udwary, M. Merski, C. A. Townsend, *J Mol Biol* **2002**, *323*, 585-598.

[35] J. M. Crawford, A. L. Vagstad, K. R. Whitworth, K. C. Ehrlich, C. A. Townsend, *Chembiochem* **2008**, *9*, 1019-1023.

[36] D. W. Brown, S. H. Lee, L. H. Kim, J. G. Ryu, S. Lee, Y. Seo, Y. H. Kim, M. Busman, S. H. Yun, R. H. Proctor, T. Lee, *Mol Plant Microbe Interact* **2014**.

[37] C. R. Huitt-Roehl, E. A. Hill, M. M. Adams, A. L. Vagstad, J. W. Li, C. A. Townsend, *ACS Chem Biol* **2015**, *10*, 1443-1449.

[38] J. M. Winter, D. Cascio, D. Dietrich, M. Sato, K. Watanabe, M. R. Sawaya, J. C. Vederas, Y. Tang, *J Am Chem Soc* **2015**, *137*, 9885-9893.

[39] J. M. Winter, M. Sato, S. Sugimoto, G. Chiou, N. K. Garg, Y. Tang, K. Watanabe, *J Am Chem Soc* **2012**, *134*, 17900-17903.

[40] Y. M. Chiang, E. Szewczyk, A. D. Davidson, N. Keller, B. R. Oakley, C. C. Wang, *J Am Chem Soc* **2009**, *131*, 2965-2970.

[41] (a) H. Zhou, J. Zhan, K. Watanabe, X. Xie, Y. Tang, *Proc Natl Acad Sci U S A* **2008**, *105*, 6249-6254; (b) H. Zhou, K. Qiao, Z. Gao, M. J. Meehan, J. W. Li, X. Zhao, P. C. Dorrestein, J. C. Vederas, Y. Tang, *J Am Chem Soc* **2010**, *132*, 4530-4531; (c) H. Zhou, K. Qiao, Z. Gao, J. C. Vederas, Y. Tang, *J Biol Chem* **2010**, *285*, 41412-41421.

[42] A. L. Vagstad, S. B. Bumpus, K. Belecki, N. L. Kelleher, C. A. Townsend, *J Am Chem Soc* **2012**, *134*, 6865-6877.

[43] S. M. Ma, J. Zhan, K. Watanabe, X. Xie, W. Zhang, C. C. Wang, Y. Tang, *J Am Chem Soc* **2007**, *129*, 10642-10643.

[44] A. T. Keatinge-Clay, D. A. Maltby, K. F. Medzihradszky, C. Khosla, R. M. Stroud, *Nat Struct Mol Biol* **2004**, *11*, 888-893.

[45] (a) G. A. Zornetzer, B. G. Fox, J. L. Markley, *Biochemistry* **2006**, *45*, 5217-5227; (b) A. Roujeinikova, W. J. Simon, J. Gilroy, D. W. Rice, J. B. Rafferty, A. R. Slabas, *J Mol Biol* **2007**, *365*, 135-145; (c) M. Leibundgut, S. Jenni, C. Frick, N. Ban, *Science* **2007**, *316*, 288-290; (d) C. Nguyen, R. W. Haushalter, D. J. Lee, P. R. L. Markwick, J. Bruegger, G. Caldara-Festin, K. Finzel, D. R. Jackson, F. Ishikawa, B. O'Dowd, J. A. McCammon, S. J. Opella, S.-C. Tsai, M. D. Burkart, *Nature* **2014**, *505*, 427-431.

[46] P. Wattana-Amorn, C. Williams, E. Ploskon, R. J. Cox, T. J. Simpson, J. Crosby, M. P. Crump, *Biochemistry* **2010**, *49*, 2186-2193.

[47] (a) R. Thomas, *Chembiochem* **2001**, *2*, 612-627; (b) H. Zhou, Y. Li, Y. Tang, *Nat Prod Rep* **2010**, *27*, 839-868.

[48] G. Bringmann, T. A. M. Gulder, A. Hamm, M. Goodfellow, H.-P. Fiedler, *Chem Commun* **2009**, 6810-6812.

[49] (a) B. D. Ames, T. P. Korman, W. Zhang, P. Smith, T. Vu, Y. Tang, S.-C. Tsai, *Proc Natl Acad Sci USA* **2008**, *105*, 5349-5354; (b) M. Y. Lee, B. D. Ames, S. C. Tsai, *Biochemistry* **2012**, *51*, 3079-3091.

[50] Y. R. Li, W. Xu, Y. Tang, *J Biol Chem* **2010**, *285*, 22762-22771.

[51] (a) S. C. Dillon, A. Bateman, *BMC Bioinformatics* **2004**, *5*, 109; (b) J. W. Labonte, C. A. Townsend, *Chem Rev* **2013**, *113*, 2182-2204.

[52] A. Keatinge-Clay, *J Mol Biol* **2008**, *384*, 941-953.

[53] M. Leesong, B. S. Henderson, J. R. Gillig, J. M. Schwab, J. L. Smith, *Structure* **1996**, *4*, 253-264.

[54] F. Hibbert, J. Emsley, in *Adv Phys Org Chem, Vol. Volume 26* (Ed.: D. Bethell), Academic Press, **1991**, pp. 255-379.

[55] S. G. Mills, P. Beak, *J Org Chem* **1985**, *50*, 1216-1224.

[56] L. C. Du, L. L. Lou, *Nat Prod Rep* **2010**, *27*, 255-278.

[57] (a) I. Fujii, A. Watanabe, U. Sankawa, Y. Ebizuka, *Chem Biol* **2001**, *8*, 189-197; (b) T. P. Korman, J. M. Crawford, J. W. Labonte, A. G. Newman, J. Wong, C. A. Townsend, S. C. Tsai, *Proc Natl Acad Sci USA* **2010**, *107*, 6246-6251.

[58] (a) A. Watanabe, Y. Ono, I. Fujii, U. Sankawa, M. E. Mayorga, W. E. Timberlake, Y. Ebizuka, *Tetrahedron Lett* **1998**, *39*, 7733-7736; (b) A. Watanabe, I. Fujii, U. Sankawa, M. E. Mayorga, W. E. Timberlake, Y. Ebizuka, *Tetrahedron Lett* **1999**, *40*, 91-94.

[59] D. L. Ollis, E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S. M. Franken, M. Harel, S. J. Remington, I. Silman, J. Schrag, J. L. Sussman, K. H. G. Verschueren, A. Goldman, *Protein Eng* **1992**, *5*, 197-211.

[60] P. Carter, J. A. Wells, *Nature* **1988**, *332*, 564-568.

[61] D. P. Frueh, H. Arthanari, A. Koglin, D. A. Vosburg, A. E. Bennett, C. T. Walsh, G. Wagner, *Nature* **2008**, *454*, 903-906.

[62] A. L. Vagstad, E. A. Hill, J. W. Labonte, C. A. Townsend, *Chem Biol* **2012**, *19*, 1525-1534.

[63] M. H. Wheeler, D. Abramczyk, L. S. Puckhaber, M. Naruse, Y. Ebizuka, I. Fujii, P. J. Szaniszlo, *Eukaryot Cell* **2008**, *7*, 1699-1711.

[64] (a) Y. Xu, T. Zhou, S. Zhang, L.-J. Xuan, J. Zhan, I. Molnar, *J Am Chem Soc* **2013**, *135*, 10783-10791; (b) M. E. Horsman, T. P. Hari, C. N. Boddy, *Nat Prod Rep* **2015**.

[65] (a) C. D. Reeves, Z. Hu, R. Reid, J. T. Kealey, *Appl Environ Microbiol* **2008**, *74*, 5121-5129; (b) H. Zhou, K. Qiao, Z. Gao, J. C. Vederas, Y. Tang, *J Biol Chem* **2010**, *285*, 41412-41421.

[66] S. Wang, Y. Xu, E. A. Maine, E. M. K. Wijeratne, P. Espinosa-Artiles, A. A. L. Gunatilaka, I. Molnár, *Chem Biol* **2008**, *15*, 1328-1338.

[67] I. Gaffoor, F. Trail, *Appl Environ Microbiol* **2006**, *72*, 1793-1799.

[68] K. Kusumoto, Y. Nogata, H. Ohta, *Curr Genet* **2000**, *37*, 104-111.

[69] J. W. Cary, K. C. Ehrlich, S. B. Beltz, P. Harris-Coward, M. A. Klich, *Mycologia* **2009**, *101*, 352-362.

[70] D. M. Geiser, J. W. Dorner, B. W. Horn, J. W. Taylor, *Fungal Genet Biol* **2000**, *31*, 169-179.

[71] J. H. Yu, T. J. Leonard, *J Bacteriol* **1995**, *177*, 4792-4800.

[72] R. E. Bradshaw, H. Jin, B. S. Morgan, A. Schwelm, O. R. Teddy, C. A. Young, S. Zhang, *Mycopathologia* **2006**, *161*, 283-294.

[73] Y. Li, Y.-H. Chooi, Y. Sheng, J. S. Valentine, Y. Tang, *J Am Chem Soc* **2011**, *133*, 15773-15785.

[74] Y.-H. Chooi, R. Cacho, Y. Tang, *Chem Biol* **2010**, *17*, 483-494.

[75] E. Szewczyk, Y. M. Chiang, C. E. Oakley, A. D. Davidson, C. C. Wang, B. R. Oakley, *Appl Environ Microbiol* **2008**, *74*, 7607-7612.

[76] T. Awakawa, K. Yokota, N. Funa, F. Doi, N. Mori, H. Watanabe, S. Horinouchi, *Chem Biol* **2009**, *16*, 613-623.

[77] Y.-M. Chiang, E. Szewczyk, A. D. Davidson, R. Entwistle, N. P. Keller, C. C. C. Wang, B. R. Oakley, *Appl Environ Microbiol* **2010**, *76*, 2067-2074.

[78] F. Y. Lim, Y. Hou, Y. Chen, J.-H. Oh, I. Lee, T. S. Bugni, N. P. Keller, *Appl Environ Microbiol* **2012**, *78*, 4117-4125.

[79] K. Throckmorton, P. Wiemann, N. P. Keller, *Toxins* **2015**, *7*, 3572-3607.

[80]   A. G. Newman, A. L. Vagstad, K. Belecki, J. R. Scheerer, C. A. Townsend, *Chem Commun* **2012**, *48*, 11772-11774.

[81]   H.-F. Tsai, I. Fujii, A. Watanabe, M. H. Wheeler, Y. C. Chang, Y. Yasuoka, Y. Ebizuka, K. J. Kwon-Chung, *J Biol Chem* **2001**, *276*, 29292-29298.

[82]   Y.-M. Chiang, K. M. Meyer, M. Praseuth, S. E. Baker, K. S. Bruno, C. C. C. Wang, *Fungal Genet Biol* **2011**, *48*, 430-437.

[83]   J.-E. Kim, K.-H. Han, J. Jin, H. Kim, J.-C. Kim, S.-H. Yun, Y.-W. Lee, *Appl Environ Microbiol* **2005**, *71*, 1701-1708.

[84]   T. Nakazawa, K. i. Ishiuchi, A. Praseuth, H. Noguchi, K. Hotta, K. Watanabe, *ChemBioChem* **2012**, *13*, 855-861.

[85]   J. F. Sanchez, Y. M. Chiang, E. Szewczyk, A. D. Davidson, M. Ahuja, C. E. Oakley, J. W. Bok, N. Keller, B. R. Oakley, C. C. C. Wang, *Mol Biosyst* **2010**, *6*, 587-593.

Chapter 2: Analysis of the Cercosporin Polyketide Synthase CTB1 Reveals a New Fungal Thioesterase Function

This chapter was adapted with permission from A. G. Newman, A. L. Vagstad, K. Belecki, J. R. Scheerer, C. A. Townsend, *Chem. Commun.* **2012**, *48*, 11772-11774.

2.1. Introduction

Cercosporin (**2**, Figure 2.1) is a photoactivated phytotoxin of polyketide origin produced by several *Cercospora* species.[1] These fungal species compose destructive plant pathogens in which cercosporin plays a central role. The perylenequinone core of the molecule is essential to its toxicity. The metabolite acts as a photosensitizing agent upon absorption of visible light, allowing for the production of singlet oxygen ($^1O_2$) and superoxide radical (O_2^-) with impressive quantum yields.[2] Cercosporin does not have a known direct cellular target, but causes indiscriminate oxidative damage to cell membranes, proteins, lipids, and nucleic acids.[3] Other perylenequinones share this photosensitizing activity with cercosporin.[4] While the putative cercosporin biosynthetic gene cluster has been identified in the tobacco pathogen *C. nicotianae*, little is known about the metabolite's assembly.[5] The gene cluster comprises eight proteins, including six enzymes believed to be responsible for the construction of the metabolite,[5-6] an MFS transporter,[7] and a zinc finger transcription factor that regulates expression of the cluster.[5b]

It is postulated that cercosporin biosynthesis is initiated by CTB1—an iterative, non-reducing polyketide synthase (NR-PKS). CTB1 comprises the characteristic basis set of six catalytic domains of the NR-PKS family found in fungi including a tandem acyl-carrier protein (ACP$_2$)—a common variant.[5a] The starter unit:ACP acyltransferase (SAT) domain of CTB1 initiates polyketide extension through the selective utilization of
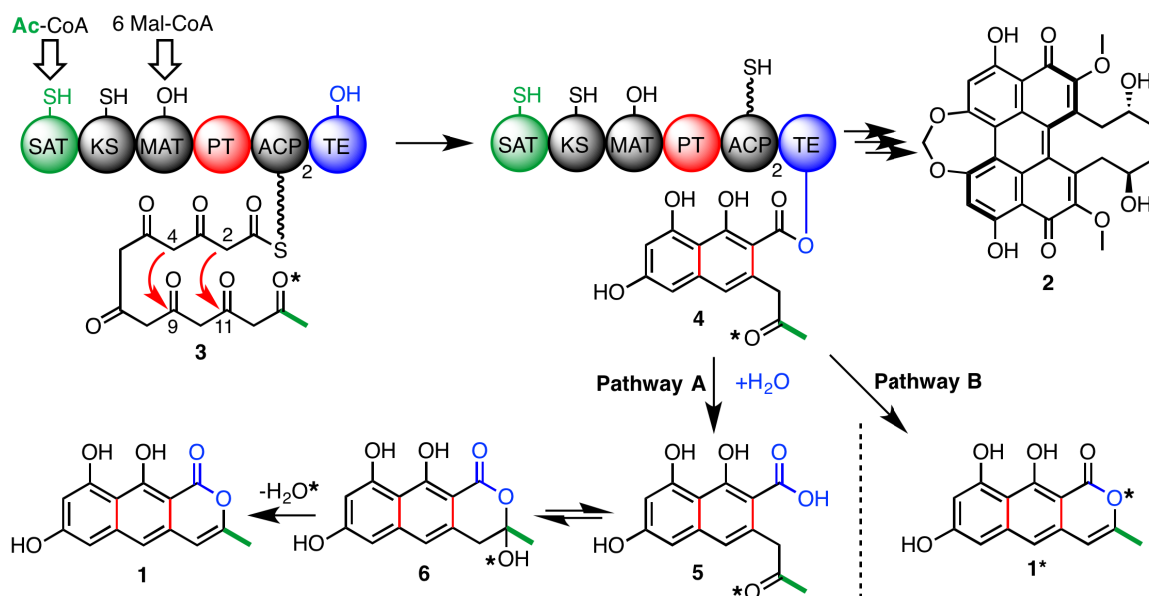
Figure 2.1 | Possible mechanisms for *nor*-toralactone formation by CTB1. Two mechanisms were considered. In pathway **a**, the product **1** is released by hydrolysis followed by pyrone formation. In pathway **b**, *nor*-toralactone (**1***) is formed directly through enol lactonization.

acetyl-CoA,[8] which is predicted to be elongated to the heptaketide **3** by the β-ketoacyl synthase (KS) domain through successive decarboxylative Claisen condensations with six malonyl units introduced by the malonyl:ACP acyltransferase (MAT) domain. It is believed that the product template (PT) domain catalyzes C4–C9 and C2–C11 aldol cyclizations and aromatizations generating an enzyme bound bicyclic intermediate. The intermediate is delivered to the thioesterase (TE) domain as bicycle **4** for final product release.[5]

TE domains catalyze product release from fatty acid synthases, non-ribosomal peptide synthetases, and polyketide synthases. In the main, these reactions occur by simple hydrolysis but in the latter two protein families can take place by macrolactamization or macrolactonization, respectively.[9] For the subset of iterative, NR-PKSs, TE domains play a critical "editing" role during polyketide extension[10] in

addition to terminal synthetic roles in carbon–carbon bond formation (Claisen/Dieckmann ring formation)[11] and cleavage (deacylation).[12]

The anticipated product of CTB1 is carboxylic acid **5**—the result of conventional TE-catalyzed hydrolysis. Herein we report that—contrary to expectation—the cleanly observed major product of CTB1 is the naphthopyrone *nor*-toralactone (**1**). The two simplest mechanisms (Figure 2.1) that can be proposed for *nor*-toralactone formation are (**a**) canonical TE-mediated hydrolysis to carboxylic acid **5**, which can be visualized to form the pseudoacid **6** and dehydrate to **1**, or (**b**) the TE-bound bicyclic intermediate **4** could be directly cyclized to **1*** by either rapid spontaneous enol lactonization or TE-catalyzed pyrone formation—an unprecedented TE function. Using an enzyme deconstruction approach, we investigated the *in vitro* biochemistry of CTB1 and further probed the mechanism of release catalyzed by the TE domain.

2.2. Results

2.2.1. Deconstruction of CTB1

Protein deconstruction has proved to be a powerful tool for determining PKS activity *in vitro*.[11a] This method requires the identification of interdomain regions in multidomain proteins, which—in this instance—was achieved using bioinformatics methods.[13] Hexahistidine constructs of mono- and multidomain protein fragments were prepared for heterologous expression in *E. coli* and nickel-affinity purification (see Appendix C). The constructs used in these experiments included a SAT-KS-MAT tridomain and PT, ACP$_2$, and TE monodomains (Appendix C, Table C.2). Additionally, S2008A and H2171Q active-site point mutants of the TE monodomain were prepared
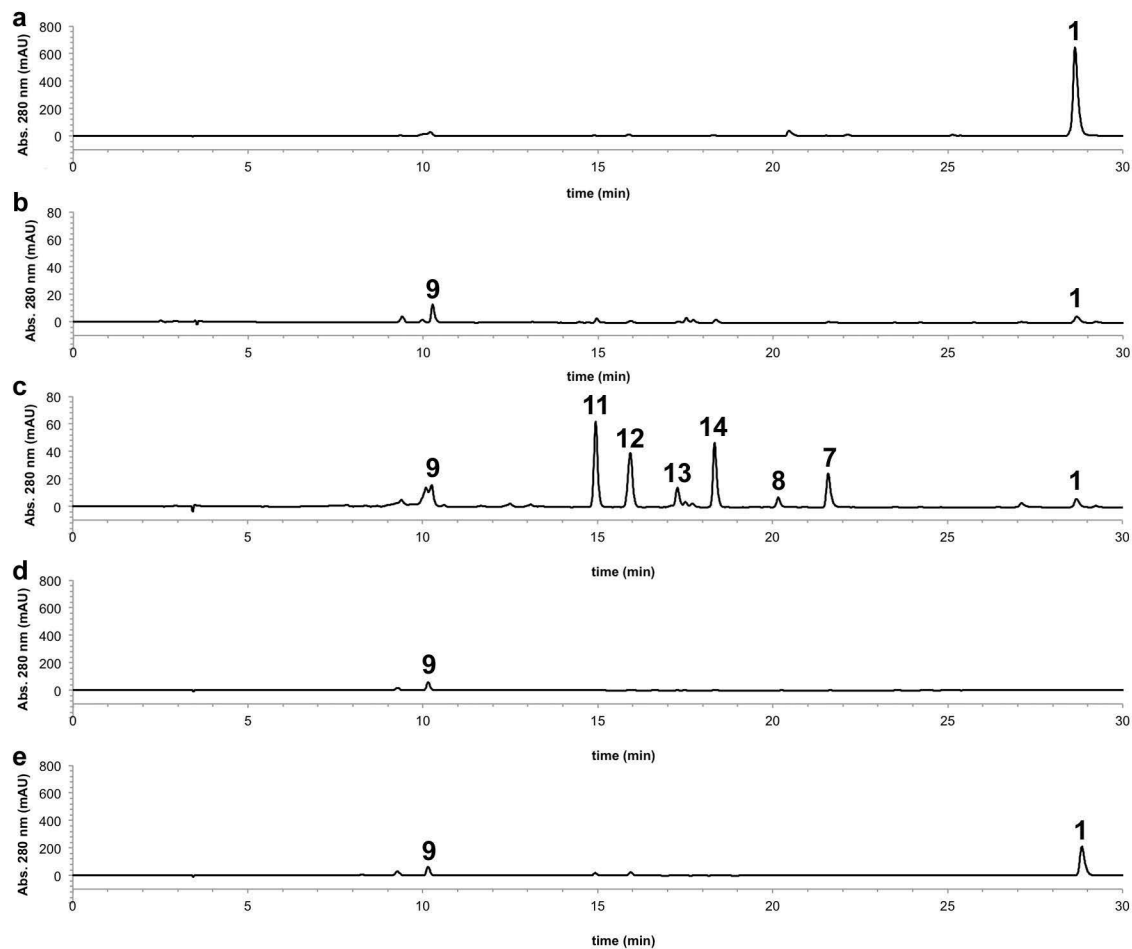
Figure 2.2 │ Extracted products of *in vitro* reconstituted CTB1 reactions. HPLC chromatograms at 280 nm are presented of reactions of a fully reconstituted CTB1 (**a**), a reconstituted CTB1 lacking the TE domain (**b**), a reconstituted minimal CTB1 (SAT-KS-MAT + ACP$_2$, **c**), a fully reconstituted CTB1 with TE-S2008A mutant (**d**), and a fully reconstituted CTB1 with TE-H2171Q mutant (**e**). For clarity, the scale of the *y*-ordinate in parts **b** and **c** are an order of magnitude less than the other chromatograms.

(Appendix C, Table C.2). Protein purity was confirmed by 12% SDS-PAGE analysis

(Appendix C, Figure C.1).

2.2.2. Reconstitution of CTB1 Activity

Recombination of deconstructed CTB1 protein fragments to generate complete

and partial NR-PKS systems *in vitro* led to reconstituted enzymatic activity and

deduction of individual domain function. The complete set of CTB1 domains (SAT-KS-

MAT + PT + ACP$_2$ + TE) was combined in equimolar concentrations with acetyl-CoA

41
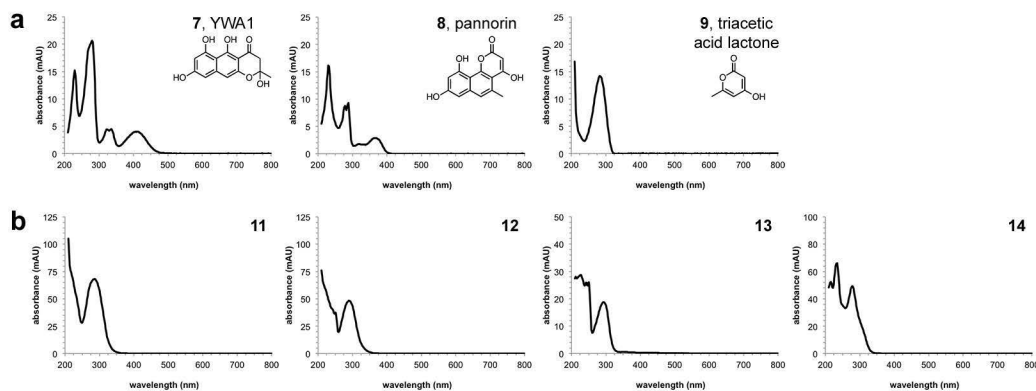
Figure 2.3│UV-visible spectra for derailment products of minimal CTB1. UV-vis spectra are presented for fully (**a**) and partially (**b**) characterized derailment products.

Table 2.1│HRMS–ESI-IT-TOF for derailment products of minimal CTB1.

Peak	*m/z*: [MH⁺] calcd for:	found:
7	YWA1, $C_{14}H_{13}O_6$, 277.0712	277.0609 (100%)
8	pannorin, $C_{14}H_{11}O_5$, 259.0607	259.0597 (100%)
9	triacetic acid lactone, $C_6H_7O_3$, 127.0395	127.0395 (100%)
11	$C_{14}H_{13}O_6$, 277.0712	277.0703 (100%), 299.0515 (MNa⁺, 40.99%)
12	$C_{14}H_{13}O_6$, 277.0712	277.0705 (100%)
13	$C_{13}H_{13}O_4$, 233.0814	233.0802 (100%)
14	$C_{14}H_{13}O_6$, 277.0712	277.0703 (100%), 299.0528 (MNa⁺, 12.45%)

and malonyl-CoA. The major product of this reaction (Figure 2.2a) was *nor*-toralactone (**1**) as determined by comparison to a standard prepared by total synthesis (See Appendix C). *In vitro* reactions of CTB1 were carried out in which the TE domain was excluded (SAT-KS-MAT + PT + ACP₂). The overall synthetic activity of CTB1 was dramatically reduced (*ca.* 50-fold) in these reactions as measured by total product output (Figure 2.2b). Naphthopyrone **1** was produced in small amounts while no other major product was observed. In keeping with this observation, a mutant TE domain in which the catalytic Ser2008 was replaced with Ala was similarly unproductive (Figure 2.2d). However, when the catalytic His2171 was mutated to Gln, the reaction proceeded at about two thirds of the wild type production (Figure 2.2e).
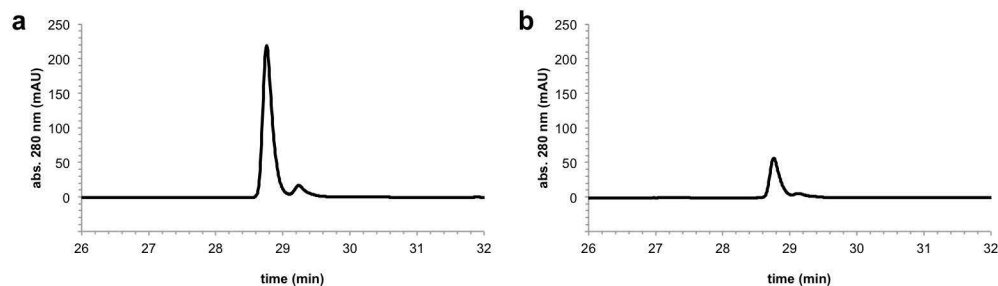
Figure 2.4│Demonstration of the "starter-unit effect" in CTB1. *Nor*-toralactone is differentially produced by a fully reconstituted CTB1 *in vitro* from acetyl-CoA and malonyl-CoA (**a**) or malonyl-CoA only (**b**).

A minimal CTB1 system in which both the PT and TE domains were excluded (SAT-KS-MAT + ACP$_2$) was proficient in acetate polyketide extension, leading to a variety of "derailment" products (Figure 2.2c). These derailment products were consistent with full-length heptaketide intermediates as determined by LCMS (Figure 2.3, Table 2.1). Two such products were structurally characterized to be YWA1 (**7**)[14] and pannorin (**8**)[15] by comparison to known chemical properties and an authentic standard, respectively. Additionally, triacetic acid lactone (**9**) was observed, a product of self-condensation from a truncated triketide intermediate (Figure 2.3a).[16] All derailment products were produced in significantly lower quantities when the PT or TE was present.

2.2.3. Selective Heavy-Isotope Incorporation into CTB1 Products

To investigate the CTB1 TE mechanism, an ^{18}O heavy-isotope label was incorporated at the O13 position of the heptaketide linear intermediate **3**. The fate of this oxygen could be used to distinguish between TE-catalyzed hydrolysis and enol lactonization (Figure 2.1). The O13 position is uniquely derived from the starter unit acetyl precursor, allowing for selective labeling of this position. Reconstitution reactions in which acetyl-CoA was withheld showed less than a quarter of the production of *nor*-
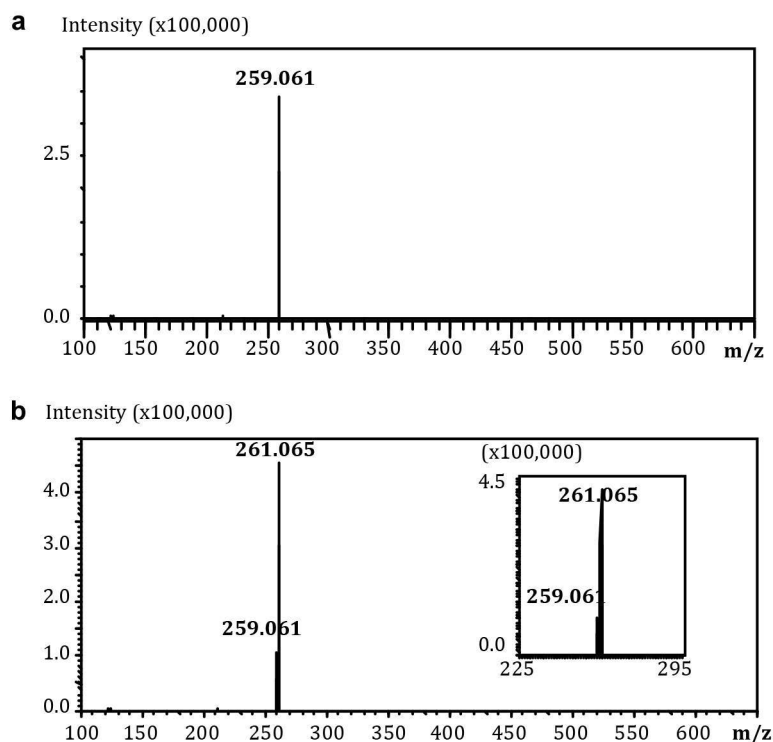
Figure 2.5 │ Observed mass shift for *nor*-toralactone in CTB1 reconstitution reactions. Reactions of fully reconstituted CTB1 (SAT-KS-MAT + PT + ACP$_2$ +TE) using malonyl-CoA and unlabeled (**a**) or heavy-isotope labeled (**b**) acetyl-CoA are indicative of heavy-isotope retention in *nor*-toralactone.

toralactone by HPLC relative to the same reaction in which both acetyl-CoA and malonyl-CoA were included (Figure 2.4).

Reconstitution reactions were carried out as before with either unlabeled acetyl-CoA or [^{18}O]acetyl-CoA. [^{18}O]Acetyl-CoA was prepared enzymatically and HPLC purified. Mass spectrometric analysis showed the [^{18}O]acetyl-CoA had an exact mass consistent with incorporation with a single heavy-atom (*m/z*: [MH$^+$] calcd for C$_{23}$H$_{39}$N$_7$O$_{16}$^{18}OP$_3$S, 812.1379; found 812.1328). A mass indicative of no heavy-atom incorporation was not detected. The products of the reconstitution reactions were analyzed by LCMS to measure heavy isotope content. Naphthopyrone **1** produced by the fully reconstituted CTB1 (SAT-KS-MAT + PT + ACP$_2$ + TE) showed a distinct mass shift consistent with the incorporation of a single ^{18}O atom (*m/z*: [MH$^+$] calcd for

44

$C_{14}H_{11}O_4{}^{18}O$ 261.0649; found 261.0651 [100%]). However, a minor portion did not contain the heavy isotope (*m/z*: [MH$^+$] calcd for $C_{14}H_{11}O_5$ 259.0607; found 259.0610 [27%], Figure 2.5).

In the derailment products of the minimal CTB1 (SAT-KS-MAT + ACP$_2$), ^{18}O incorporation was the minor component relative to the unlabeled component. YWA1 (**7**) showed very little ^{18}O incorporation (*m/z*: [MH$^+$] calcd for $C_{14}H_{13}O_6$ 277.0712; found 277.0691 [100%], 279.0725 [3%]). No ^{18}O incorporation was observed in the derailment product pannorin (**8**, *m/z*: [MH$^+$] calcd for $C_{14}H_{11}O_5$ 259.0607; found 259.0596 [100%]; *m/z*: [MH$^+$] calcd for $C_{14}H_{11}O_4{}^{18}O$ 261.0649; mass not found). On the other hand, triacetic acid lactone (**9**) was substantially enriched in ^{18}O (*m/z*: [MH$^+$] calcd for $C_6H_7O_2{}^{18}O$ 129.0438; found 129.0430 [100%], 127.0392 [37%]).

2.3. Discussion

The proposed product of CTB1 is carboxylic acid **5**, making the appearance of *nor*-toralactone (**1**) unexpected. While it is tempting to interpret pyrone formation as resulting from direct enol lactonization, spontaneous intramolecular cyclization and dehydration of the bicyclic intermediate **4** could produce *nor*-toralactone without the intervention of the TE domain (Figure 2.1, pathway **a**). Indeed, enol lactonization is the most commonly observed spontaneous off-loading mechanism for NR-PKSs lacking a TE domain.[17] In the classic example of the norsolorinic acid synthase PksA from *Aspergillus parasiticus*, the product norpyrone (**15**) is preferentially formed over norsolorinic acid anthrone (**16**) when the TE domain is withheld (Figure 2.6).[11a, 18] In fact, C–O bond formation is irreversible, as PksA TE was unable to hydrolyze the pyrone of norpyrone to form norsolorinic acid anthrone—its native product—despite the

45
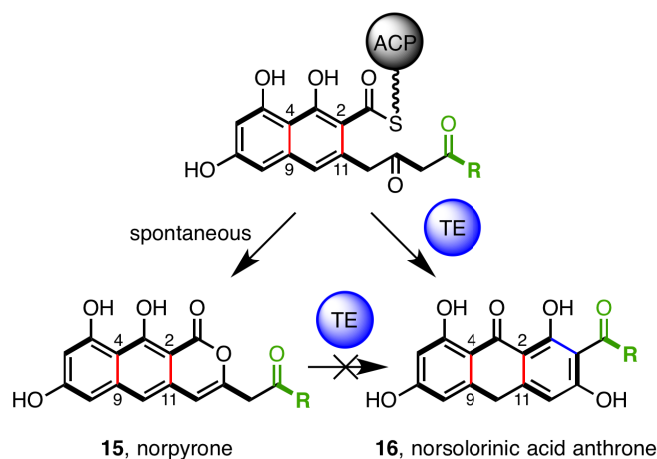
Figure 2.6 | Spontaneous versus TE-catalyzed release of the PksA penultimate intermediate.

predicted stabilizing force of C–C bond formation and aromatization.[11a] We sought to resolve the CTB1 TE domain's mechanistic ambiguity, given its peculiar activity relative to established NR-PKS chemistry.

As in other NR-PKSs, spontaneous pyrone formation was the predicted release mechanism in reactions without the TE domain. However, the low production of *nor*-toralactone in reactions lacking the CTB1 TE domain indicated a slow rate of spontaneous release, strongly suggesting a central role for the TE domain in accelerating pyrone formation. In keeping with this observation, the abolishment of *nor*-toralactone production in the TE S2008A active site mutant strengthens the argument of the TE's role in product evolution. The high reactivity of the H2171Q catalytic triad mutant was surprising. In conventional TE mechanisms, the catalytic His is believed to serve as a general base during the formation of the acyl-enzyme intermediate and then again to facilitate product-releasing nucleophilic attack of the acyl-enzyme intermediate.[9, 11b] A different residue may serve this role in the mutant TE thereby allowing for the increased residual activity. Alternatively, functionally related enzymes with the α/β-hydrolase family have been shown to be relatively tolerant to the His to Gln mutation.[19] It is
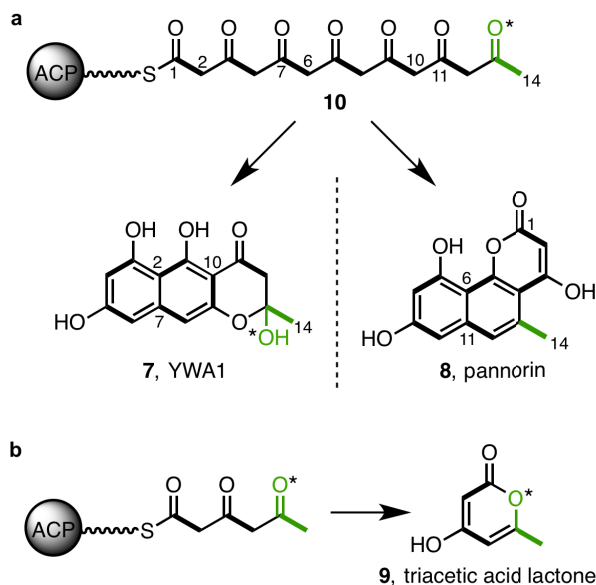
Figure 2.7 │ Origin of CTB1 derailment products. Spontaneous cyclization of linear heptaketide intermediate **10** led to the production of YWA1 (**7**) and pannorin (**8**, **a**). A common truncation product of NR-PKSs is triacetic acid lactone (**9**, **b**). The ketide unit derived from the acetyl-CoA starter unit is displayed in green. The ^{18}O-labeled position is denoted with an asterisk.

proposed that these enzymes do not utilize the classical serine protease mechanism.

Rather, the His-Ser dyad promotes the formation of a reactive intermediate that directly

activates the catalytic Ser for nucleophilic attack. It is unclear if such a mechanism is at

play in CTB1.

Derailment products of a minimal NR-PKS lacking the PT and TE domains are to

be expected. Without the catalytic guidance of these tailoring domains, a thermodynamic

distribution of low-energy spontaneous aldol cyclizations and intramolecular

transesterifications of the linear poly-β-ketone intermediate will predominate.[11a] The

mass signatures of the observed minimal CTB1 derailment products are consistent with

this view of derailment evolution. Specifically, YWA1 (**7**) and pannorin (**8**) can be

envisioned as deriving from the same linear heptaketide intermediate **10** (Figure 2.7).

Further, the uncharacterized derailment products had masses consistent with varying

levels of dehydration or decarboxylation of the same linear intermediate **10**. All minimal CTB1 derailment products were produced at significantly lower quantities when the PT or TE was present, underscoring the central roles of these domains in both directed cyclization and overall catalytic efficiency.

Despite solidifying its overall role in catalysis, these observations did not address the mechanism by which the TE domain catalyzes the synthesis of *nor*-toralactone. As noted, two possible mechanisms were considered (Figure 2.1). The fate of the polyketide oxygen at C13 is different for each mechanism. In pathway **a** (hydrolysis), the oxygen would be lost as water, while in pathway **b** (enol lactonization) it would be retained as the endocyclic oxygen of the pyrone moiety. In NR-PKSs, the SAT domain is responsible for the observed "starter-unit effect" by selectively loading the ACP with a specific acyl-thioester substrate.[8, 20] While the starter unit for CTB1 is the relatively simple acetyl-CoA, other SATs select more distinctive starter units—usually produced by an upstream biosynthetic module. The results often have a profound effect on the product output of the given NR-PKS.[21] In the case of PksA, the starter unit is a hexanoyl thioester produced by a dedicated fungal fatty acid synthase.[22] In reconstitution reactions in which the occupancy of individual domains of PksA were assayed, it was determined that polyketide biosynthesis will not commence until the proper hexanoyl starter unit is transferred to the enzyme, underscoring the critical role the starter unit effect plays on overall turnover.[10] Because the O13 atom in the CTB1 linear intermediate **3** is uniquely derived from its starter unit acetyl-CoA, we were able to selectively label it through the use of [^{18}O]acetyl-CoA. CTB1 demonstrated a significant starter unit effect with *nor*-toralactone production decreased to less than 25% when acetyl-CoA was withheld. The

residual activity was believed to arise from acetyl-CoA contamination of malonyl-CoA stocks through spontaneous decarboxylation, but could also arise from PKS-mediated decarboxylation following malonyl loading. The strength of the starter-unit effect allowed for mechanistic determination by observing selective heavy-isotope incorporation into the CTB1 products.

The heavy-isotope incorporation data support pathway **b** and direct enol lactonization by the TE domain. The ^{18}O heavy isotope was clearly incorporated into *nor*-toralactone. The fraction of *nor*-toralactone in which the heavy isotope was not incorporated was estimated at about a quarter by relative mass intensity. This fraction was attributed to unlabeled acetyl incorporation, as the proportion was similar to the residual production in reactions where acetyl-CoA was withheld. Furthermore, heavy-isotope incorporation data for the derailment products was entirely consistent with mechanism **b**. In YWA1 (**7**), ^{18}O was incorporated into a very minor proportion of molecules as determined by relative mass intensity. As the expected labeled oxygen atom is part of the hemiketal moiety in YWA1, the loss of the heavy isotope as water was expected (Figure 2.7a). In pannorin (**8**), no heavy isotope was retained, because it was eliminated as water during the aldol cyclization/aromatization needed to access this scaffold (Figure 2.7a). Finally, the heavy isotope is retained as the endocyclic oxygen of triacetic acid lactone (**9**, Figure 2.7b). As with *nor*-toralactone, the percentage of heavy-isotope incorporation in triacetic acid lactone roughly matches that of unlabeled acetyl-CoA usage.

The isotope incorporation data are unambiguous that *nor*-toralactone is the direct product of CTB1 (pathway **b**). Furthermore, the strong dependence of the TE domain in

nor-toralactone evolution eliminates the possibility of pathway **b** occurring spontaneously. The mechanism by which the TE domain of CTB1 favors pyrone formation over hydrolysis is not precisely known. It is likely that the thioesterase promotes keto/enol tautomerization in the active site thereby favoring enol lactonization. In contrast, TE substrates undergoing Claisen cyclization for product release contain a β-diketone side chain, not a methyl ketone as in *nor*-toralactone, whose lower pK_a (*ca.* 9 *vs.* 19) doubtless plays a determining role in the catalysis of carbon–carbon bond formation.[11]

2.4. Conclusions

The NR-PKS of cercosporin biosynthesis in *C. nicotianae*, CTB1, catalyzes the formation of the naphthopyrone *nor*-toralactone. The formation of *nor*-toralactone is unexpected and contrary to the proposed biosynthetic pathway for cercosporin. Nevertheless, direct formation of *nor*-toralactone by CTB1 is unambiguous, as demonstrated by heavy-isotope incorporation experiments. The data support pyrone formation through direct enol lactonization catalyzed by the TE domain—a hitherto unknown chemical paradigm for NR-PKS thioester domains. The underpinning mechanism by which enol lactonization is favored in the CTB1 TE domain remains unknown and warrants further investigation. The new understanding of CTB1 catalysis presented here must be reconciled with the currently held biosynthetic pathway for cercosporin.

2.5. Experimental Methods

2.5.1. Reagents and Biological Strains

All reagents and primers were purchased from Sigma-Aldrich (St. Louis, MO, USA), unless otherwise specified. Magnesium chloride, sodium chloride, potassium phosphate monobasic, potassium phosphate dibasic, and deoxynucleotides (dNTPs) were purchased from Thermo Scientific (Waltham, MA, USA). Isopropyl-β-D-thiogalactoside (IPTG), kanamycin, and nickel (high density) agarose beads were purchased from Gold Biotechnology (St. Louis, MO, USA). Imidazole was purchased from Acros Organics (Geel, Belgium). Sodium acetate (^{18}O, 95%) was from Cambridge Isotope Laboratories (Cambridge, MA, USA). Yeast extract and tryptone were purchased from Boston BioProducts (Ashland, MA, USA). Agar was purchased from bioWORLD (Dublin, OH, USA). Plasmids pET-24a(+) and pET-28a(+) were purchased from EMD Millipore (Darmstadt, Germany). All enzymes used for manipulation of DNA and 10 × HF Buffer were purchased from New England Biolabs (Ipswich, MA, USA). Bio-Rad protein assay dye for determining protein concentration by the Bradford assay was purchased from Bio-Rad laboratories (Hercules, CA, USA). Bacterial strain *Escherichia coli* BL21(DE3) was purchased from Sigma-Aldrich and made electrocompetent by standard protocols (Appendix B).[23] Genomic DNA purified from *C. nicotianae* (ATCC© 18366™) was provided by Dr. M. E. Daub (North Carolina State University, NC, USA).

2.5.2. Preparation of Expression Constructs

DNA manipulations were carried out in accordance with standard methods (Appendix B).[24] Details of the expression constructs used in this study are presented in Table C.3 (Appendix C). Primers used for cloning are presented in Table C.1 (Appendix

C). All plasmids were maintained in *E. coli* BL21(DE3) cells stored in 20% glycerol at −80 °C. The SAT-KS-MAT and PT proteins had a C-terminal $6 \times$ His tag. The ACP_2 and TE proteins had an N-terminal $6 \times$ His tag. All plasmids were screened for proper assembly by restriction enzyme digest and the complete insert was confirmed by DNA sequencing. All sequencing was conducted at the Sequencing and Synthesis Core Facility, Johns Hopkins University, School of Medicine (Baltimore, MD, USA).

The SAT-KS-MAT tridomain expression construct (pECTB1-NKA6) was generated from three separate DNA fragments that were spliced together using overlap extension PCR (Appendix B). The first fragment (containing the 3′ end of exon 3) was amplified by PCR (Appendix B) from genomic *C. nicotianae* DNA template with primers CTB1-ex3-5 and CTB1-ex3-3. The second fragment was a codon-optimized synthon (Appendix C) spanning exons 4 and 5 constructed through polymerase cycling assembly (PCA, Appendix B). The primers used in the construction of this synthon are described in Table C.2 (Appendix C). The third fragment (containing the 5′ end of exon 6) was amplified by PCR from genomic *C. nicotianae* DNA template with primers CTB1-ex6-5 and CTB1-MAT6-3. Overlap extension PCR of these three fragments with outside primers CTB1-ex3-5 and CTB1-MAT6-3 was used to generate the SAT-KS-MAT insert. This product was digested with *HindIII* and *NotI* and inserted into analogous sites of pECTB1-SAT[8] to generate pECTB1-NKA6.

The PT monodomain expression construct (pECTB1-PT) was made by overlap extension PCR of three DNA fragments. The first fragment was amplified by PCR from genomic *C. nicotianae* DNA template with primers CTB1-PT-5 and CTB1-ex6-3. The second fragment was amplified by PCR from genomic *C. nicotianae* DNA template with

primers CTB1-ex7-5 and CTB1-ex7-3. The third fragment was amplified by PCR from genomic *C. nicotianae* DNA template with primers CTB1-ex8-5 and CTB1-PT-3. Overlap extension PCR was carried out with these three fragments with primers CTB1-PT-5 and CTB1-PT-3 to generate the PT insert. The resulting fragment was digested with *NdeI* and *NotI* and inserted into corresponding restriction sites in pET-24a(+) to make pECTB1-PT.

The ACP$_2$ monodomain expression construct (pECTB1-ACP) was made by PCR using *C. nicotianae* genomic DNA template with primers CTB1-ACP-5 and CTB1-ACP-3s. The amplified product was digested with *NdeI* and *NotI* and inserted into the corresponding restriction sites in pET-28a(+) to generate pECTB1-ACP.

An ACP$_2$-TE didomain expression construct (pECTB1-TA) was made through overlap extension PCR of two DNA fragments. The first fragment was amplified by PCR from *C. nicotianae* genomic DNA with primers CTB1-ACP-5 and CTB1-ex8-3. The second fragment was amplified by PCR from *C. nicotianae* genomic DNA with primers CTB1-ex9-5 and CTB1-3. Overlap extension PCR of these two fragments with primers CTB1-ACP-5 and CTB1-3 generated the ACP$_2$-TE insert. This amplified product was digested with *NdeI* and *NotI* and inserted into the corresponding restriction sites in pET-24a(+) to generate pECTB1-TA.

The TE monodomain expression construct (pECTB1-TE) was created through PCR amplification of pECTB1-TA template with primers CTB1-TE-5 and CTB1-3-Stop. The amplified product was digested with *NdeI* and *NotI* and inserted into the corresponding restriction sites in pET-28a(+) to generate pECTB1-TE.

53

The TE point-mutant expression constructs (pECTB1-TE-S2008A and pECTB1-TE-H2171Q) were generated by overlap extension PCR. The 5′ fragments were amplified by PCR using pECTB1-TE template with primers CTB1-TE-5 and CTB1-S2008A-3 or CTB1-H2171Q-3, respectively. The 3′ fragments were amplified by PCR using pECTB1-TE template with primers CTB1-S2008A-5 or CTB1-H2171Q-5, respectively, and CTB1-3-Stop. Overlap extension PCR of these two fragments with primers CTB1-TE-5 and CTB1-3-Stop generated the point-mutant insert. These amplified products were digested with *NdeI* and *NotI* and inserted into the corresponding sites of pET-28a(+) to generate pECTB1-TE-S2008A and pECTB1-TE-H2171Q, respectively.

2.5.3. *Protein Expression and Purification*

Expression constructs were transformed into *E. coli* BL21(DE3) by electroporation (Appendix B). Proteins were prepared by heterologous expression and purified by nickel affinity chromatography (Appendix B). Eluted protein fractions were dialyzed into 100 mM potassium phosphate pH 7.0, 5% glycerol at 4 °C overnight with three buffer exchanges. Purified protein concentrations were determined by the Bradford assay using bovine serum albumin (New England Biolabs) as a standard, as described in Appendix B.[25] Excess protein was flash frozen in liquid nitrogen and stored at −80 °C until further use.

2.5.4. *[^{18}O]Acetyl-CoA Synthesis and Purification*

[^{18}O]Acetyl-CoA was prepared enzymatically as follows: 5 mM CoASH, 10 mM sodium acetate (^{18}O$_2$, 95%), 10 mM ATP, 1 mM TCEP, and 0.4 unit/mL acetyl-CoA synthetase in 25 mM potassium phosphate pH 7.0 at 25 °C for 2 h. Protein was removed by filtration through an Amicon Ultra centrifugal filter unit with a 10 kDa nominal

molecular weight limit (EMD Millipore). [18O]Acetyl-CoA was purified by HPLC on an Agilent 1200 (Agilent Technologies, Santa Clara, CA, USA) system fitted with a Prodigy 5u ODS3 100Å (250 × 4.60 mm, 5 μm) column (Phenomenex, Torrance, CA, USA) using the following mobile phase gradient: hold at 5% solvent A, 5 min; linear gradient from 5% solvent A to 35% solvent A, 20 min; at a flow rate of 1 mL/min; where solvent A was acetonitrile + 0.1% TFA and solvent B was water + 0.1% TFA. [18O]Acetyl-CoA eluted at 12.2 min. Purified [18O]acetyl-CoA was lyophilized, dissolved in 100 mM potassium phosphate pH 7.0 to yield a 25 mM solution, and stored at −80 °C. High-resolution mass analysis to determine heavy-isotope incorporation was conducted on a Shimadzu LCMS-IT-TOF (Columbia, MD, USA).

2.5.5. In Vitro ACP Phosphopantetheinylation

The ACP fragment needed to be activated with 4′-phosphopantetheine prior to reconstitution reactions. This was accomplished *in vitro* using Svp, a promiscuous phosphopantetheinyl transferease as previously described.[11a] Briefly, the ACP activation reaction was as follows: 100 μM CTB1-ACP$_2$, 2 μM Svp, 400 μM CoASH, and 20 mM MgCl$_2$ in 100 mM potassium phosphate pH 7.0, 5% glycerol at 25 °C for 1 h.

2.5.6. In Vitro Reconstitution Reactions

Proteins used in the reconstitution reactions were purified immediately before use. Reaction conditions were as follows: 10 μM each of the CTB1 protein fragments to be used, 0.5 mM acetyl-CoA, 2 mM malonyl-CoA, 1 mM TCEP in 100 mM potassium phosphate, 5% glycerol, pH 7.0. The CoA substrates were added simultaneously to initiate reactions. Reactions were conducted at room temperature for 4 h. To demonstrate

the starter unit effect, acetyl-CoA was excluded. To test the mechanism of pyrone formation, [^{18}O]acetyl-CoA was used instead of the unlabeled acetyl-CoA.

Reactions were quenched with the addition of concentrated HCl, and the products were extracted thrice with 2 × reaction volume with ethyl acetate. The organic fractions were combined, evaporated under vacuum, and dissolved in 25% acetonitrile (*aq.*) to a final volume equivalent to the reaction volume. The extracts were filtered through 0.2 μm PTFE and analyzed by HPLC and LCMS-IT-TOF.

2.5.7. *Analytical Methods*

HPLC was carried out on an Agilent 1200 fitted with a Prodigy 5u ODS3 100Å (250 × 4.60 mm, 5 μm) column and UV-vis diode array detector (DAD) using the following mobile phase gradient: 50 μL injections; into 5–85% solvent A, 40 min; at 1 mL/min; where solvent A was acetonitrile + 0.1% formic acid and solvent B was water + 0.1% formic acid. Polyketide products were detected by monitoring at 280 nm and the full UV-vis spectrum was collected for each product.

LCMS analysis was carried out on a Shimadzu LCMS-IT-TOF fitted with a Luna C18(2) 3μ (150 × 2.0 mm, 3 μm) column (Phenomenex). The mobile phase gradient for LC was described above, except with a flow rate of 0.2 mL/min. Mass detection was carried out in the positive ion mode. We thank Dr. K. L. Fiedler and Prof. R. J. Cotter of the Middle Atlantic Mass Spectrometry Laboratory at the Johns Hopkins School of Medicine (Baltimore, MD, USA) and the Mid-Atlantic Regional Office of Shimadzu Scientific Instruments, Inc. (Columbia, MD, USA) for use of their Shimadzu LC-IT-TOF.

2.6. References

[1]   (a) A. Okubo, S. Yamazaki, K. Fuwa, *Agric Biol Chem* **1975**, *39*, 1173-1175; (b)
        M. E. Daub, *Phytopathology* **1982**, *72*, 370-374; (c) M. E. Daub, S. Herrero, K. R.
        Chung, *FEMS Microbiol Lett* **2005**, *252*, 197-206.
[2]   M. E. Daub, R. P. Hangarter, *Plant Physiol* **1983**, *73*, 855-857.
[3]   (a) M. E. Daub, *Plant Physiol* **1982**, *69*, 1361-1364; (b) M. E. Daub, S. P. Briggs,
        Plant Physiol **1983**, *71*, 763-766; (c) M. E. Daub, M. Ehrenshaft, *Annu Rev
        Phytopathol* **2000**, *38*, 461-490.
[4]   C. A. Mulrooney, B. J. Morgan, X. Li, M. C. Kozlowski, *J Org Chem* **2010**, *75*,
        16-29.
[5]   (a) M. Choquer, K. L. Dekkers, H. Q. Chen, L. H. Cao, P. P. Ueng, M. E. Daub,
        K. R. Chung, *Mol Plant Microbe Interact* **2005**, *18*, 468-476; (b) H. Q. Chen, M.
        H. Lee, M. E. Daub, K. R. Chung, *Mol Microbiol* **2007**, *64*, 755-770.
[6]   (a) K. L. Dekkers, B. J. You, V. S. Gowda, H. L. Liao, M. H. Lee, H. H. Bau, P.
        P. Ueng, K. R. Chung, *Fungal Genet Biol* **2007**, *44*, 444-454; (b) H. Q. Chen, M.
        H. Lee, K. R. Chung, *Microbiology* **2007**, *153*, 2781-2790; (c) C. Staerkel, M. J.
        Boenisch, C. Kroger, J. Bormann, W. Schafer, D. Stahl, *BMC Plant Biology* **2013**,
        13, 50.
[7]   M. Choquer, M. H. Lee, H. J. Bau, K. R. Chung, *FEBS Lett* **2007**, *581*, 489-494.
[8]   J. M. Crawford, A. L. Vagstad, K. R. Whitworth, K. C. Ehrlich, C. A. Townsend,
        Chembiochem **2008**, *9*, 1019-1023.
[9]   L. C. Du, L. L. Lou, *Nat Prod Rep* **2010**, *27*, 255-278.
[10]   A. L. Vagstad, S. B. Bumpus, K. Belecki, N. L. Kelleher, C. A. Townsend, *J Am
        Chem Soc* **2012**, *134*, 6865-6877.
[11]   (a) J. M. Crawford, P. M. Thomas, J. R. Scheerer, A. L. Vagstad, N. L. Kelleher,
        C. A. Townsend, *Science* **2008**, *320*, 243-246; (b) T. P. Korman, J. M. Crawford,
        J. W. Labonte, A. G. Newman, J. Wong, C. A. Townsend, S. C. Tsai, *Proc Natl
        Acad Sci USA* **2010**, *107*, 6246-6251.
[12]   A. L. Vagstad, E. A. Hill, J. W. Labonte, C. A. Townsend, *Chem Biol* **2012**, *19*,
        1525-1534.
[13]   (a) D. W. Udwary, M. Merski, C. A. Townsend, *J Mol Biol* **2002**, *323*, 585-598;
        (b) D. T. Jones, *J Mol Biol* **1999**, *292*, 195-202; (c) D. W. A. Buchan, S. M.
        Ward, A. E. Lobley, T. C. O. Nugent, K. Bryson, D. T. Jones, *Nucleic Acids Res*
        2010, *38*, W563-W568.
[14]   I. Fujii, A. Watanabe, U. Sankawa, Y. Ebizuka, *Chem Biol* **2001**, *8*, 189-197.
[15]   H. Ogawa, K. Hasumi, K. Sakai, S. Murakawa, A. Endo, *J Antibiot* **1991**, *44*, 762-
        767.
[16]   (a) M. Yalpani, K. Willecke, F. Lynen, *Eur J Biochem* **1969**, *8*, 495-&; (b) K. R.
        Srinivasan, S. Kumar, *Biochemistry* **1981**, *20*, 3400-3404.
[17]   J. M. Crawford, C. A. Townsend, *Nature Rev Microbiol* **2010**, *8*, 879-889.
[18]   J. M. Crawford, T. P. Korman, J. W. Labonte, A. L. Vagstad, E. A. Hill, O.
        Kamari-Bidkorpeh, S. C. Tsai, C. A. Townsend, *Nature* **2009**, *461*, 1139-U1243.
[19]   (a) A. C. Ruzzini, S. Ghosh, G. P. Horsman, L. J. Foster, J. T. Bolin, L. D. Eltis, *J
        Am Chem Soc* **2012**, *134*, 4615-4624; (b) A. C. Ruzzini, G. P. Horsman, L. D.
        Eltis, *Biochemistry* **2012**, *51*, 5831-5840.

[20]   J. M. Crawford, A. L. Vagstad, K. C. Ehrlich, C. A. Townsend, *Bioorg Chem* **2008**, *36*, 16-22.

[21]   (a) J. M. Crawford, B. C. R. Dancy, E. A. Hill, D. W. Udwary, C. A. Townsend, *Proc Natl Acad Sci USA* **2006**, *103*, 16728-16733; (b) Z. Z. Gao, J. J. Wang, A. K. Norquay, K. J. Qiao, Y. Tang, J. C. Vederas, *J Am Chem Soc* **2013**, *135*, 1735-1738; (c) H. Zhou, K. Qiao, Z. Gao, J. C. Vederas, Y. Tang, *J Biol Chem* **2010**, *285*, 41412-41421; (d) H. Zhou, K. Qiao, Z. Gao, M. J. Meehan, J. W. Li, X. Zhao, P. C. Dorrestein, J. C. Vederas, Y. Tang, *J Am Chem Soc* **2010**, *132*, 4530-4531.

[22]   (a) A. Watanabe, Y. Ebizuka, *Tetrahedron Lett* **2002**, *43*, 843-846; (b) J. Foulke-Abel, C. A. Townsend, *Chembiochem* **2012**, *13*, 1880-1884; (c) Y. Ma, L. H. Smith, R. J. Cox, P. Beltran-Alvarez, C. J. Arthur, T. J. Simpson, *Chembiochem* **2006**, *7*, 1951-1958.

[23]   W. J. Dower, J. F. Miller, C. W. Ragsdale, *Nucleic Acids Res* **1988**, *16*, 6127-6145.

[24]   J. Sambrook, D. W. Russell, *Molecular cloning: A laboratory manual*, Cold Spring Harbor Laboratory Press, Plainview, NY, **2001**.

[25]   M. M. Bradford, *Anal Biochem* **1976**, *72*, 248-254.

Chapter 3: Systematic Domain-Swaps of Iterative, Non-reducing Polyketide Synthases Provide a Mechanistic Understanding and Rationale for Catalytic Reprogramming

This chapter was adapted with permission from A. G. Newman, A. L. Vagstad, P. A. Storm, C. A. Townsend, *J. Am. Chem. Soc.* **2014**, *136*, 7348-7362.

3.1.   Introduction

Polyketides are a structurally and functionally diverse family of natural products containing environmental toxins and pigments as well as pharmaceutical agents. Several drugs on the market are of polyketide origin including the antibiotics tetracycline and erythromycin, immunosuppressant rapamycin, anticholesterol drug lovastatin, and anticancer drug epothilone B.[1] The direct engineering of polyketide biosynthetic pathways for the production of non-native metabolites has been an attractive goal for those wishing to expand the potential polyketide drug pool. However, such approaches require a greater and nuanced understanding of the underlying mechanisms at play in polyketide assembly. As a consequence, the methodology of rational redirection of polyketide pathways is still in its infancy.[2]

Many of the known polyketides are assembled in a linear fashion by large, multidomain proteins dubbed type I polyketide synthases (PKS).[3] While the modular type I PKSs have been extensively studied for 20 years, the biochemistry of fungal, iterative, non-reducing PKSs (NR-PKS) is only now beginning to be elucidated.[4] The NR-PKSs are responsible for the biosynthesis of a variety of aromatic polyketide products and share a common domain architecture that is intrinsically linked to their function (Figure 3.1). The mode of biosynthesis is analogous to that of fatty acids by animal fatty acid synthases (FAS), but simplified.[5] The three amino-terminal domains,
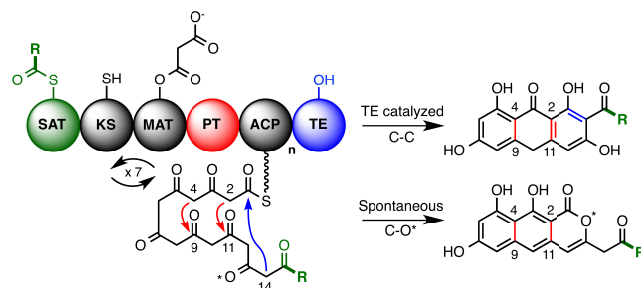
Figure 3.1 | The core domain architecture of NR-PKSs. Enzyme bound intermediates and products of TE-directed Claisen cyclization or spontaneous O–C bond closure are highlighted.

the starter unit:acyl-carrier protein transacylase (SAT), ketosynthase (KS) and malonyl acyl transferase (MAT) domains are responsible for the initiation and polyketide elongation phases. The SAT domain selects a precursor or starter unit substrate as an acyl thioester while the MAT domain introduces ketide extender units from malonyl-CoA. The KS works in collaboration with the acyl-carrier protein (ACP) to catalyze the decarboxylative Claisen condensation of these substrates generating a linear, ACP-bound poly-β-ketone intermediate. The carboxy-terminal domains of NR-PKSs control the final stage of biosynthesis, which includes regiospecific aldol cyclizations/aromatizations by the product template (PT) domain and release through either hydrolysis,[6] Claisen (Dieckmann) cyclization,[7] or pyrone formation[8] by the TE. In this way, the four factors governing chemical diversity in aromatic polyketides are entirely controlled by the enzyme, with the amino-terminal half determining starter unit selection and chain length, and the carboxy-terminal half controlling regiospecific cyclization/aromatization and mechanism of product release.

Protein deconstruction, in which the NR-PKS is dissected into smaller mono- to multidomain fragments, has been a crucial tool for mechanistic understanding of these enzymes.[7a, 9] The method enables the rapid and selective *in vitro* recombination of NR-
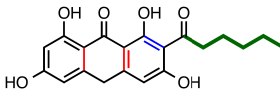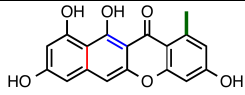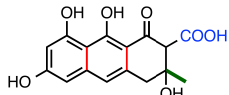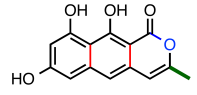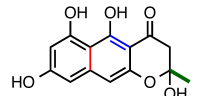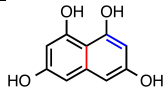
PKS activity, allowing for individual domain functions to be deduced and assayed. In these reactions, the enzymes work *in trans*, mimicking the discrete type II PKSs. The approach also allows for the recombination of an unnatural NR-PKS through domain-swapping from functionally analogous domains of different NR-PKSs. Genetic domain-swapping and the creation of chimeric modular, type I PKSs has been extensively examined *in vivo*, primarily with the DEBS PKSs of erythromycin biosynthesis.[10] These systems have been uniquely attractive because the domains of these canonical modular enzymes are only used once in an assembly line fashion. Additionally, the KS domains often demonstrate sufficient substrate promiscuity to allow for a degree of synthetic flexibility. Similarly, combinatorial studies of discrete type II PKSs have led to the generation of aromatic metabolite libraries.[11] Such approaches have focused on site-specific reductions and cyclizations catalyzed by ketoreductase (KR) and aromatase/cyclase (ARO/CYC) proteins accepting alternative chain-length polyketide intermediates.[12]

Recently, our laboratory has explored *in vitro* domain-swapping as a method of developing non-cognate NR-PKSs.[13] We postulated that swapping *N*- and C-terminal halves of NR-PKSs could lead to the generation of novel products. In this way starter unit selection and chain-length control would be decoupled from the tailoring steps of late stage polyketide biosynthesis. Using the N-terminal SAT-KS-MAT tridomain fragment from the cercosporin biosynthetic NR-PKS, CTB1,[8, 14] we were able to efficiently complement enzymatic activity with the C-terminal half of four non-cognate NR-PKSs, representing a variety of cyclization and release pathways (Appendix D, Figure D.2). These limited data codified several proposed "rules" for efficient redirection of

biosynthesis, which are further tested and validated in experiments described herein. While the MAT and ACP domains behave consistently regardless of the parent synthase, the KS domain exerts stringent chain-length control, as has been demonstrated in multiple systems. The PT domain catalyzes cyclization using the bound phosphopantetheine thioester as the benchmark from which to establish regiochemistry. In this way, aromatization always occurs with the proper "register" (*i.e.* C2–C7, C4–C9, or C6–C11) for the given PT, even if it accepts a linear polyketide intermediate of non-native length. This observation is in keeping with the complementation of cyclization register observed by Tang and coworkers with respect to the AptA and VrtA PTs.[15] Finally, the TE domain is crucial for catalytic turnover and will only exert its effect if a compatible intermediate is generated.

With the success of *in vitro* C-terminal domain-swapping with the CTB1 SAT-KS-MAT, we wanted to systematically investigate the potential of this approach. We elected to study six enzymes that represent the wide span of known chemical diversity controlled by NR-PKSs: alkyl starter unit selection (acetyl or hexanoyl), chain-length control (C_{12} to C_{20}), PT cyclization modes (C2–C7, C4–C9, or C6–C11), and TE release mechanism (hydrolysis, pyrone formation, or Claisen cyclization). The native activity of each enzyme investigated has been determined: *Aspergillus parasiticus* PksA, *Giberella fujikuroi* Pks4, *Aspergillus terreus* ACAS, *Cercospora nicotianae* CTB1, *Aspergillus nidulans* wA, and *Colletotichum lagenaria* Pks1 each produce norsolorinic acid anthrone (**1,** noranthrone),[7a] pre-bikaverin (**2**),[16] atrochrysone carboxylic acid (**3**),[17] *nor*-toralactone (**4**),[8] YWA1 (**5**),[18] and 1,3,6,8-tetrahydroxynaphthalene (THN, **6**),[19] respectively (Table 3.1).

Table 3.1 | Native products of NR-PKSs used in this study.

Protein	SAT Starter Unit Selection	Chain Length	PT Cyclization Register	TE Release Mechanism	Product Formed
PksA	Hexanoyl	C_{20}	C4–C9 C2–C11	C14–C1	noranthrone, **1**
Pks4	Acetyl	C_{18}	C2–C7	C10–C1	pre-bikaverin, **2**
ACAS	Acetyl	C_{16}	C6–C11 C4–C13	Hydrolysis[a]	atrochrysone carboxylic acid, **3**
CTB1	Acetyl	C_{14}	C4–C9 C2–C11	O13–C1	*nor*-toralactone, **4**
wA	Acetyl	C_{14}	C4–C9 C2–C11	C10–C1	YWA1, **5**
Pks1	Acetyl	C_{12}	C2–C7	C10–C1[b]	THN, **6**

[a] The TE of the ACAS system exists as a discreet β-lactamase type TE, ACTE.
[b] The Pks1 TE has an additional deacetylase activity.

3.2. Results

3.2.1. Deconstruction of NR-PKSs

The dissection of NR-PKSs into component mono- to multidomain fragments was patterned on the previously reported deconstruction of PksA.[7a] The percent similarity between the NR-PKSs used in this study ranged from 43% to 61% as calculated from global pairwise protein sequence alignments (Table 3.2). Equivalent cut sites for the other NR-PKSs used in this study were selected based on primary sequence alignments. Alternative cut sites were selected to optimize yields for proteins that suffered from low expression, insolubility or instability. SAT-KS-MAT tridomain, PT, ACP_n, and TE

Table 3.2│Percent identity/similarity/gaps for global pairwise alignment of NR-PKSs. Alignment conducted using EMBOSS Needle with default settings (http://www.ebi.ac.uk/Tools/psa/).

	PksA	ACAS	Pks4	CTB1	wA
ACAS	30.3/45.9/23.1				
Pks4	36.9/55.6/8.9	32.8/48.6/20.3			
CTB1	36.8/53.8/11.5	29.4/42.9/25.5	36.0/52.4/11.9		
wA	38.5/54.8/12.2	32.3/56.7/26.0	43.0/60.1/10.5	37.5/54.9/8.8	
Pks1	35.8/52.2/11.8	33.7/47.1/25.3	38.2/53.6/11.2	35.3/53.3/7.9	43.3/60.8/6.2

monodomain fragments were cloned as hexahistidine fusions for heterologous expression in *E. coli* followed by nickel affinity purification (Appendix D, Figure D.1). The tandem ACPs from CTB1, wA, and Pks1 were prepared as intact didomain fragments, here referred to as ACP$_2$. The discrete β-lactamase type TE of atrochrysone biosynthesis, ACTE,[17] was expressed separately. SAT-KS-MAT tridomain fragments for ACAS and wA could not be obtained despite repeated attempts to optimize cut-site selection, fusion tag identity and location, and culture conditions. Details of individual protein constructs used in this study are presented in Table D.3 (Appendix D). All constructs were sequence verified.

3.2.2. Non-cognate Minimal NR-PKS Compatibility

Given the successful redirection of biosynthesis in combinatorial reactions of CTB1,[13] we sought to systematically evaluate if non-cognate NR-PKS domains, in general, could predictably alter biochemistry. Combinatorial reactions were modeled on the *in vitro* reconstitution of PksA, using 10 μM protein concentrations and acyl-*N*-acetylcysteamine (SNAC) thioester substrates.[7a] Extracted products from these reactions were analyzed by reverse phase HPLC and LC-ESI-MS. Product identification was accomplished by comparing UV-visible spectra and masses to those of authentic standards or literature values. As in the previous study,[13] selected sets of PT, ACP$_n$, and
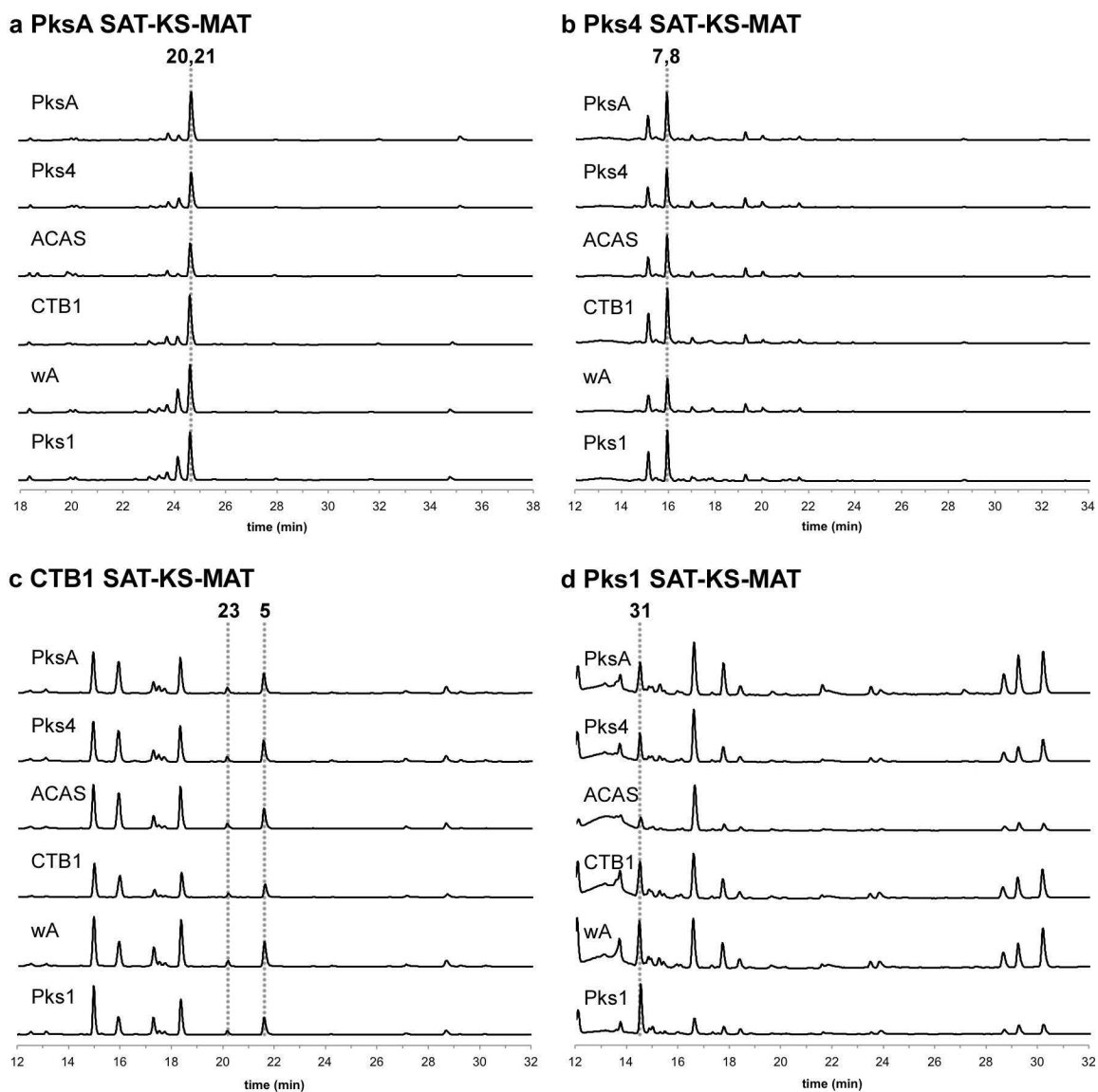
Figure 3.2 | Product profiles for combinatorial minimal NR-PKSs. The 280 nm chromatograms for reactions of PksA SAT-KS-MAT (**a**), Pks4 SAT-KS-MAT (**b**), CTB1 SAT-KS-MAT (**c**), and Pks1 SAT-KS-MAT (**d**) with the indicated ACP_n are presented.

TE monodomains from a single parent synthase were recombined with a non-cognate SAT-KS-MAT tridomain. By systematically varying the SAT-KS-MAT tridomain, we were able to achieve 72 individual combinations, representing the complete set of two-part non-cognate configurations.

We first established that all minimal NR-PKSs were active, regardless of the identity of the ACP_n domain. We define a minimal NR-PKS as the requisite set of

domains for proficient ketide homologation accomplished in the initial stages of polyketide biosynthesis. In practice, the minimal NR-PKS consists of the SAT-KS-MAT and ACP$_n$ domains. The combination of SAT-KS-MAT with non-cognate ACP$_n$ gave nearly identical product profiles in similar yields to that of the cognate reaction in all cases (Figure 3.2). It is noteworthy that the native ACP monodomains (PksA, Pks4 and ACAS) were able to efficiently complement activity with SAT-KS-MAT tridomains from proteins with native tandem ACPs (CTB1 and Pks1). It has been established that in wild-type wA both ACPs are independently functional.[18] These data further strengthen the view that independent functionality is generally true of tandem ACPs. The role of multiple ACPs may then be to improve biosynthetic efficiency, as has been shown in polyunsaturated fatty acid synthases, which contain up to six tandem ACPs.[20] The complete compatibility of ACPs indicates that the recognition motifs for this domain are highly conserved among the NR-PKSs used in the current study. It is expected that ACP compatibility is a universal feature of NR-PKSs and an essential precondition of the planned domain-swapping experiment.

Overall, the products of the minimal NR-PKSs were consistent, by mass, with shunt products of full-length poly-β-ketones produced by the respective SAT-KS-MAT protein, with one exception. The major products of the minimal Pks4 are a pair of octaketide (C_{16}) truncations coincident with the actinorhodin type II PKS products, SEK4
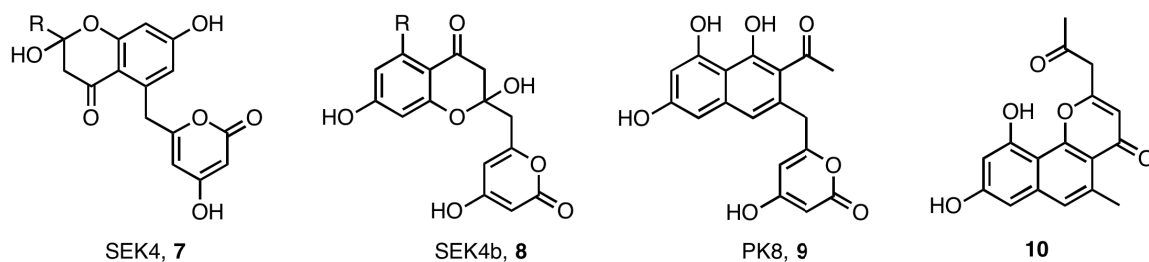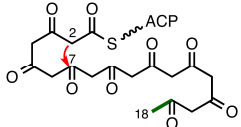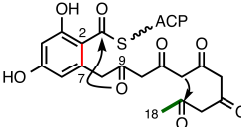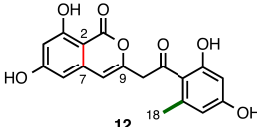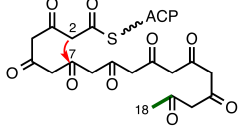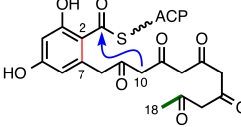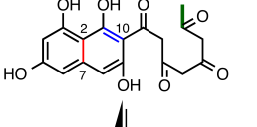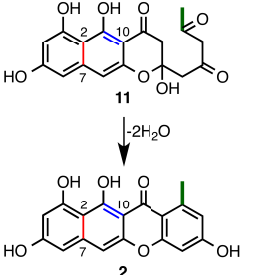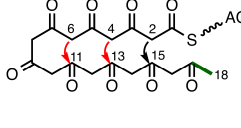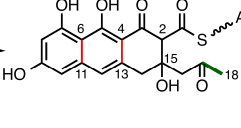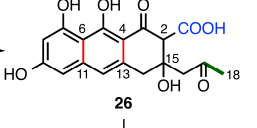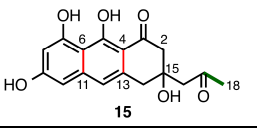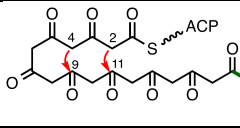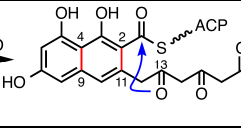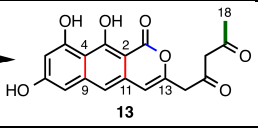
Figure 3.3 │ Observed and predicted derailment products of minimal Pks4.

(**7**) and SEK4b (**8**, Figure 3.3).[21] Pks4 is expected to yield C_{18}-chain-length, nonaketide

products (**9, 10**, Figure 3.3), as has been observed in the reactions of 50 μM Pks4 KS-

MAT and 10 μM Pks4 ACP.[22] It should be noted that in the current study, the

concentration of Pks4 SAT-KS-MAT is 10 μM, indicating that enzyme concentration

might be a factor influencing chain length for deconstructed synthases. It is likely that

Pks4 exhibits reduced processivity as the growing linear intermediate approaches the full

chain length. The reduction in processivity is in agreement with a similar effect in PksA,

where truncated heptaketide shunt products were observed *in vitro*.[23] This effect might

be an artifact of the enzyme deconstruction approach. If the Pks4 KS domain has a

decreased affinity for the C_{16}-linear poly-β-ketone intermediate, the KS and ACP

domains could freely dissociate in the deconstructed system allowing for derailing

cyclizations and off-loading. For the intact system and perhaps for deconstructed

enzymes at higher concentrations, dissociation is less likely.
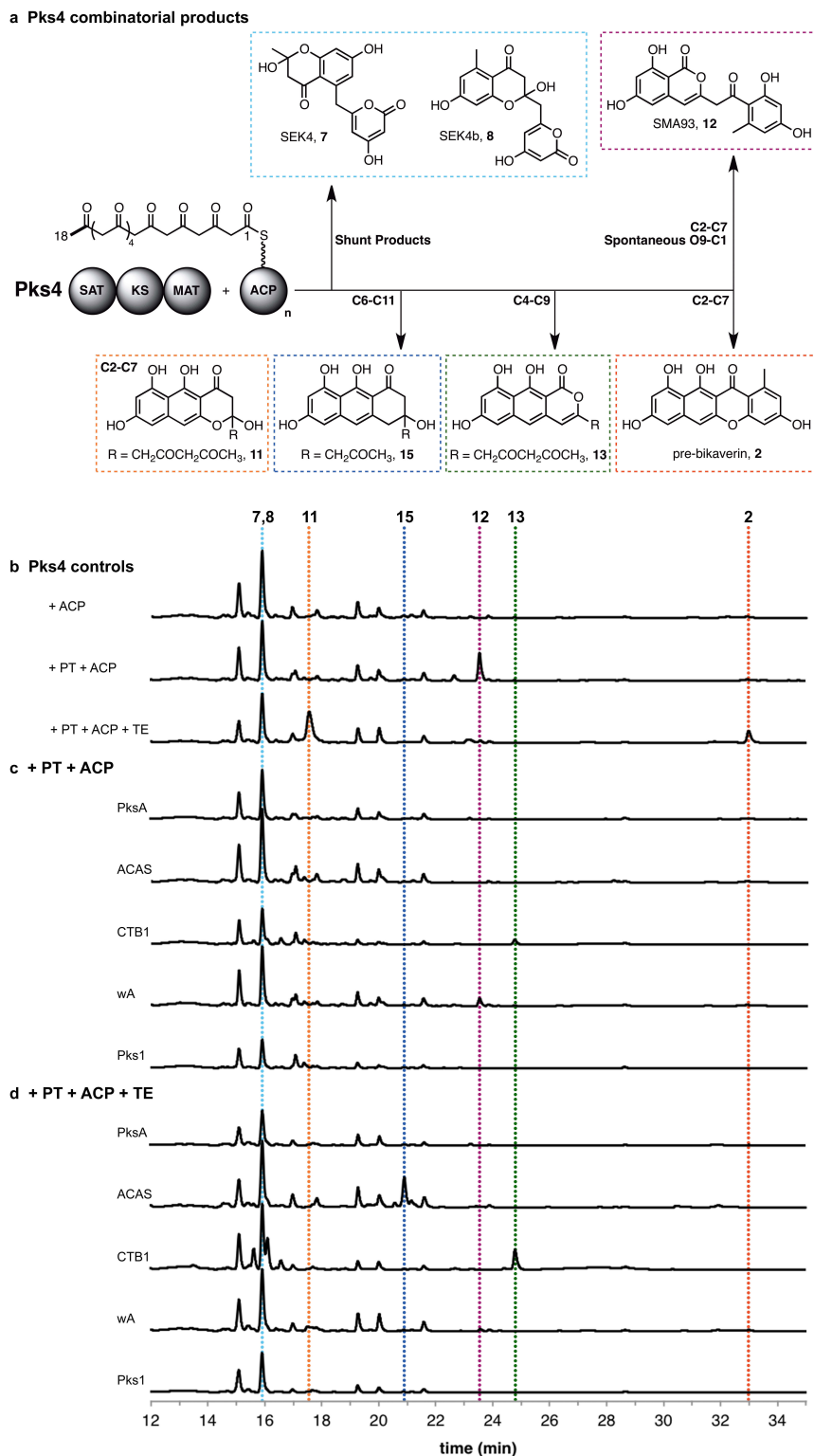
Figure 3.4│Product analysis of combinatorial reactions with Pks4 SAT-KS-MAT. Proposed products of combinatorial reactions are presented (**a**). The chromatograms at 280 nm for products of cognate control reactions (**b**), non-cognate combinatorial reactions with the indicated PT and ACP$_n$ (**c**), and non-cognate combinatorial reactions with the indicated PT, ACP$_n$, and TE (**d**) are presented.

Table 3.3 │ Intermediates and products produced by combinatorial reactions with Pks4 SAT-KS-MAT.

SAT-KS-MAT	PT + ACP	TE	Linear Poly-β-ketone	PT-Cyclized Intermediate	Released Product
Pks4	Pks4		*(structure)*	*(structure)*	**12**
Pks4	Pks4	Pks4	*(structure)*	*(structure)*	**11**, **2**
Pks4	ACAS	ACTE	*(structure)*	*(structure)*	**26**, **15**
Pks4	CTB1	CTB1	*(structure)*	*(structure)*	**13**

3.2.3. *Combinatorial Reactions of Pks4 SAT-KS-MAT*

HPLC product profiles for combinatorial reactions utilizing Pks4 SAT-KS-MAT are presented in Figure 3.4. Reactions for successful combinations are delineated in Table 3.3. As has been reported, the Pks4 PT catalyzes C2–C7 aldol cyclization/aromatization while the Pks4 TE catalyzes C10–C1 Claisen/Dieckmann cyclization on a nonaketide (C_{18}) linear intermediate to generate pre-bikaverin (**2,** Table 3.3).[16] The third ring (O9–
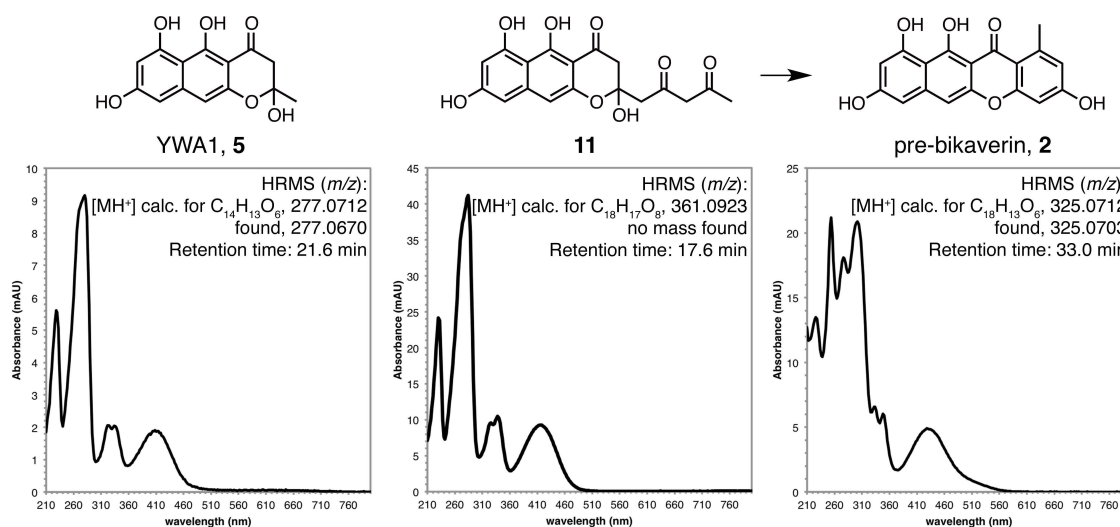
Figure 3.5│UV-vis spectra and HRMS for YWA1-core containing molecules. Product **11** converts to pre-bikaverin.

C13) and fourth ring (C12–C17) are postulated to form spontaneously, followed by two

dehydrations. Indeed, pre-bikaverin is observed upon complete reconstitution of a

cognate Pks4 (Figure 3.4b). Interestingly, an additional product **11** of this reaction was

observed to have a virtually identical UV-vis spectrum to that of YWA1 (**5**) suggesting

that they share the same core architecture (Figures 3.4b and 3.5).[18] The mass of this

product was found to correspond to a formula of $C_{18}H_{16}O_8$ and there is an apparent

complete conversion of it to pre-bikaverin with time. Given these data, we postulate that

this product is the C_{18}-YWA1 analog **11**, in which the methyl substituent bears an

acetoacetyl side chain extension (Table 3.3). This hypothesis would be consistent with

proper PT and TE cyclization followed by hemi-ketal formation prior to

cyclization/aromatization of the fourth ring.

Neither pre-bikaverin nor the C_{18}-YWA1 analog **11** is detected in reactions of

Pks4 lacking the TE. Instead, SMA93 (**12**)[24] is observed (Figure 2B). SMA93 arises

from proper C2–C7 PT-catalyzed cyclization of the C_{18}-linear intermediate followed by
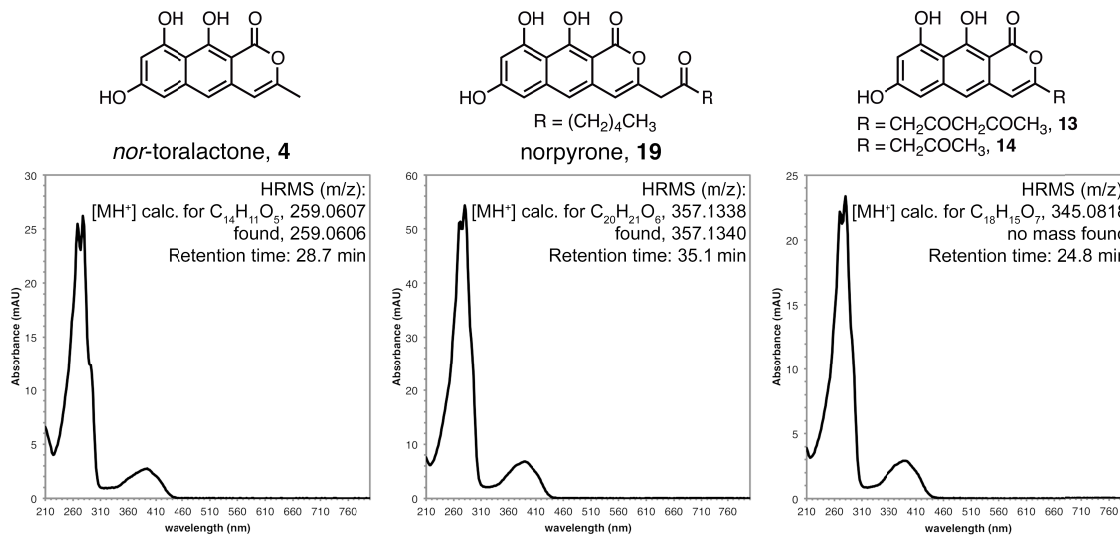
Figure 3.6│ UV-vis spectra and HRMS for naphthopyrone-core containing molecules.

spontaneous O9–C1 closure, a common release mechanism (Table 3.3). In reactions of

Pks4 SAT-KS-MAT with wA PT and ACP$_2$, small quantities of SMA93 were also

identified, indicating that wA PT directed C2–C7 cyclization (Figure 3.4c). However, wA

TE could not complement C10–C1 cyclization, and only vanishingly small amounts of

pre-bikaverin were observed (Figure 3.4d).

On the other hand, combinations with CTB1 (C$_{14}$) PT, ACP$_2$ and TE redirected

biosynthesis towards a product of similar retention time and spectrum to that of *nor*-

toralactone (**4**), the native CTB1 product (Figure 3.4d).[8] Smaller quantities of this

product are observed in reactions lacking the CTB1 TE domain (Figure 3.4c). The

product has the characteristic spectrum of a naphthopyrone and can be compared to both

nor-toralactone and norpyrone (**19,** Figure 3.6).[7a, 8] In keeping with these data, we

propose a C$_{18}$-*nor*-toralactone analog **13** arising from CTB1 PT-catalyzed C4–C9 and

C2–C11 cyclization, followed by CTB1 TE-assisted O13–C1 pyrone formation (Table

3.3). Because a mass for this product could not detected, the C$_{16}$-*nor*-toralactone analog
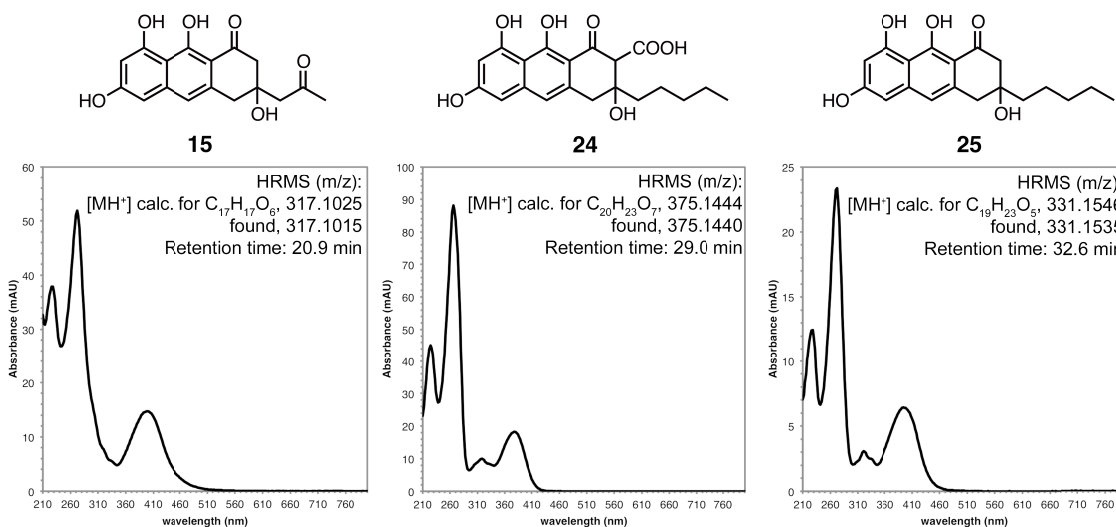
Figure 3.7 │ UV-vis spectra and HRMS for atrochrysone-core containing molecules.

14 (in which the methyl group has been extended with a single acetyl substituent) cannot be ruled out. Two additional unique products were observed in these reactions, eluting at 15.6 minutes and 16.1 minutes (Figure 3.4d). We did not conduct structural characterization of these products, but speculate that they arise from CTB1 PT- and TE-influenced spontaneous cyclization and release.

ACAS (C_{16}) was also able to complement Pks4 SAT-KS-MAT, producing a product with a retention time of 20.9 minutes and a molecular formula of $C_{17}H_{16}O_6$ as determined by LC-ESI-MS (Figure 3.4d). The UV-vis spectrum for this product is in agreement with that for atrochrysone, the decarboxylated native product of ACAS, atrochrysone carboxylic acid (Table 3.1).[17, 25] If the ACAS PT domain catalyzed its native C6–C11 and C4–C13 cyclizations followed by ACTE-catalyzed hydrolytic release on the C_{18} linear poly-β-ketone intermediate, the resulting product, **26**, would contain the atrochrysone core architecture (Table 3.3). We propose the structure **15**, which would arise from decarboxylation of acid **26** and would satisfy both the mass and spectral data

72

(Figure 3.7). A similar decarboxylation is observed in reactions of native ACAS, with atrochrysone carboxylic acid decomposing spontaneously to atrochrysone.[17]

Notably, reactions of Pks4 SAT-KS-MAT with PksA (C_{20}) PT, ACP and TE domains showed no evidence for redirection of chemistry (Figure 3.4d). This observation was unexpected because PksA normally processes a C_{20}-length intermediate and could, therefore, easily accommodate the expected C_{18}-length nonaketide intermediate of Pks4. Instead, reactions with all combinations of PksA domains showed activity similar to that of the minimal Pks4 alone. The inability of the PksA domains to complement activity with the Pks4 intermediate likely arises from the unique hexanoyl starter unit of PksA biosynthesis.[26] The crystal structure of PksA PT shows structurally distinct, hydrophobic binding pocket deep in its active site that presumably binds the hexyl moiety of the PksA linear poly-β-ketone.[27] It is probable that this binding pocket is incompatible with the poly-β-ketone of acetyl-initiated polyketides, nullifying binding and cyclization by the PksA PT.
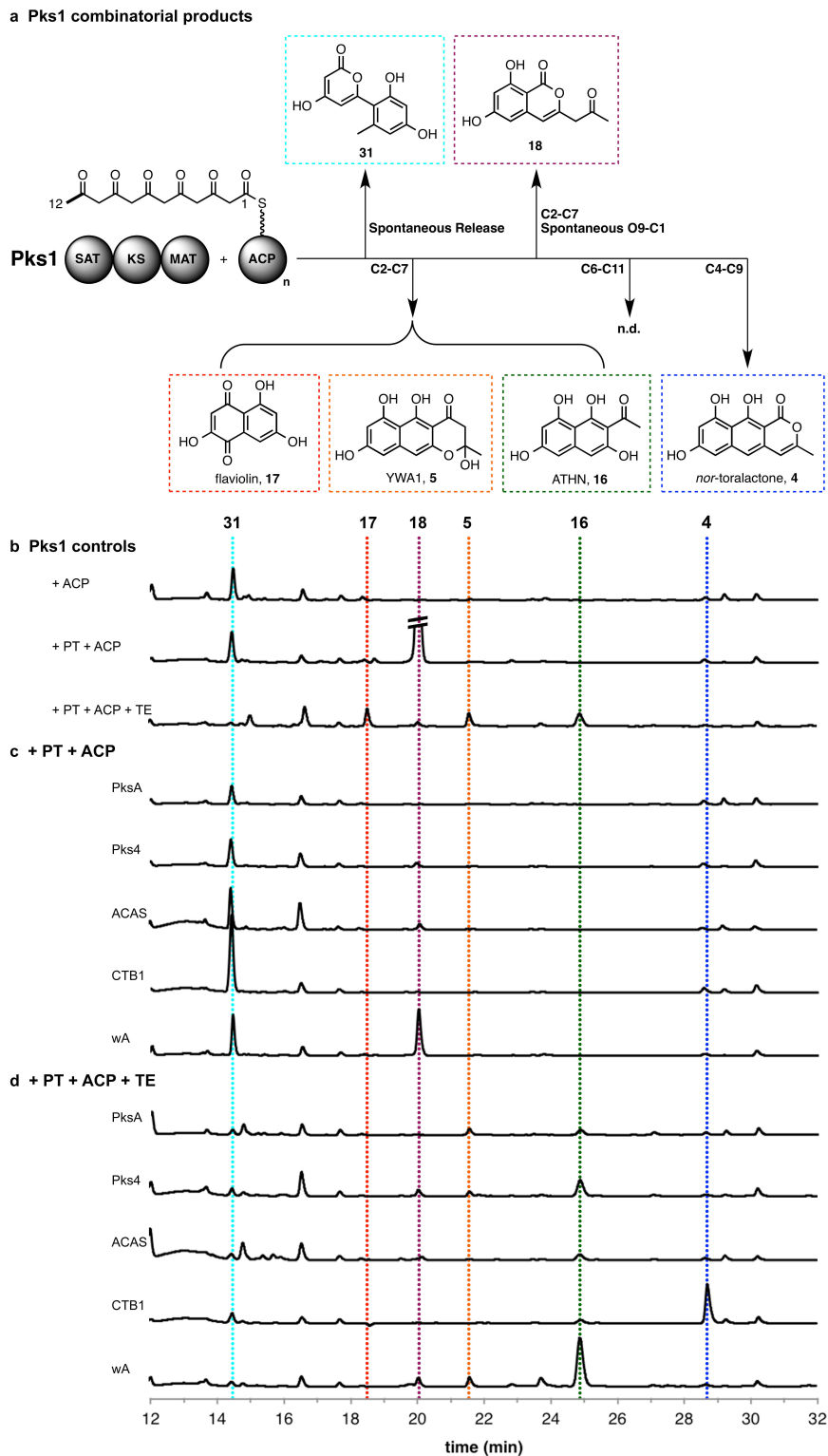
Figure 3.8│ Product analysis of combinatorial reactions with Pks1 SAT-KS-MAT. Proposed products of combinatorial reactions are presented (**a**). The chromatograms at 280 nm for products of cognate control reactions (**b**), non-cognate combinatorial reactions with the indicated PT and ACP_n (**c**), and non-cognate combinatorial reactions with the indicated PT, ACP_n, and TE (**d**) are presented.

Table 3.4│Intermediates and products produced by combinatorial reactions with Pks1 SAT-KS-MAT.

SAT-KS-MAT	PT + ACP	TE	Linear Poly-β-ketone	PT-Cyclized Intermediate	Released Product
Pks1	Pks1 or wA				**18**
Pks1	Pks1	Pks1			**6** → [O] → **17**
Pks1	Pks4 or wA	Pks4 or wA			**16**
Pks1	CTB1	CTB1			**4**

3.2.4. Combinatorial Reactions of Pks1 SAT-KS-MAT

HPLC product profiles for combinatorial reactions utilizing Pks1 SAT-KS-MAT are presented in Figure 3.8. Reactions for successful combinations are delineated in Table 3.4. The native product of Pks1 is 1,3,6,8-tetrahydroxynaphthalene (THN, **6**), an apparent pentaketide (C_{10}) product.[19] However, as has been previously demonstrated, Pks1 in fact catalyzes the formation of a hexaketide (C_{12}) intermediate from the condensation of an acetyl starter unit with five malonyl equivalents. The Pks1 PT domain catalyzes C2–C7 cyclization with the TE domain catalyzing the formation of the second ring through C10–C1 Claisen condensation cyclization, generating 2-acetyl-1,3,6,8-tetrahydroxynaphthalene (ATHN, **16**). The Pks1 TE domain has an additional, unique activity and will catalyze concomitant deacetylation to form THN directly. THN is

further auto-oxidized to the naphthoquinone flaviolin (**17**), which is the observed product of *in vitro* reactions (Table 3.4).[19] These activities are clearly demonstrated in the present Pks1 control reactions (Figure 3.8b). In reactions lacking the TE domain, the major product is an isocoumarin hexaketide product **18** that arises from PT-catalyzed C2–C7 cyclization followed by spontaneous O9-C1 release (Table 3.4).[19] The fully reconstituted system, although inefficient, produces ATHN and flaviolin as well as the heptaketide product YWA1 (**5**). The deconstructed Pks1 shows a marked reduction in efficiency versus the intact or minimally deconstructed (SAT-KS-MAT-PT-ACP$_2$ + TE) proteins, which produce flaviolin nearly exclusively and in high yields.[19] The appearance of heptaketide products in reactions with the TE is concurrent with a reduction of overall biosynthetic capacity (relative to reactions lacking the TE). This observation indicates that chain-length control, while highly regulated, is likely tied to overall turnover.

Indeed, reactions of Pks1 SAT-KS-MAT with CTB1 (C$_{14}$) PT, ACP$_2$ and TE produce *nor*-toralactone (**4**) as the clear major product (Figure 3.8d). *Nor*-toralactone is created by CTB1 PT-mediated C4–C9 and C2–C11 aldol cyclizations followed by CTB1 TE-catalyzed O13–C1 bond closure of a heptaketide intermediate (Table 3.4) and is the native product of CTB1.[8] The generation of *nor*-toralactone is attended by a simultaneous reduction in shunt product formation indicating that both the PT and TE domains of CTB1 have a marked influence over chain-length control in this particular non-cognate PKS. It is likely that the CTB1 PT domain is capturing both hexaketide (native, C$_{12}$) and heptaketide (native + C$_2$, C$_{14}$) linear intermediates, but only efficiently catalyzing cyclization of the C$_{14}$ intermediate. In native CTB1, the TE domain affects a

dramatic enhancement of overall turnover.[8] A similar kinetic role for the CTB1 TE is probably at play in the present study. In reactions of Pks1 SAT-KS-MAT with CTB1 PT and ACP$_2$ only, very small quantities of *nor*-toralactone are formed while the $C_{12}H_{10}O_5$ hexaketide shunt product is greatly enhanced (Figure 3.8c). This observation follows native CTB1 reactions lacking the TE domain where overall biosynthetic efficiency is greatly reduced.[8] We propose that after CTB1 captures and cyclizes the heptaketide intermediates, the resulting enzyme-bound bicyclic acyl intermediate is slowly released through spontaneous pyrone formation. Therefore, the only products that result in enzymatic turnover are the hexaketide shunt products that have a higher rate of spontaneous release. This effect is completely masked by the highly efficient CTB1 TE domain, which quickly releases the bicyclic intermediates through pyrone formation, effectively enriching *nor*-toralactone in the product pool.

Reactions of Pks1 SAT-KS-MAT with wA (C_{14}) PT, ACP$_2$, and TE domains stand in contrast to the results of the Pks1/CTB1 non-cognate system. While the native wA produces and functions on a heptaketide intermediate (as in CTB1),[18] the Pks1/wA non-cognate system does not alter the native Pks1 hexaketide chain-length control. Reactions of Pks1 SAT-KS-MAT with wA PT and ACP$_2$ greatly enhance the production of the isocoumarin **18** indicating that the wA PT domain is catalyzing its expected C2–C7 cyclization of the non-native hexaketide linear poly-β-ketone (Figure 3.8c). In reactions including the wA TE, the chemistry is redirected towards ATHN formation, consistent with catalyzed release through C10–C1 Claisen cyclization (Figure 3.8d). As with the Pks1/CTB1 non-cognate system, inclusion of the wA TE led to a parallel reduction in shunt product formation. This effect was not observed in Pks1/wA reactions lacking the

TE domain. As with CTB1, this finding implicates the wA TE domain in the many-fold enhancement of overall efficiency through kinetic competition.

It is evident that reactions of Pks1 SAT-KS-MAT with Pks4 (C_{18}) PT, ACP, and TE redirected chemistry towards ATHN, albeit without the efficiency of the Pks1/wA non-cognate pair (Figure 3.8d). Reactions lacking the Pks4 TE domain produced the hexaketide isocoumarin **18** (Figure 3.8c). This outcome represents the expected Pks4 PT- and TE-catalyzed cyclizations but on a much smaller intermediate (Table 3.4). In fact, the hexaketide intermediate of Pks1 is a third smaller than the nonaketide intermediate of Pks4. This behavior is a testament to the considerable versatility of the Pks4 PT and TE domains, even if the redirection is inefficient. It also stands in stark contrast to the activity of the Pks1/CTB1 non-cognate pairs, where there is apparent chain-length discrimination. There is no evidence for the Pks4 domains exerting any influence over Pks1 KS chain-length control.

This result is even more remarkable when it is considered that neither PksA nor ACAS was able to complement Pks1 SAT-KS-MAT, with these reactions producing only shunt products (Figure 3.8). As with Pks4, both PksA and ACAS accept larger intermediates, C_{20} and C_{16}, respectively.[7a, 17] The fact that both of these enzymes demonstrate chain-length discrimination while Pks4 PT and TE do not further confirms the flexibility of Pks4. Similarly, it has been previously shown that the Pks4 PT and TE domains also process the heptaketide intermediate produced by CTB1 SAT-KS-MAT (Appendix D, Figure D.2). The Pks4 PT and TE may be inherently promiscuous with respect to substrate selectivity. While it is known that both domains process C_{18} intermediates, the exact chemical natures of these intermediates are unclear. The native
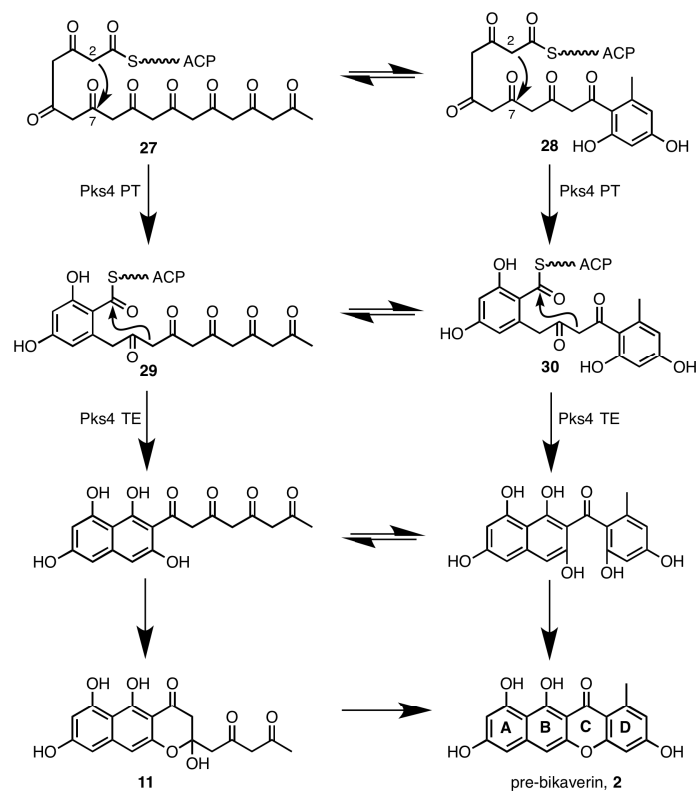
Figure 3.9 │ Possible intermediates of the Pks4 PT and TE domains.

metabolite of Pks4 is pre-bikaverin. It is certain that the PT and TE domains set the A and

B rings of pre-bikaverin, however the exact timing of the C and D ring formation is

unclear (Figure 3.9). It is possible that the Pks4 PT could accept a linear C_{18} poly-β-

ketone **27**, a monocyclic C_{18} intermediate **28**, or both. Either one of these species could

converge on pre-bikaverin. If it is the case that both of the species exist in the population

of reactive intermediates and the PT domain accepts either as a substrate, it may explain

the apparent lack of chain-length discrimination by Pks4 PT. Similarly, Pks4 TE could

accept either a monocyclic intermediate **29**, a "pre-cyclized" intermediate **30**, or both,

explaining its lack of specificity.

In all combinatorial reactions lacking a TE domain, an additional derailment

product, 6-(2′,4′-dihydroxy-6′-methylphenyl)-4-hydroxy-2-pyrone (**31**), is observed. The

derailment arises from improper cyclization of the C_{12} intermediate with C-O bond closer release yielding the pyrone. While this product has been observed in type III PKSs, it occurs here from uncatalyzed cyclization.[28] Higher levels of production in reactions lacking a TE domain underscore the importance of the TE domain for overall catalytic efficiency in PKSs. Without a TE domain forcing catalytic turnover, off-loading of derailments like product **31** can outpace even PT-catalyzed cyclizations. Further, the TE domain's editing role reduces the accumulation of this product by ensuring off-pathway routes are minimized.
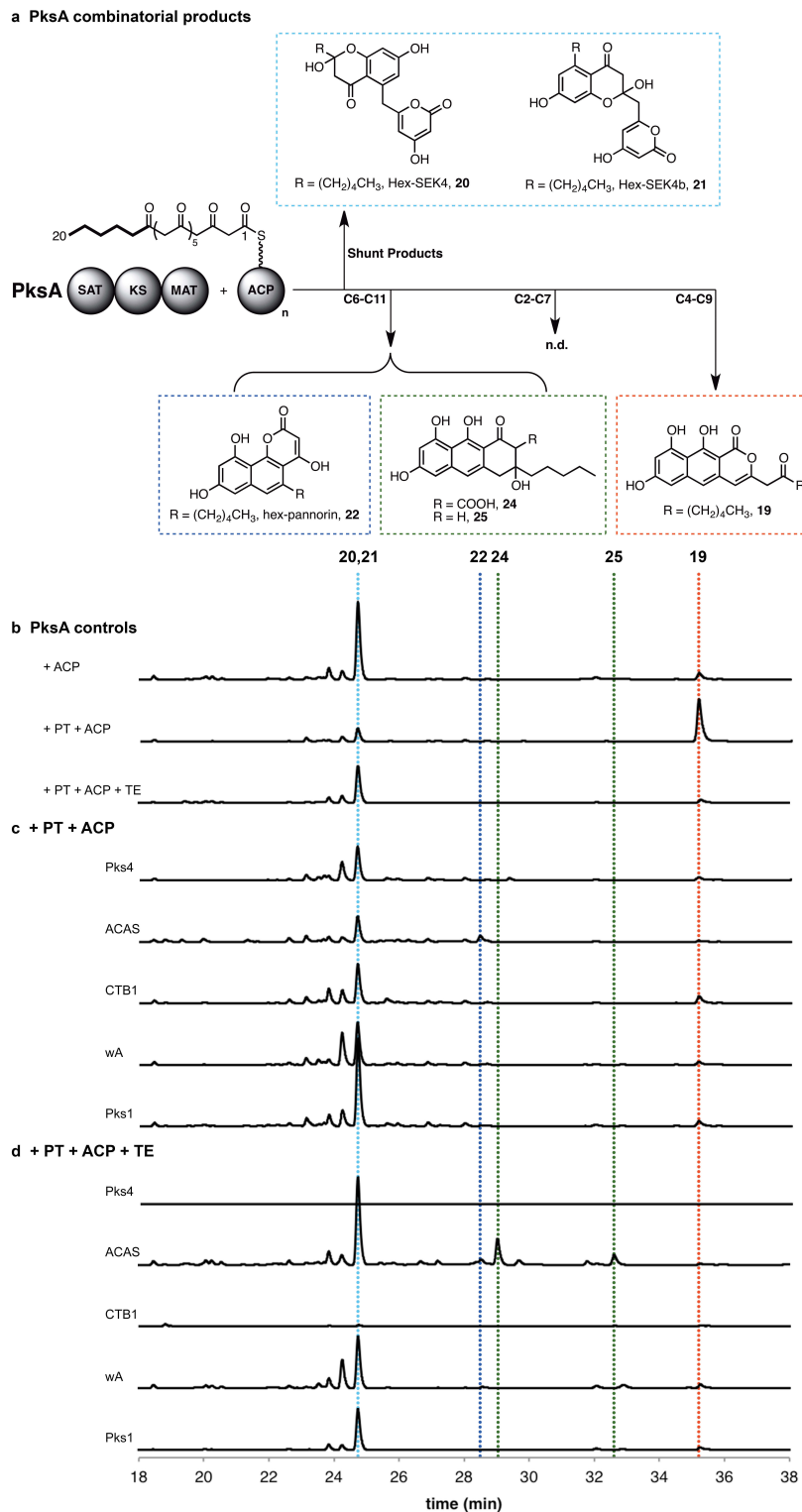
Figure 3.10│Product analysis of combinatorial reactions with PksA SAT-KS-MAT. Proposed products of combinatorial reactions are presented (**a**). The chromatograms at 280 nm for products of cognate control reactions (**b**), non-cognate combinatorial reactions with the indicated PT and ACP_n (**c**), and non-cognate combinatorial reactions with the indicated PT, ACP_n, and TE (**d**) are presented.

Table 3.5 | Intermediates and products produced by combinatorial reactions with PksA SAT-KS-MAT.
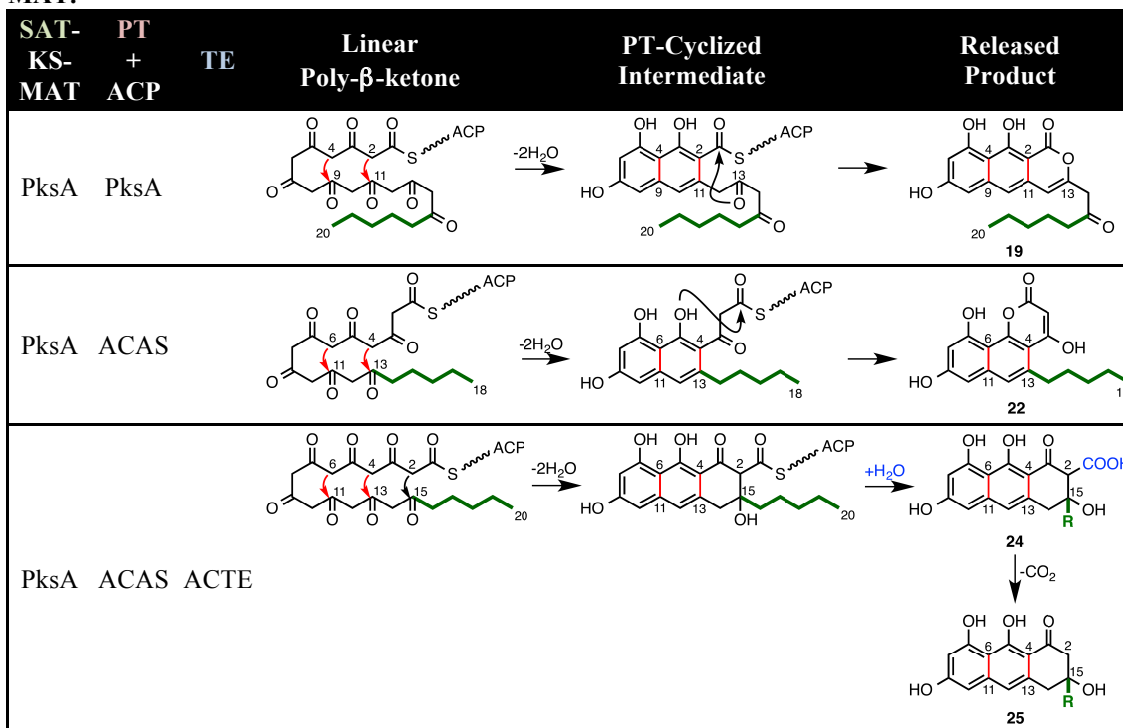
3.2.5. Combinatorial Reactions of PksA SAT-KS-MAT

HPLC product profiles for combinatorial reactions utilizing PksA SAT-KS-MAT are presented in Figure 3.10. Reactions for successful combinations are delineated in Table 3.5. Of the NR-PKSs investigated in this study, PksA is unique in that it accepts an abbreviated fatty acid hexanoyl starter unit.[26] As such, hexanoyl-SNAC was used as the starter unit in non-cognate combinations using the PksA SAT-KS-MAT tridomain. It is well established that PksA produces noranthrone (**1**), which is undetectable in the current assay due to low solubility, from the condensation and cyclization of hexanoyl with seven units of malonyl.[7a] Thus, the linear octaketide poly-β-ketone intermediate has a C_{20} chain length. The PksA PT domain catalyzes C4–C9 and C2–C11 aldol cyclizations and the PksA TE domain catalyzes release through C14–C1 Claisen cyclization. Reactions
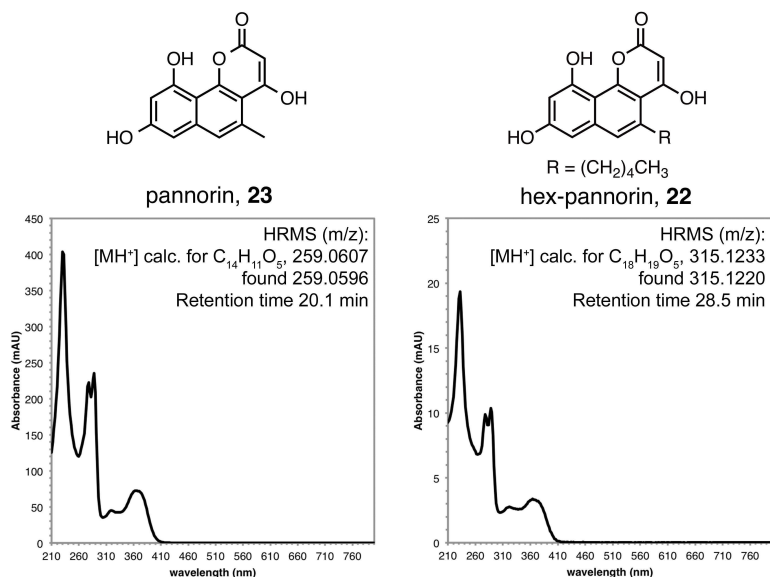
Figure 3.11 | UV-vis spectra and HRMS are presented for pannorin (23) and hex-pannorin (22).

lacking the PksA TE domain will make the spontaneous release product norpyrone (**19**, Table 3.5).[7a] Furthermore, the derailment products of the minimal PksA have been characterized as derivatives of SEK4 (**7**) and SEK4b (**8**) carrying a hexanoyl side chain derived from the starter unit (hex-SEK4 **20**, hex-SEK4b **21**).[23] Control reactions consisting of a cognate PksA system form these expected products (Figure 3.10b).

Of all the non-cognate pairs with PksA SAT-KS-MAT, only ACAS (C_{16}) was able to complement catalysis. Reactions of PksA SAT-KS-MAT with ACAS PT and ACP led to the formation of a pannorin-like molecule that we dubbed hex-pannorin (**22**, Figure 3.10c). Hex-pannorin displays the same spectral features as pannorin (**23**), as is to be expected for products sharing the same core architecture (Figure 3.11).[29] Interestingly, hex-pannorin arises from a chain length one extension unit shorter (C_{18}) than expected from the parent PksA SAT-KS-MAT, implicating ACAS PT in some level of chain-length control. This product would arise from proper C6–C11 and C4–C13 aldol

cyclizations by the ACAS PT domain followed by spontaneous pyrone formation with release from the enzyme (Table 3.5). A similar reaction occurs with the non-cognate pair of CTB1 SAT-KS-MAT with ACAS PT and ACP to make pannorin itself (Appendix D, Figure D.2).[13] Previous evidence suggests that the minimal PksA generates its linear C_{20} intermediate in a highly processive fashion.[23] These data imply that the ACAS PT can sample maturing chain-length intermediates, particularly those approaching the full, native chain length, a result that is in keeping with other non-cognate pairs described above. The low level of production could signify a slow off-loading rate, a feature that could also explain the decreased amounts of hex-SEK4 and hex-SEK4b derailment products. Intriguingly, the addition of ACTE to the system rescued the PksA KS chain-length control and led to the formation of a C_{20} product (Figure 3.10d) displaying a spectrum consistent with the atrochrysone core architecture (Figure 3.7).[25] Assuming native processing of the C_{20} linear intermediate by the ACAS PT and ACTE domains, the resulting product would be an atrochrysone carboxylic acid analog bearing a hexanoyl side chain (**24**, Table 3.5). Additionally, we observed a species **25** analogous to atrochrysone, likely arising from decarboxylation of acid **24** and displaying a characteristic spectrum for a molecule bearing the atrochrysone core architecture (Figure 3.7 and Table 3.5).

It is noteworthy that combinations of PksA SAT-KS-MAT containing the PT, ACP_n, and TE domains of either Pks4 or CTB1 led to an elimination of nearly all production, even shunt products (Figure 3.10d). This observation points towards the intrinsic editing role for all TE domains of these PKSs. As has been described for PksA, the TE domains of NR-PKSs not only catalyze the final release of product, but also
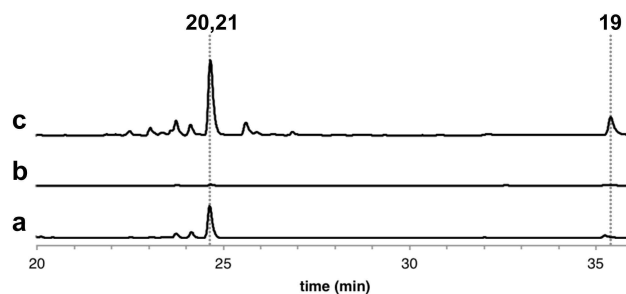
Figure 3.12│Evidence for the effect of TE-mediated editing in non-cognate combinatorial reactions. Chromatograms at 280 nm for products of reactions of PksA SAT-KS-MAT and PksA PT, ACP, and TE (**a**); or CTB1 PT, ACP$_2$, and TE (**b**); or CTB1 PT, PksA ACP, and CTB1 TE (**c**).

intervene when catalysis has been stalled.[23] Derailment products are expected to accumulate on the ACP domain of a given NR-PKS if they do not have an efficient, spontaneous off-loading mechanism. The TE domain can catalyze the hydrolysis of these products, freeing the ACP active site for another round of productive catalysis. It is likely that in the case of the PksA/Pks4 or PksA/CTB1 non-cognate pairs, the hexanoyl-loaded ACPs appear as improperly loaded, derailment products to the non-cognate TEs. Thus, they hydrolyze the starter unit more rapidly than extension can occur, thereby shutting down the catalytic cycle at the initiation stage.

These competing kinetic processes are complicated by reactions in which the ACP is held cognate with the SAT-KS-MAT tridomain. In reactions of PksA SAT-KS-MAT and ACP with CTB1 (C$_{14}$) PT and TE, production is restored (Figure 3.12). Remarkably, this non-cognate system is proficient in norpyrone (**19**) formation as well as enhanced in shunt production. Norpyrone would be the product of proper CTB1 PT and TE-mediated cyclization, but on the C$_{20}$-PksA linear intermediate, a 43% longer chain than the native CTB1 intermediate. Although seemingly contradictory to the observed CTB1 TE editing of the previous reaction, this result is entirely consistent with the current understanding of TE editing. Editing must always be negotiated through the ACP

85

domain. The non-cognate PksA ACP and CTB1 TE pair likely interacts poorly relative to the cognate pair. Moreover, it is established that the hexanoyl-loaded PksA ACP initiates rapid and processive extension by the PksA SAT-KS-MAT, meaning the CTB1 TE cannot compete with the cognate minimal PksA. Overall, the competing interdomain kinetics would imply an enhanced efficiency towards extension with the CTB1 TE editing occurring later in the catalytic cycle after the C_{20} linear intermediate has been constructed. Thus, the editing by the CTB1 TE domain is suppressed through effective native substrate channeling, allowing for later stage redirection of chemistry towards norpyrone.

3.2.6. Reactions of Reassembled and Chimeric NR-PKSs

In general, soluble full-length NR-PKSs are difficult to obtain by heterologous expression in *E. coli*, a key experimental advantage of more reliably expressed, smaller dissected fragments for the deconstruction approach. The rapid reconstitution of component parts exemplified in 72 native and non-native combinations described in this study is a second experimental advantage to both deduce the function of each domain and observe what limits are imposed by their heterocombination on product determination and overall flux. While not the focus of this study, we prepared intact Pks1 and a CTB1– Pks1 chimera to assess the penalty deconstruction has on net synthetic efficiency.

Equivalent reactions containing either 10 μM full-length Pks1 or 10 μM each of Pks1 SAT-KS-MAT, PT, ACP_2, and TE (4-part combination) were compared and found to each produce flaviolin (**17**) as a major product (Figure 3.13). Additionally, THN (**6**) remained the principal product for intact Pks1. Because THN oxidizes to flaviolin spontaneously, quantification of their relative productivity can be approximated.
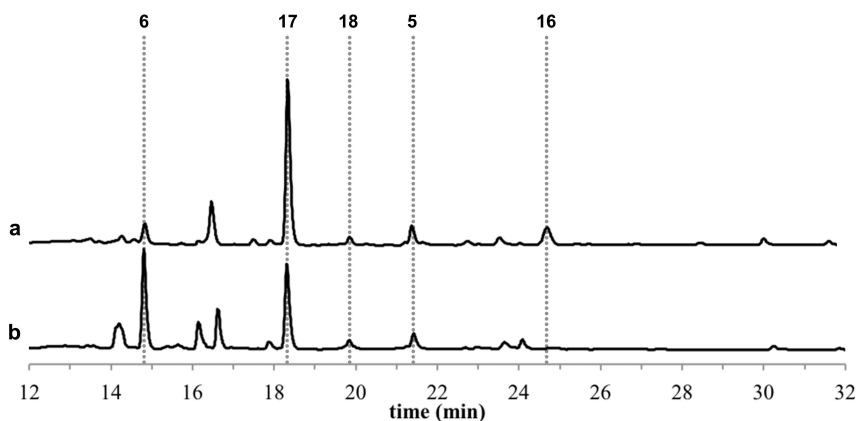
Figure 3.13 | Comparison of deconstructed and intact Pks1 reactions. Chromatograms at 280 nm are presented for products of deconstructed Pks1 (**a**) and intact Pks1 (**b**).

Interestingly, ATHN (**16**) is not observed in the intact system, indicating that it is a more efficient enzyme. Derailment products **18** and YWA1 (**5**) are observed in comparable yields for both systems. Additionally, the intact Pks1 was able to produce flaviolin in reactions containing 1 μM of protein. At these concentrations deconstructed systems do not produce any detectable products. These results are in keeping with previous experiments with deconstructed NR-PKSs where it was shown for PksA that the extent of deconstruction is inversely related to overall productivity.[7a] Taken together, these observations suggest that at the extreme of *holo*-Pks1 and its 4-part reconstitution at 10 μM a penalty of only 2- to 3-fold is exacted, but will be less as 3- or 2-part combinations.[7a, 19] All remain faithful in product synthesis.

In a further test of the penalty for deconstruction, we prepared a chimeric, full-length NR-PKS harboring CTB1 SAT-KS-MAT and Pks1 PT-ACP$_2$-TE. While this protein, dubbed M4P6, was soluble, we were hampered by low yields. Nevertheless, we were able to conduct *in vitro* reactions with 1 μM M4P6. The chimeric protein produced YWA1 (**5**) in much higher yields than the deconstructed system, which contained 10
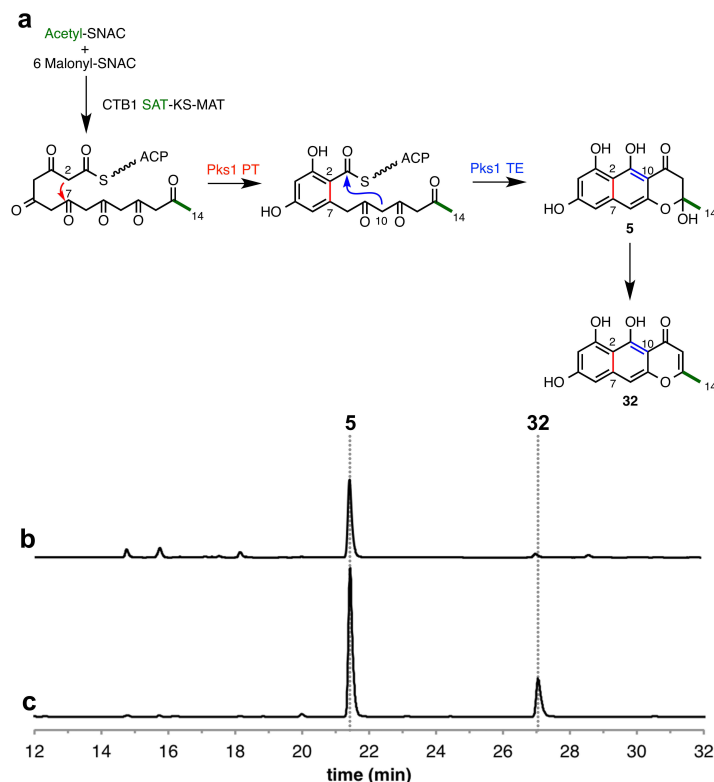
Figure 3.14│Comparison of a deconstructed and a chimeric CTB1/Pks1 combinatorial NR-PKS. The biosynthetic logic of the CTB1 SAT-KS-MAT/Pks1 PT-ACP$_2$-TE two-part heterocombination (**a**) reflects the activity of a the deconstructed system (CTB1 SAT-KS-MAT + Pks1 PT + ACP$_2$ + TE (**b**) and the chimeric M4P6 (**c**).

times the concentration of reconstituted protein domains (Figure 3.14). The dehydration

product of YWA1, *nor*-rubrofusarin (**32**) was the only other product observed in the

reaction of M4P6. Assuming product yield scales linearly with enzyme concentration

under a given set of reaction conditions and excess substrates, this outcome represents a

ca. 20-fold increase in YWA1 production by M4P6 over the 4-part deconstructed system

as estimated by HPLC.

3.3.   *Discussion*

The central facet of successful combinatorial biosynthesis in NR-PKSs revealed

in this study is that production of a non-native species is contingent upon enzyme-

directed cyclization and release outpacing spontaneous reactions and TE-mediated

88

editing. Therefore, an intimate understanding of the underlying mechanisms and relative rates governing the NR-PKS catalytic cycle is crucial for designing an efficient combinatorial enzyme. Recently, it was shown in our laboratory that fidelity of biosynthesis in PksA has been attributed to three main characteristics: processivity of extension, coordinated domain interactions leading to balanced active site occupancy, and TE-mediated editing of spurious intermediates.[23] The results of the current study suggest that these features may be generally true for the NR-PKSs, as successful complementation required the coordinated control of each of these mechanisms, while failure of just one led to the derailment of efficient redirected catalysis.

The complete interchangeability of ACPs in the minimal NR-PKS underscores the importance of the processivity of extension in these enzymes. Helix II of the ACP has been implicated in guiding client domain interactions.[30] While conservation of these recognition residues may explain some non-cognate compatibility, the uniform behavior of the mixed minimal systems with respect to the identity of the SAT-KS-MAT tridomain demonstrates the importance of rapid and complete extension. In PksA, transfer of the starter unit from the SAT domain to the ACP is both slow and crucial for proper extension.[23] It is not until the hexanoyl starter unit is delivered to the KS active site via the ACP that extension commences despite the MAT domain maintaining a high steady-state level of malonyl occupancy. Indeed, when acetyl and malonyl are improperly loaded on the ACP, extension will not begin, a behavior rationalized as being mediated through negative cooperation between the SAT and KS domains. However, once the correct hexanoyl starter unit is loaded onto the PksA KS, extension to the full-length C_{20} linear poly-β-ketone occurs rapidly and without detectable accumulation of intermediate chain

lengths.[23] The non-cognate minimal NR-PKSs show similar activity. ACPs that normally accept acetyl starter units will accept hexanoyl starter units and *vice versa*. In either case, extension is complete and truncation products of intermediate chain lengths are not observed. It is fortuitous that we carried out all our reactions with an intact SAT-KS-MAT tridomain fragment from a single parent NR-PKS, as the interaction of these domains govern correct extension.[23] We hypothesize that, at the very least, a matched SAT and KS pair is crucial in the development of a non-cognate NR-PKS. All efforts in our laboratory to express and purify a deconstructed SAT-KS didomain or KS monodomain have failed to date, making this hypothesis unverifiable by *in vitro* recombination.

This is not to say the KS domain exclusively controls the chain length. Early studies of combinatorial synthesis in bacterial type II PKS also yielded contradictory results.[12a] While the majority of minimal type II PKSs demonstrated complete chain-length control,[31] the *Streptomyces coelicolor* spore pigment *whiE* minimal PKS produced a wide assortment of polyketide products ranging in length from 14 to 24 carbons.[11] This observation was in contrast to the prevailing hypothesis that chain length in type II PKSs was controlled by the minimal PKS, consisting of a KS, a chain-length control factor (CLF), and an ACP.[31] In the case of *whiE*, an aromatase/cyclase (ARO/CYC) was implicated in chain-length stabilization, extending the role for downstream, tailoring enzymes in the central process of polyketide extension.[32] In NR-PKSs, the PT domain may exert similar control of chain length by binding and sampling growing chains, especially as they near their mature "programmed" length, as has been suggested for the PksA PT.[7a, 23] Acyl-ACP species have been shown to negotiate client

domain interactions in modular PKSs and FAS.[33] Indeed, in some of the combinatorial reactions containing the CTB1 and ACAS PT domains, the non-cognate PT influenced the final chain length, directing formation of a longer or shorter intermediate, respectively.

Efficient redirection of chemistry depends on the PT domain not only capturing a linear intermediate but also catalyzing the correct aldol cyclization(s). The cyclization register in successful combinations is always set by the PT domain and determined by the acyl-ACP thioester benchmark, as has been previously observed.[13, 34] PT domains, with some exceptions (*e.g.* PksA PT), display a degree of substrate promiscuity and can proficiently accept chains with plus or minus one extension unit. This permissiveness is in keeping with previous investigations of NR-PKS PTs.[13, 15] Some PTs however can also accept linear substrates with dramatically different chain lengths. Surprisingly, this level of promiscuity is usually in the direction of longer intermediates implying that some PT active sites are unexpectedly malleable. The PksA PT active site was completely incompatible with any non-cognate intermediate, suggesting that the starter unit effect in this case is carried over to the cyclization stage of biosynthesis. The PksA PT structure contains a hexyl-binding region shown to be crucial for catalytic cyclization.[27] It is likely that acetyl-initiated linear poly-β-ketones cannot bind this site, terminating any catalytic redirection by the PksA PT towards these intermediates.

The TE domain is the most crucial of all. As has been demonstrated in multiple NR-PKSs, the TE domain has roles both biosynthetic and editorial.[7b, 8, 19, 23] The transesterification of PT-cyclized intermediates onto the TE represents a likely slow step in the overall catalytic cycle. However, in most cases the TE-catalyzed reaction is rapid

and drives the overall biosynthetic program. This factor is most evident in the reactions of CTB1, where the inclusion of the TE domain enhances *nor*-toralactone production about 50-fold relative to the reaction without the TE domain.[8] Without the TE domain, off-loading is inefficient, effectively stalling the catalytic cycle and allowing for the accumulation of derailment products. Furthermore, the TE domain will only process intermediates comprised of their native substrate's core cyclization architecture. As has been recently suggested,[35] the TE domain serves as the catalytic lynchpin that determines success or failure in biosynthetic redirection.

The editing role of the TE domain of NR-PKSs has only recently been appreciated; however, it is essential for maintaining high catalytic turnover.[23] By intercepting spurious derailment products, the TE domain clears the ACP domain allowing for it to reenter the kinetically more productive, programmed catalytic cycle. Nowhere is this effect more cleanly demonstrated than with the combinatorial reactions of PksA and CTB1 (Figure 3.12). The CTB1 TE domain recognizes the hexanoyl loaded CTB1 ACP as an aberrant species and effectively hydrolyses it, thereby shutting down any catalysis at an early stage. The fact that changing the identity of the ACP relative to the TE rescues extension highlights a fundamental factor that is likely at play in all reactions. TE-mediated editing can effectively shut down redirection of chemistry at any stage in the catalytic cycle. Furthermore, a desired intermediate may be deemed unsuitable by the TE domain and eliminated, compromising the success of the combinatorial approach to the synthesis of unnatural polyketides. Thus, the kinetics of redirected synthesis must outpace the rate of TE editing. These interactions are difficult

to measure in these multistep processes, but are the theoretical factors determining combinatorial success and failure.

The deconstruction approach brings practical advantages of generally reliable protein over-production in *E. coli* and the rapid combinatorial assembly of functional units into reconstituted native and non-native catalytic systems. Although not assured at the outset, the reconstituted NR-PKSs have been found to recapitulate their wild-type synthetic capabilities. Similarly, heterocombination of catalytic components reassemble to elicit chemistry in a largely predictable fashion as we demonstrate here. While examples are limited, as one might intuitively expect, *holo* native and chimeric NR-PKSs show improved synthetic efficiency over their reconstituted component parts and give a reduced incidence of shunt and truncated products. Although only two cases are compared, the data suggest that these improvements will be greater for reconstituted heterodomains over native systems. These observations auger well for their rational assembly in single multidomain designed proteins for non-natural product synthesis.

3.4. Conclusions

In summary, we find that non-cognate "domain-swapped" NR-PKSs will redirect biosynthesis to non-native products if certain underlying constraints governing NR-PKS programming are met. While ketide extension appears to be complete in all non-cognate reactions, the interplay of downstream domains (namely the PT and TE domains) determines catalytic turnover to non-native products. When these domains capture suitable intermediates and efficiently carry out their programmed catalysis, redirection of chemistry takes place. Effective redirection requires that these processes outpace the rate of spontaneous derailments and TE-mediated editing. For more effective engineered NR-

PKSs, a greater understanding of the mechanisms and internal kinetics of the catalytic program is required.

3.5. Experimental Methods

3.5.1. Reagents and Biological Strains

All reagents and primers were purchased from Sigma-Aldrich (St. Louis, MO, USA), unless otherwise specified. Magnesium chloride, sodium chloride, potassium phosphate monobasic, potassium phosphate dibasic, and deoxynucleotides (dNTPs) were purchased from Thermo Scientific (Waltham, MA, USA). Isopropyl-β-D-thiogalactoside (IPTG), kanamycin, and nickel (high density) agarose beads were purchased from Gold Biotechnology (St. Louis, MO, USA). Imidazole was purchased from Acros Organics (Geel, Belgium). Sodium acetate (^{18}O, 95%) was from Cambridge Isotope Laboratories (Cambridge, MA, USA). Yeast extract and tryptone were purchased from Boston BioProducts (Ashland, MA, USA). Agar was purchased from bioWORLD (Dublin, OH, USA). Plasmids pET-24a(+) and pET-28a(+) were purchased from EMD Millipore (Darmstadt, Germany). All enzymes used for manipulation of DNA and 10 × HF Buffer were purchased from New England Biolabs (Ipswich, MA, USA). Bio-Rad protein assay dye for determining protein concentration by the Bradford assay was purchased from Bio-Rad laboratories (Hercules, CA, USA). Bacterial strain *Escherichia coli* BL21(DE3) was purchased from Sigma-Aldrich and made electrocompetent by standard protocols (see Appendix B).[36]

3.5.2. Preparation of Expression Constructs

DNA manipulations were carried out in accordance with established procedures.[37] Details of the expression plasmids for deconstructed NR-PKSs used in this

study are summarized in Table D.3 (Appendix D). Cut sites for protein deconstruction were guided by a variety of bioinformatics analyses including multiple sequence alignment, secondary structure prediction and the UMA algorithm for predicting interdomain regions.[9] Primers used for cloning new constructs used in this study are presented in Tables D.1 and D.2 (Appendix D). All expression plasmids were maintained in *E. coli* BL21(DE3) cells stored in 20% glycerol at −80 °C.

A C-terminal 6 × His tagged Pks4 SAT-KS-MAT construct was cloned by overlap extension PCR of exon fragments from genomic DNA (Appendix B). The resulting spliced fragment was used as template in PCR with primers Pks4-SAT-s1 and Pks4-NKA-3 (Appendix B). The resulting PCR product was digested with NdeI and NotI and ligated into pET-24a(+) to yield pEPks4-NKA-1. This plasmid was cut with NdeI and EcoRI and inserted into plasmid pEPks4-SATn-1[38] to yield pEPks4-NKAn. The resulting plasmid contains the N-terminal codon optimized sequence of pEPks4-SATn-1.

A full-length *C*-terminal 6 × His tagged CTB1 construct was cloned by overlap extension PCR. Overlap extension PCR DNA fragments were prepared by PCR from pCR-CTB1-ex #3 (a pCR®-Blunt plasmid containing CTB1 exon 3) template with primers CTB1-S3+1 and CTB1-ex6-3 and pCR-CTB1-ex7-9 #3 (a pCR®-Blunt plasmid containing CTB1 exons 7–9) template with primers CTB1-ex7-5 and CTB1-3. The resulting PCR products were spliced together using overlap extension PCR with outside primers CTB1-S3+1 and CTB1-3. The resulting overlap extension PCR product was digested with NheI-HF and NotI-HF and inserted into pECTB1-NKA6 to give pECTB1.

A C-terminal 6 × His tagged CTB1 SAT-KS-MAT/Pks1 PT-ACP$_2$-TE chimeric full-length PKS was cloned using Gibson assembly cloning to give plasmid pEM4P6-

1.[39] The crossover sites (CTB1 S1293 and Pks1 P1290) were selected to be five amino acids upstream of a conserved threonine at the N-terminal end of the PT domain. The CTB1 SAT-KS-MAT fragment was PCR amplified from pECTB1 template with T7 and M4P6-3 primers. The Pks1 PT-ACP$_2$-TE fragment was PCR amplified from pEPks1alt template with M4P6-5 and T7 Term primers. Gibson assembly was conducted using these two DNA fragments and NdeI and NotI linearized pET-24a(+) in equimolar amounts, as described in Appendix B.

3.5.3. Protein Expression and Purification

Proteins were prepared by heterologous expression in *E. coli* BL21(DE3) and purified by nickel-affinity purification as delineated in Appendix B. Eluted protein fractions were dialyzed at 4 °C overnight with three buffer exchanges into 100 mM potassium phosphate pH 7.0, 10% glycerol. Purified protein concentrations were determined by the Bradford assay using bovine serum albumin (New England Biolabs) as a standard as described in Appendix B.[40] ACP containing proteins were activated to *holo* form through the action of the promiscuous phosphopantetheinyl transferase Svp, as previously described (See Chapter 2.5.5.).[7a] Excess protein was flash frozen in liquid nitrogen and stored at −80 °C until further use.

3.5.4. In Vitro Reactions

In vitro reactions were conducted using *S*-acyl-*N*-acetylcysteamine (acyl-SNAC) substrates in place of acyl-CoAs, as previously described.[8, 13] Acetyl- and hexanoyl-SNACs were synthesized from the corresponding acid chlorides. The malonyl-CoA synthetase from *Rhizobium leguminosarum*, MatB, was used to produce HPLC purified malonyl-SNAC, as previously described.[23, 41] Purified enzyme fragments were

selectively recombined *in vitro* at 10 μM final concentration for each protein in 100 mM

potassium phosphate pH 7.0, 1 mM *tris*(2-carboxyethyl)phosphine (TCEP), and 10%

glycerol. Reactions were initiated with the simultaneous addition of either 0.5 mM

hexanoyl-SNAC (for reactions containing PksA SAT-KS-MAT) or 0.5 mM acetyl-SNAC

(all other reactions) and 2 mM malonyl-SNAC (all reactions) and were conducted for 4 h

at room temperature. Reactions were quenched by acidification with HCl, and products

were extracted into ethyl acetate. Organic fractions were combined, dried *in vacuo*, and

dissolved in 20% acetonitrile (*aq.*) at a final volume equivalent to the initial reaction

volume.

3.5.5. *Product Profile Analysis*

Reactions were analyzed by reverse phase HPLC on an Agilent 1200 instrument

(Agilent Technologies, Santa Clara, CA). Solvent A was water + 0.1% formic acid.

Solvent B was acetonitrile + 0.1% formic acid. Purified reaction extracts were injected

onto a linear gradient of 5% to 85% solvent B over 30 min at 1 mL/min on a Prodigy 5u

ODS3 column (4.6 × 250 mm, 5 μm [Phenomenex, Torrence, CA, USA]).

Chromatograms were recorded at 280 nm, 4 nm bandwidth with a background reference

of 650 nm, 100 nm bandwidth. Mass data, unless otherwise stated, were collected on a

Shimadzu LC-IT-TOF (Shimadzu Corporation, Kyoto, Japan) in positive ion mode fitted

with a Luna C18(2) column (2.0 × 150 mm, 3 μm [Phenomenex]) using a linear gradient

of 5% to 85% solvent B over 30 min at 0.2 mL/min.

3.5.6. *Polyketide Product Chemical Characterization*

Due to low titers, product identification by NMR spectroscopy was not feasible.

Notwithstanding, as NR-PKSs act through defined chemistry, distinct UV-vis absorption

profiles and high resolution MS data are sufficient for metabolite characterization.

Unequivocal structural assignment was available for the majority of products through

comparison to synthetic standards and literature values of fully characterized materials.

Provisional assignments based upon available data are made for selected species for

which unambiguous identification was not possible. A complete collection of

characterization data for each product is presented in Appendix D.

3.6.   References

[1]   D. J. Newman, G. M. Cragg, *J Nat Prod* **2012**, *75*, 311-335.
[2]   (a) H. G. Floss, *J Biotechnol* **2006**, *124*, 242-257; (b) B. Wilkinson, J. Micklefield, *Nature Chem Biol* **2007**, *3*, 379-386.
[3]   J. Staunton, K. J. Weissman, *Nat Prod Rep* **2001**, *18*, 380-416.
[4]   (a) J. M. Crawford, C. A. Townsend, *Nature Rev Microbiol* **2010**, *8*, 879-889; (b) R. J. Cox, T. J. Simpson, in *Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates, Vol. 459* (Ed.: D. A. Hopwood), Elsevier Academic Press Inc, San Diego, CA, **2009**, pp. 49-78.
[5]   T. Maier, M. Leibundgut, N. Ban, *Science* **2008**, *321*, 1315-1322.
[6]   J. F. Sanchez, Y. M. Chiang, E. Szewczyk, A. D. Davidson, M. Ahuja, C. E. Oakley, J. W. Bok, N. Keller, B. R. Oakley, C. C. C. Wang, *Mol Biosyst* **2010**, *6*, 587-593.
[7]   (a) J. M. Crawford, P. M. Thomas, J. R. Scheerer, A. L. Vagstad, N. L. Kelleher, C. A. Townsend, *Science* **2008**, *320*, 243-246; (b) T. P. Korman, J. M. Crawford, J. W. Labonte, A. G. Newman, J. Wong, C. A. Townsend, S. C. Tsai, *Proc Natl Acad Sci USA* **2010**, *107*, 6246-6251.
[8]   A. G. Newman, A. L. Vagstad, K. Belecki, J. R. Scheerer, C. A. Townsend, *Chem Commun* **2012**, *48*, 11772-11774.
[9]   D. W. Udwary, M. Merski, C. A. Townsend, *J Mol Biol* **2002**, *323*, 585-598.
[10]   (a) C. Khosla, S. Kapur, D. E. Cane, *Curr Opin Chem Biol* **2009**, *13*, 135-143; (b) L. B. Pickens, Y. Tang, Y. H. Chooi, in *Annual Review of Chemical and Biomolecular Engineering, Vol 2, Vol. 2* (Ed.: J. M. Prausnitz), Annual Reviews, Palo Alto, CA, **2011**, pp. 211-236.
[11]   Y. M. Shen, P. Yoon, T. W. Yu, H. G. Floss, D. Hopwood, B. S. Moore, *Proc Natl Acad Sci USA* **1999**, *96*, 3622-3627.
[12]   (a) B. S. Moore, J. Piel, *Antonie Van Leeuwenhoek* **2000**, *78*, 391-398; (b) Y. Tang, T. S. Lee, H. Y. Lee, C. Khosla, *Tetrahedron* **2004**, *60*, 7659-7671.
[13]   A. L. Vagstad, A. G. Newman, P. A. Storm, K. Belecki, J. M. Crawford, C. A. Townsend, *Angew Chem Int Ed Engl* **2013**, *52*, 1718-1721.
[14]   M. Choquer, K. L. Dekkers, H. Q. Chen, L. H. Cao, P. P. Ueng, M. E. Daub, K. R. Chung, *Mol Plant Microbe Interact* **2005**, *18*, 468-476.

[15] Y. R. Li, W. Xu, Y. Tang, *J Biol Chem* **2010**, *285*, 22762-22771.
[16] S. M. Ma, J. Zhan, K. Watanabe, X. Xie, W. Zhang, C. C. Wang, Y. Tang, *J Am Chem Soc* **2007**, *129*, 10642-10643.
[17] T. Awakawa, K. Yokota, N. Funa, F. Doi, N. Mori, H. Watanabe, S. Horinouchi, *Chem Biol* **2009**, *16*, 613-623.
[18] I. Fujii, A. Watanabe, U. Sankawa, Y. Ebizuka, *Chem Biol* **2001**, *8*, 189-197.
[19] A. L. Vagstad, E. A. Hill, J. W. Labonte, C. A. Townsend, *Chem Biol* **2012**, *19*, 1525-1534.
[20] H. Jiang, R. Zirkle, J. G. Metz, L. Braun, L. Richter, S. G. Van Lanen, B. Shen, *J Am Chem Soc* **2008**, *130*, 6336-6337.
[21] (a) H. Fu, S. Ebertkhosla, D. A. Hopwood, C. Khosla, *J Am Chem Soc* **1994**, *116*, 4166-4170; (b) H. Fu, D. A. Hopwood, C. Khosla, *Chem Biol* **1994**, *1*, 205-210.
[22] W. J. Zhang, Y. R. Li, Y. Tang, *Proc Natl Acad Sci USA* **2008**, *105*, 20683-20688.
[23] A. L. Vagstad, S. B. Bumpus, K. Belecki, N. L. Kelleher, C. A. Townsend, *J Am Chem Soc* **2012**, *134*, 6865-6877.
[24] S. M. Ma, J. X. Zhan, X. K. Xie, K. J. Watanabe, Y. Tang, W. J. Zhang, *J Am Chem Soc* **2008**, *130*, 38-39.
[25] M. Gill, P. M. Morgan, *ARKIVOC* **2001**, *2001*, 145-156.
[26] J. M. Crawford, B. C. R. Dancy, E. A. Hill, D. W. Udwary, C. A. Townsend, *Proc Natl Acad Sci USA* **2006**, *103*, 16728-16733.
[27] J. M. Crawford, T. P. Korman, J. W. Labonte, A. L. Vagstad, E. A. Hill, O. Kamari-Bidkorpeh, S. C. Tsai, C. A. Townsend, *Nature* **2009**, *461*, 1139-U1243.
[28] (a) I. Abe, S. Oguro, Y. Utsumi, Y. Sano, H. Noguchi, *J Am Chem Soc* **2005**, *127*, 12709-12716; (b) K. Springob, S. Samappito, A. Jindaprasert, J. Schmidt, J. E. Page, W. De-Eknamkul, T. M. Kutchan, *FEBS J* **2007**, *274*, 406-417.
[29] H. Ogawa, K. Hasumi, K. Sakai, S. Murakawa, A. Endo, *J Antibiot* **1991**, *44*, 762-767.
[30] P. Wattana-Amorn, C. Williams, E. Ploskon, R. J. Cox, T. J. Simpson, J. Crosby, M. P. Crump, *Biochemistry* **2010**, *49*, 2186-2193.
[31] R. McDaniel, S. Ebertkhosla, H. Fu, D. A. Hopwood, C. Khosla, *Proc Natl Acad Sci USA* **1994**, *91*, 11542-11546.
[32] T. W. Yu, Y. M. Shen, R. McDaniel, H. G. Floss, C. Khosla, D. A. Hopwood, B. S. Moore, *J Am Chem Soc* **1998**, *120*, 7749-7759.
[33] (a) L. Tran, M. Tosin, J. B. Spencer, P. F. Leadlay, K. J. Weissman, *Chembiochem* **2008**, *9*, 905-915; (b) E. Ploskon, C. J. Arthur, A. L. P. Kanari, P. Wattana-amorn, C. Williams, J. Crosby, T. J. Simpson, C. L. Willis, M. P. Crump, *Chem Biol* **2010**, *17*, 776-785.
[34] Y. Q. Xu, T. Zhou, Z. F. Zhou, S. Y. Su, S. A. Roberts, W. R. Montfort, J. Zeng, M. Chen, W. Zhang, M. Lin, J. X. Zhan, I. Molnar, *Proc Natl Acad Sci USA* **2013**, *110*, 5398-5403.
[35] Y. Xu, T. Zhou, S. Zhang, L.-J. Xuan, J. Zhan, I. Molnar, *J Am Chem Soc* **2013**, *135*, 10783-10791.
[36] W. J. Dower, J. F. Miller, C. W. Ragsdale, *Nucleic Acids Res* **1988**, *16*, 6127-6145.

[37]   J. Sambrook, D. W. Russell, *Molecular cloning: A laboratory manual*, Cold Spring Harbor Laboratory Press, Plainview, NY, **2001**.

[38]   J. M. Crawford, A. L. Vagstad, K. C. Ehrlich, C. A. Townsend, *Bioorg Chem* **2008**, *36*, 16-22.

[39]   D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison, III, H. O. Smith, *Nature Methods* **2009**, *6*, 343-U341.

[40]   M. M. Bradford, *Anal Biochem* **1976**, *72*, 248-254.

[41]   J. H. An, Y. S. Kim, *Eur J Biochem* **1998**, *257*, 395-402.

Chapter 4: Attempts to Rationally Engineer a Topopyrone Synthase Through Domain-Swapped, Chimeric Polyketide Synthases

4.1. Introduction

Polyketides are a large, structurally diverse family of natural products encompassing both pharmaceutical and toxic agents.[1] Polyketide synthases (PKS) are the enzymes responsible for polyketide biosynthesis. The mode of biosynthesis for most polyketides is through linear, assembly line construction catalyzed by large, multidomain type I PKSs.[2] Alternatively, in fungi, iterative, non-reducing PKSs (NR-PKS) synthesize aromatic polyketide products in a mode analogous to animal fatty acid synthases (FAS).[3] Reliable and rational engineering of PKSs has been a long-standing goal for those wishing to expand the pool of potential therapeutic polyketides; however, a limited understanding of PKS functionality has hampered engineering efforts. Recent advances in both NR-PKS enzymology[4] and engineering,[5] however have enabled the possibility of accessing custom designed NR-PKSs for the production of specific polyketide products.

NR-PKSs have a canonical domain architecture.[4] The three N-terminal domains—the starter unit:acyl carrier protein transacylase (SAT), ketosynthase (KS), and malonyl-CoA:acyl carrier protein transacylase (MAT)—together with the acyl carrier protein (ACP) compose a "minimal" NR-PKS that is requisite for extension of acetate units into an enzyme-bound poly-β-ketone.[5-6] The SAT selects a precursor (or starter unit) derived from an acyl-thioester[7] while the MAT introduces ketide extension units derived from malonyl-CoA.[6b] The KS, together with the ACP, catalyzes decarboxylative Claisen condensation to generate the linear, poly-β-ketone intermediate covalently bound

to the ACP.[4, 6b] The C-terminal domains, the product template (PT) and thioester (TE),

control the final stages of biosynthesis. The PT domain catalyzes regiospecific,

intramolecular aldol cyclization(s)/aromatization(s) of the ACP-bound poly-β-ketone.[8]

The TE releases the final product through one of several specific mechanisms—

Dieckmann cyclization/aromatization,[9] macrolactonization,[10] enol lactonization,[11] or

hydrolysis.[12] Thus, individual domains of the NR-PKS order the four determining

chemical features governing aromatic polyketide structure: starter-unit selection (SAT),

chain-length control (KS), cyclization register (PT), and product release (TE).

Coordinating domain specificity is the basis of most NR-PKS engineering efforts.

Currently, "domain-swapping" is the most successful method that has been applied

experimentally.[5, 13] In this approach, interchanging individual homologous domains

from different native NR-PKS imparts the swapped domain's native catalytic specificity

to the overall catalytic cycle. Investigating two-part heterologous combinations of a

representative set of deconstructed NR-PKSs codified a set of rules governing catalytic

proficiency of domain-swapped NR-PKS (see Chapter 3).[5, 13f] Any catalytically

proficient NR-PKS must balance the rate of productive chemistry with the rate of

spontaneous derailment and TE-mediated editing[5]—an intrinsic role of the TE

domain.[6b] Further, individual domains retain their native specificity while displaying an

expanded substrate selectivity—but within limits. For example, the PksA SAT domain

from *Aspergillus parasiticus*, which natively accepts a hexanoyl precursor, can be made

to select linear, aliphatic precursors;[14] or, the Pks1 PT domain from *Collectotichum*

lagenarium, which natively processes a hexa-β-ketone (C_{12}) intermediate, can catalyze

aldol cyclization with its native regiospecificity on a hepta-β-ketone intermediate (C_{14}).[5]

topopyrone A, **1a**   topopyrone B, **1b**   topopyrone C, **1c**   topopyrone D, **1d**
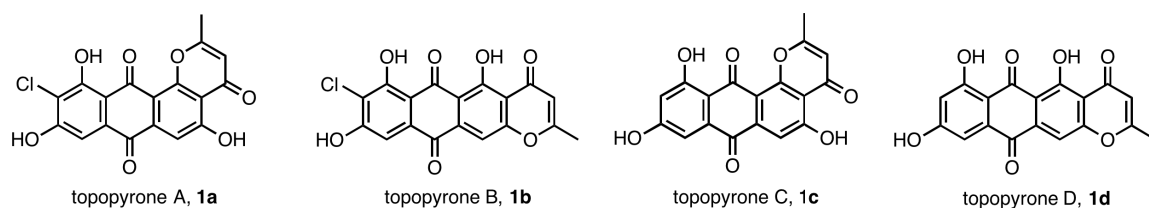
Figure 4.1 │ Structures of topopyrones A, B, C, D. The topopyrones are polyketide products of the fungus *Phoma* sp. BAUA2861.

The off-loading TE domain is of crucial importance in any domain-swapped paradigm as it serves as the final, and most important, decision gate of the catalytic cycle.[5, 13d]

Here, we attempt to rationally engineer a NR-PKS for the production of a targeted polyketide scaffold topopyrone (**1a–d**, Figure 4.1)—a family of potent human topoisomerase I inhibitors.[15] This goal represents a refinement from the usual NR-PKS engineering efforts, which are designed to test the limits of engineering by systematically swapping domains and observing the resulting chemistry. In the current study, we preselect the final product and reverse engineered the NR-PKS to achieve the desired scaffold. We adopt a domain-swapping approach such that the N-terminal and C-terminal domains come from the native Pks4 of *Gibberalla fujikuroi* and PksA of *A. parasiticus*, respectively. The N-terminal SAT-KS-MAT domains from Pks4 control starter-unit selection and chain-length limitation while the C-terminal PT-ACP-TE domains from PksA control cyclization register and product release. We attempt to further optimize engineered catalysis through rational PT active-site alteration. To date, no such attempt to design a NR-PKS for the production of a specific metabolite has been successful. A long-standing goal of the field is to use these enzymes for the fermentative production of novel metabolites and molecules of commercial interest. Engineering a topopyrone synthase would represent a significant step towards achieving that eventual goal.

4.2. Results

4.2.1. Generation and In Vitro Activity of Chimeric NR-PKSs

To test the feasibility of chimeric NR-PKS design, two-part heterocombination expression constructs were prepared by fusing the N-terminal tridomain (SAT-KS-MAT) and C-terminal tridomain (PT-ACP$_n$-TE) of non-cognate parent NR-PKSs. Six parent NR-PKS (see Appendix A) were used: PksA (Q12053, *A. parasiticus*), Pks4 (CAB92399, *G. fujikuroi*), ACAS (EAU31624, *Aspergillus terreus*), CTB1 (AAT69682, *Cercospora nicotianae*), wA (CAA46695, *Aspergillus nidulans*), and Pks1 (BAA18956, *C. lagenaria*). Crossover site selection was guided by a multiple sequence alignment of NR-PKSs and the tertiary structure of PksA PT with the intent to maintain as much of the MAT-PT linker of the parent sequences as possible in the final chimeric construct. Because the N-terminus of the PT domain is better defined than the C-terminus of the MAT domain, the crossover site was selected to be 4 amino acids upstream of the highly conserved PksA Thr1309 which marks the start of the PksA PT domain (Figure E.1, Appendix E). Using this crossover site, the complete set of two-part heterocombination expression constructs was prepared. Test expressions were conducted for all 30 members of this library. Twenty were found to induce expression of soluble protein (Table 4.1).

Table 4.1│Two-part heterocombination chimeric NR-PKSs. Each cell presents the SDS-PAGE results of an *E. coli* expression test of an individual two-part NR-PKS chimera. Rows indicate the identity of the SAT-KS-MAT. Columns indicate the identity of the PT-ACP-TE. The lanes from each gel represent the uninduced, induced, and elution bands for the indicated chimeric NR-PKS, from left to right. Constructs that do not express protein are not shown. The name for each chimeric NR-PKS is also indicated.

		PT-ACP$_n$-TE					
		PksA	Pks4	ACAS	CTB1	wA	Pks1
SAT-KS-MAT	**PksA**		M1P2	M1P3	M1P4 n.d.	M1P5	M1P6
	Pks4	M2P1		M2P3	M2P4 n.d.	M2P5	M2P6
	ACAS	M3P1	M3P2		M3P4 n.d.	M3P5 n.d.	M3P6
	CTB1	M4P1	M4P2	M4P3		M4P5	M4P6
	wA	M5P1 n.d.	M5P2 n.d.	M5P3 n.d.	M5P4 n.d.		M5P6 n.d
	Pks1	M6P1	M6P2	M6P3	M6P4 n.d.	M6P5	
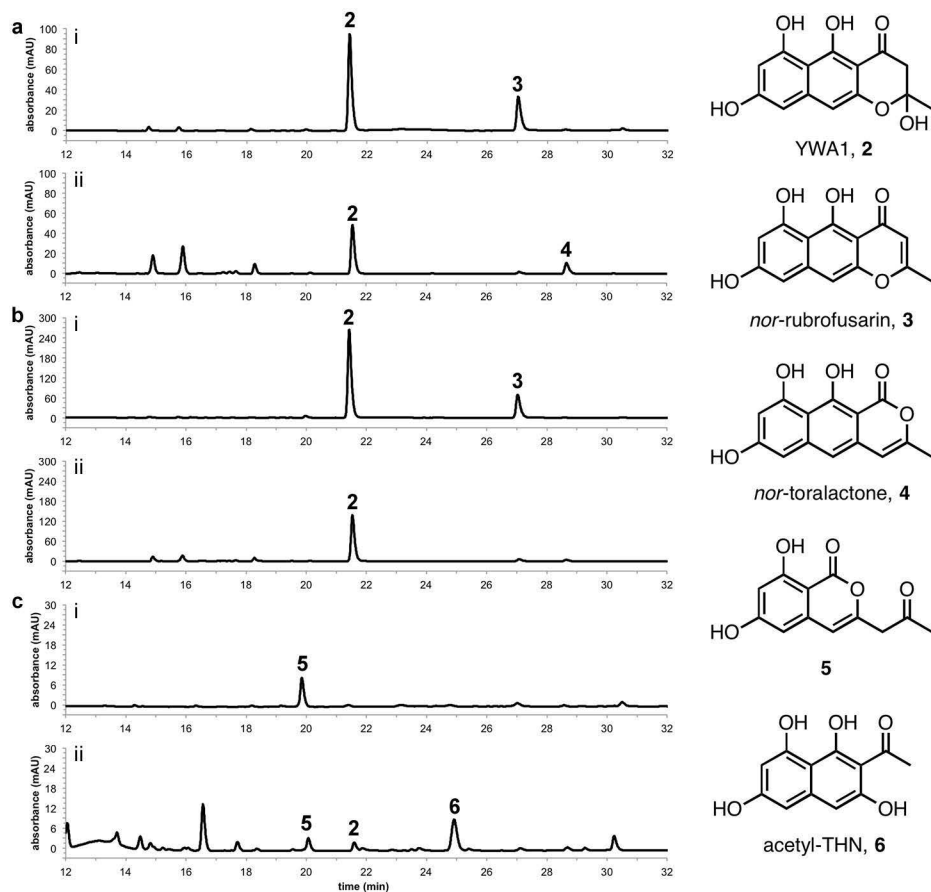
Figure 4.2│*In vitro* reactions of two-part intact and deconstructed NR-PKS heterocombinations. Chromatograms at 280 nm are presented for heterocombinations of (a) M4P2 (CTB1 SAT-KS-MAT + Pks4 PT-ACP-TE), (b) M4P6 (CTB1 SAT-KS-MAT + Pks1 PT-ACP$_2$-TE), and (c) M6P2 (Pks1 SAT-KS-MAT + Pks4 PT-ACP-TE) for (i) intact, chimeric enzymes at 1 µM and (ii) deconstructed enzymes (SAT-KS-MAT + PT + ACP$_n$ + TE) at 10 µM each proteins.

105

Three chimeric NR-PKSs were chosen to probe the limits of protein activity. The three examples investigated each corresponded to a set of domains that successfully catalyzed non-native polyketide formation when deconstructed NR-PKSs were used. The chimeras tested included M4P2 (CTB1 SAT-KS-MAT + Pks4 PT-ACP-TE), M4P6 (CTB1 SAT-KS-MAT + Pks1 PT-ACP$_2$-TE), and M6P2 (Pks1 SAT-KS-MAT + Pks4 PT-ACP-TE). Due to low protein yields from heterologous expressions, *in vitro* reactions of the fully intact chimeric proteins were conducted at 1 µM as opposed to the customary 10 µM for reconstitution reactions of deconstructed NR-PKSs. Product assignments were based upon HPLC retention times, UV spectral profile, and MS (Figures E.2–4, Appendix E). All spectra were identical to literature values for these products.

The reaction of M4P2 (CTB1/Pks4) produced YWA1 (**2**) and its dehydration product, *nor*-rubrofusarin (**3**, Figure 4.2a). Despite the low protein concentration, yields of this product were dramatically improved over the analogous deconstructed heterocombination (CTB1 SAT-KS-MAT + Pks4 PT + ACP + TE). Assuming that product output scales linearly under the given reaction conditions, the production of YWA1 by M4P2 represented a 20–30-fold increase over the reconstituted reaction, as estimated by HPLC peak areas. The reaction of M4P6 (CTB1/Pks1) also produced YWA1 and *nor*-rubrofusarin (Figure 4.2b). Again yields of YWA1 were increased 20–30-fold over the reconstituted reaction. In both cases shunt product formation was nearly eliminated. The reaction of M6P2 (Pks1/Pks4) produced isocoumarin **5** as its primary product (Figure 4.2c). This result diverged from the analogous deconstructed heterocombination (Pks1 SAT-KS-MAT + Pks4 PT + ACP + TE), which produced a variety of derailment products in addition to the anticipated redirection product—acetyl-

106
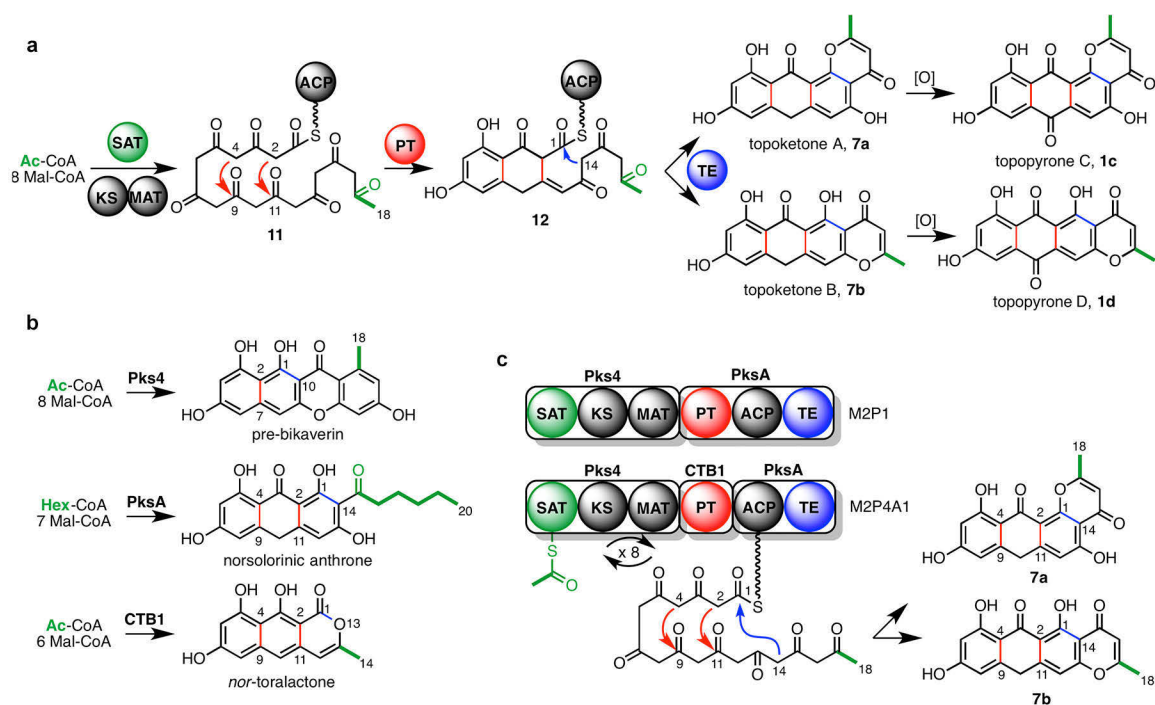
Figure 4.3│Rationale for topopyrone synthase design. (a) The underpinning NR-PKS logic of topopyrone biosynthesis. (b) Native products of parent NR-PKSs used to generate topopyrone synthases. Bonds are color coded according to the domains that generate them. (c) The M2P1 (top) and M2P4A1 (bottom) chassis for topopyrone synthases designed through domain-swapping.

tetrahydroxynaphthalene (acetyl-THN, **6**)—and small amounts of isocoumarin **5**—an

anticipated spontaneously released product following PT directed cyclization.

4.2.2. *Attempts to Engineer a Topopyrone Synthase*

Retrosynthetic analysis of topopyrones C and D (**1c, d**) served as the rationale for

a designed topopyrone synthase (Figure 4.3a). Topopyrones C and D were envisioned as

the oxidation products of anthropyrones **7a, b**—here named topoketones A and B,

respectively. Topoketones A and B were potential NR-PKS targets, arising from a

common acetyl-primed nona-β-ketone (C_{18}) adopting the C4–C9, C2–C11 PT-catalyzed

cyclization pattern and C14–C1 TE-catalyzed Claisen cyclization mode. The final ring—

the pyrone moiety—could form spontaneously through hemiketal formation followed by

dehydration. The SAT-KS-MAT from Pks4 was the only member of our parent NR-PKS

107

library that natively produced the desired acetyl-primed linear C_{18}-β-ketone intermediate (Figure 4.3b). Therefore, all designed topopyrone synthases contained the Pks4 SAT-KS-MAT tridomain. The PT domains from both PksA and CTB1 were selected to set the aromatic scaffold, as their native cyclization pattern was the desired C4–C9, C2–C11 cyclization motif (Figure 4.3b). The TE domain from PksA was the only member capable of performing the desired C14–C1 Claisen cyclization and was therefore used in all designed topopyrone synthases (Figure 4.3b). Two chimeric NR-PKS chassis were selected as potential topopyrone synthases: M2P1 (Pks4 SAT-KS-MAT/PksA PT-ACP-TE) and M2P4A1 (Pks4 SAT-KS-MAT/CTB1 PT/PksA ACP-TE, Figure 4.3c).

Additional active site manipulation was conducted on the PT domains of both the M2P1 and M2P4A1 chassis. The PksA PT domain—the only PT domain for which a structure exists—contains an unusually deep active site comprising three distinct binding
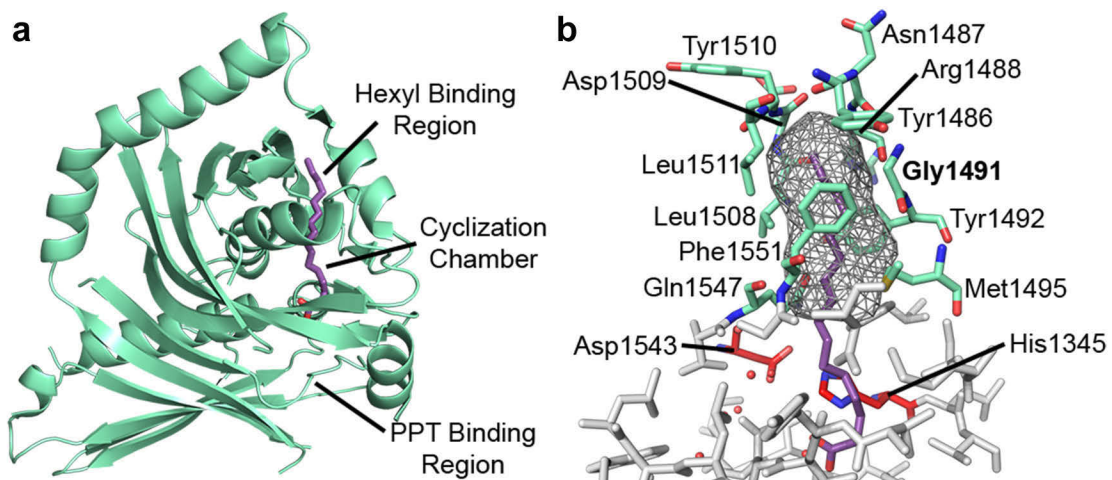
Figure 4.4 │ 1.8 Å resolution crystal structure of the PksA PT domain and the hexyl-binding region.[8]
(a) Overall structure of PksA PT monomer displaying the three distinct regions of the active site. Palmitate (in purple) occupies the cyclization chamber and hexyl-binding region. (b) Highlight of the hexyl-binding region of the PksA PT active site. Residues forming the hexyl-binding region are displayed in green. Residues forming the cyclization chamber are in white. The crucial catalytic dyad (His1345, Asp1543) is shown in red. Palmitate bound in the active site is shown in purple. The binding pocket of the hexyl-binding region is shown as a gray cage. Gly1491 forms part of the hexyl-binding region's wall. Mutations at this position would project side chain substituents into the bottom of hexyl-binding pocket.

108

subsites: the phosphopantetheine (PPT) binding region, the cyclization chamber, and the hexyl-binding region (Figure 4.4a). The hexyl-binding region was of particular interest in topopyrone synthase design. In the native protein, this site binds the hexanoyl starter-unit derived portion of PksA's linear octa-β-ketone intermediate (C_{20}, Figure 4.4b). We reasoned that this deep hydrophobic site would be detrimental to binding of an acetyl-primed nona-β-ketone (C_{18}) in the topopyrone synthases. A series of mutations was therefore created at the critical Gly1491 position. Gly1491 composes part of the wall of the hexyl-binding pocket. More sterically-demanding mutations to Gly1491 would project side chain substituents into the bottom of the hexyl-binding cavity, thereby occluding access to it. The residues that comprise the hexyl-binding region are well conserved among NR-PKSs with the exception of Gly1491, which is only found in norsolorinic acid synthases like PksA. In NR-PKSs that employ an acetyl-selective SAT, an aliphatic residue occupies this position. The mutations introduced side-chains of increasing size at this position for the M2P1 chassis (G1491A, G1491V, G1491I, and G1491L). The opposite approach was employed in the M2P4A1 chassis. In this case, Ile1468 occupies the homologous Gly1491 position. A series of mutants was generated at that position to further open up the bottom of the CTB1-derived PT active site for the M2P4A1 chassis (I1468V, I1468A, and I1468G). We rationalized that the extra space at the bottom of the PT active site would allow for better binding of the larger linear nona-β-ketone intermediate (C_{18}) as opposed to the native hepta-β-ketone intermediate (C_{14}). Nine potential topopyrone synthases were generated in total: M2P1, M2P1-G1491A, M2P1-G1491V, M2P1-G1491I, M2P1-G1491L, M2P4A1, M2P4A1-I1468V, M2P4A1-I1468A, and M2P4A1-I1468G. It should be noted that the numbering here does not
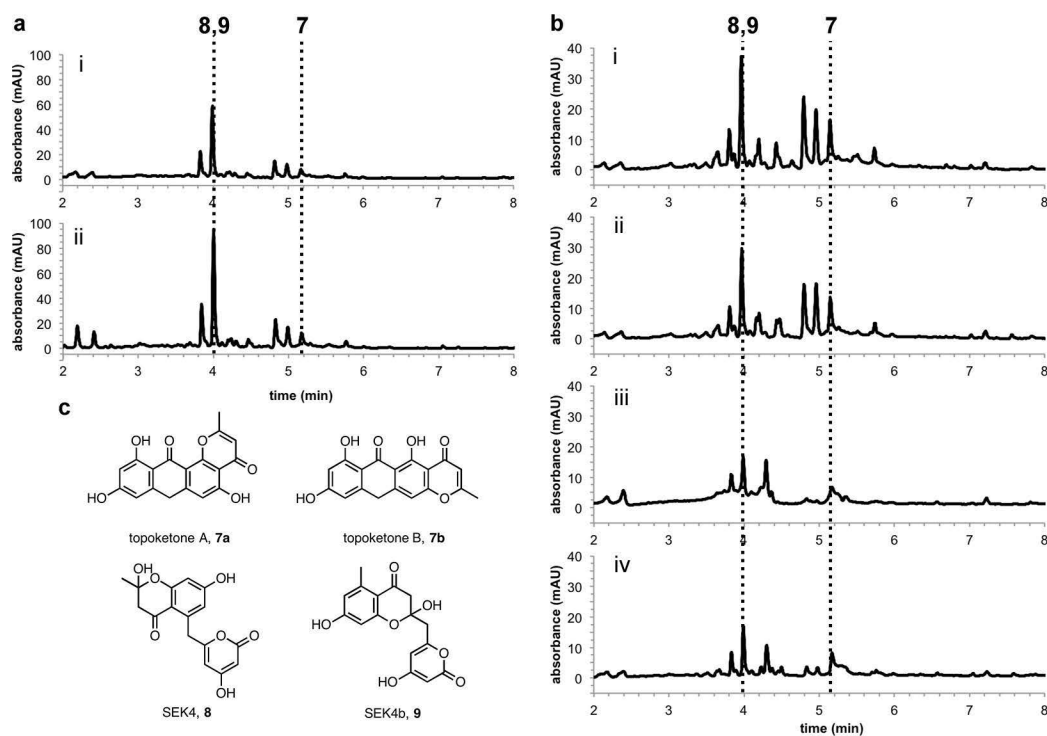
Figure 4.5│Reactions of the designed topopyrone synthase M2P1-G1491L. (a) 280 nm chromatograms for extracted products of *in vitro* reactions of 25 μM M2P1-G1491L: (i) extracted products and (ii) products sparged with air. (b) 280 nm chromatograms of extracted products of *in vitro* reactions of 25 μM M2P1-G1491L control mutants: (i) M2P1[P-]-G1491L, (ii) M2P1[P-]-G1491L + 100 μM PksA PT-G1491L, (iii) M2P1[P-T-]-G1491L , and (iv) M2P1[P-T-]-G1491L + 100 μM each PksA PT-G1491L and PksA TE. (c) Metabolites of M2P1-G1491L reactions. Derailment products SEK4 (**8**) and SEK4b (**9**) coelute.

indicate the exact sequence position of the mutation in the individual chassis, but rather corresponds to the relative native domain sequence position.

Initial reactions conducted at 1 μM protein did not show any clear indication of either topoketone or topopyrone formation by HPLC. Nevertheless, vanishingly small amounts of material corresponding to the topoketone exact mass were observed by LCMS—a far more sensitive method of analysis. We selected the best performing enzyme—M2P1-G1491L—for scaled biosynthesis. Protein concentrations were increased to 25 μM, requiring tremendously large culture volumes for heterologous expression—25 L per 250 μL *in vitro* reaction. Overall biosynthesis was increased in the high concentration reactions and a minor peak, which we assigned to topoketone, was

110

observed by both HPLC and LCMS (Figure 4.5a). Nevertheless, the major metabolites

were the SEK4 (**8**) and SEK4b (**9**) derailment products. Spectral characterizations were

positive for SEK4 and SEK4b (Figure E.5, Appendix E) and highly indicative of

topoketone (Figure E.6, Appendix E)—although unambiguous assignment for this

metabolite could not be made from the data collected. Several other products were

observed corresponding to nona-β-ketone derailment products. In an attempt to generate

topopyrone, air was bubbled through the extracted products of the M2P1-G1491L

reaction; however, no new products were observed and there was no appreciable turnover

of the likely topoketone peak (Figure 4.5a). Attempts to synthesize topopyrone from

analogous deconstructed enzymes were unsuccessful. These reactions comprised 10 μM

each protein with the compositions presented in Table 4.2.

Table 4.2 │ Composition of attempted deconstructed topopyrone synthases. Control reactions lacking
the TE or both the PT and TE were also conducted. In each case, biosynthesis of topopyrone was not
observed by HPLC.

Reaction	SAT-KS-MAT	PT	ACP	TE
1	Pks4		Pks4	
2	Pks4		PksA	
3	Pks4		CTB1	
4	Pks4	PksA-G1491L	Pks4	
5	Pks4	PksA-G1491L	PksA	
6	Pks4	PksA-G1491L	CTB1	
7	Pks4	CTB1-I1468G	Pks4	
8	Pks4	CTB1-I1468G	PksA	
9	Pks4	CTB1-I1468G	CTB1	
10	Pks4	PksA-G1491L	Pks4	PksA
11	Pks4	PksA-G1491L	PksA	PksA
12	Pks4	PksA-G1491L	CTB1	PksA
13	Pks4	CTB1-I1468G	Pks4	PksA
14	Pks4	CTB1-I1468G	PksA	PksA
15	Pks4	CTB1-I1468G	CTB1	PksA
16	Pks4	PksA-G1491L	PksA	wA
17	Pks4	PksA-G1491L	PksA	Pks4
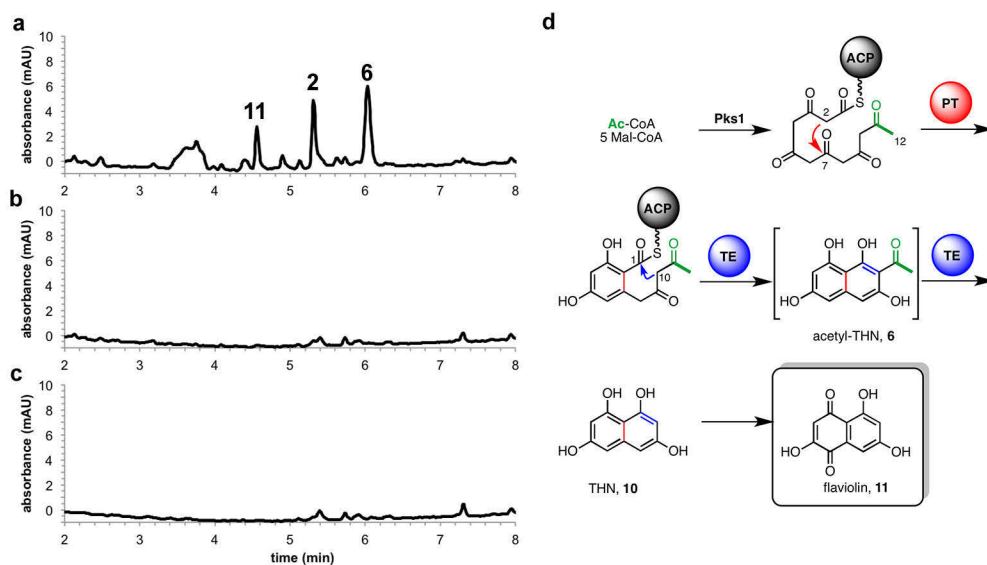
Figure 4.6│Reactions of Pks1 control mutants. 280 nm chromatograms of extracted products of (a) 1 μM Pks1, (b) 1 μM Pks1[P-], and (c) 1 μM Pks1[P-T-]. (d) The biosynthesis of Pks1 product. Pks1 produces a hexa-β-ketone linear intermediate. The PT domain catalyzes C2–C7 cyclization. The TE domain catalyzes C10–C1 Claisen cyclization generating acetyl-tetrahydroxynaphthalene (acetyl-THN, **6**) as an intermediate. The TE domain has an additional deacetylase activity and converts acetyl-THN to tetrahydroxynaphthalene (THN, **10**). THN spontaneously oxidizes to flaviolin (**11**), the primary observed product.

Given the small amounts of product formation, we were unsure if the observed topoketone was enzymatically synthesized or a product of derailed catalysis. To test for derailment, we generated two M2P1-G1491L control mutants. The first—M2P1[P-]-G1491L—corresponded to a PT null enzyme in which the catalytic His1345 of the PT active site was mutated to Ala. The second—M2P1[P-T-]-G1491L—corresponded to a PT and TE null enzyme in which both the PT catalytic His1345 and TE catalytic Ser1937 were mutated to Ala. We reasoned that if topoketone biosynthesis was enzymatic, it would be missing in the metabolite profiles of both null mutants, as the PT or TE domains would be inactive. Both control mutants showed the same profile of derailment products with evident formation of topoketone (Figure 4.5b). The overall biosynthetic capability was greater for M2P1[P-]-G1491L, which produced amounts of material comparable to M2P1-G1491L. The biosynthetic capability of M2P1[P-T-]-G1491L was

greatly reduced. The addition of 100 μM of PksA PT-G1491L or PksA TE to these reactions did not appreciably alter the product output (Figure 4.5b). All product formation was eliminated in analogous reactions of Pks1 control mutants (Pks1[P-] and Pks1[P-T-], Figure 4.6).

4.2.3. Fermentation of NR-PKS products in E. coli

To evaluate the ability of fungal NR-PKSs to produce polyketide products in *E. coli*, we created an *E. coli* BL21(DE3) strain harboring the pEPks1alt plasmid for inducible expression of Pks1. Six culture conditions were examined: Luria-Bertani broth (LB), LB supplemented with 10% glycerol, and Terrific Broth (TB) in either normal or baffled shake flasks. We used OD_{600} as a measure of cell growth and collected samples at various times after induction for up to 3 days (Figure 4.7a). Metabolites were extracted from the supernatant of collected samples and analyzed by HPLC for the presence of flaviolin (**11**), the spontaneous oxidation product of tetrahydroxynaphthalene (THN)—the native metabolite of Pks1. The flaviolin titer (mg/mL, Figure 4.7c) was calculated and
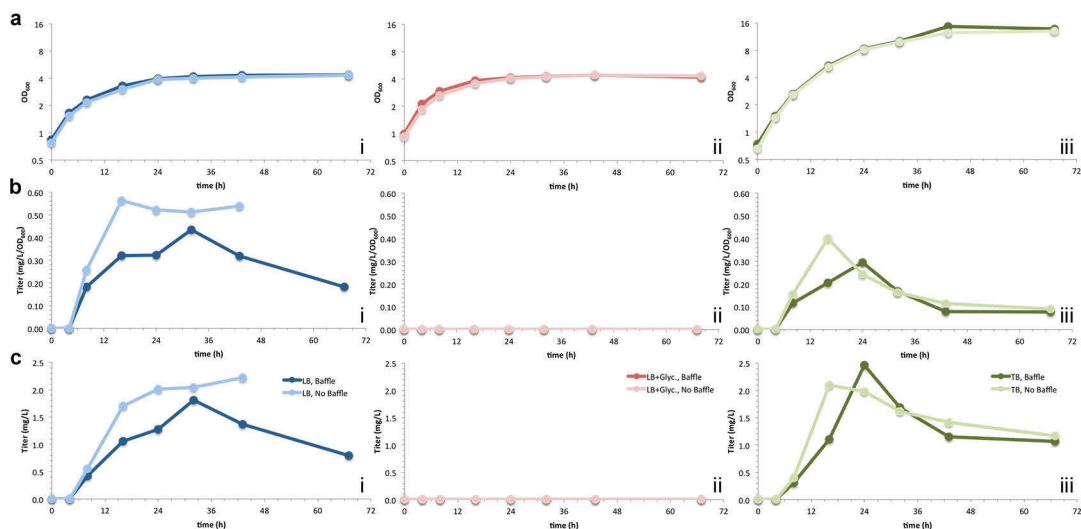
Figure 4.7 │ Flaviolin titers from *E. coli* fermentations. (a) Growth curves presented on a $\log_2$ scale, with flaviolin titers (b) normalized by cell density or (c) in mg/L. Cultures were conducted in (i) LB, (ii) LB + 10% glycerol, and (iii) TB.

normalized by cell density (mg/mL/OD$_{600}$, Figure 4.7b).

The presence or absence of baffles in the shake flask did not have an effect on cell growth. Cultures in LB and LB with 10% glycerol displayed similar growth curves. As expected, culture in TB supported higher density growth. In all cases, the stationary phase of growth was achieved, although the onset of this stage of growth occurred much later for cultures in TB. Flaviolin titers varied significantly for all culture conditions. No flaviolin was produced in cultures grown in LB with 10% glycerol. SDS-PAGE analysis indicated that Pks1 expression was not induced in these cultures. Modest titers of flaviolin were observed for both cultures in LB and TB reaching a maximum of 2.2 and 2.5 mg/mL, respectively. In general, cultures grown in flasks without baffles showed greater flaviolin titers. The largest flaviolin titers were coincident with the onset of stationary phase. Titers of flaviolin were relatively stable for cultures of LB in normal shake flasks, while both TB conditions and LB cultures in baffled flasks displayed a decrease in flaviolin titer after 24 and 32 h, respectively. Higher cell densities for cultures in TB did not translate to higher normalized titers of flaviolin. In fact, the cultures in TB performed worse than those in LB with respect to cell density-normalized flaviolin titers.

4.3. Discussion

The results presented herein further demonstrate the capabilities of the domain-swapping approach for rational engineering of NR-PKSs. Previous approaches in our laboratory have been to use deconstructed NR-PKSs to perform generate domain-swapped NR-PKSs.[4-5, 13f, 16] Historically, protein deconstruction has been the only method available to reliably access NR-PKSs and to rigorously investigate their biochemistry. While the approach is powerful, it is limited in its applications. Of

114

particular concern are the concentration demands of the deconstruction approach. Biochemical activity for a deconstructed NR-PKS is only observed above a certain protein concentration threshold—typically 10 μM for each protein component.[4] These limitations are of minor concern when investigating the *in vitro* chemistry of a domain-swapped NR-PKS, but are prohibitive in many settings. For example, fermentative production of non-native polyketides—an eventual goal of the field—using a deconstructed NR-PKS would not be a viable approach. Instead, using an "intact" domain-swapped NR-PKS would be a much more attractive means of achieving *in vivo* polyketide production.

Chimeric NR-PKSs have been used previously to produce polyketide metabolites in fungal hosts.[13b, c] These studies however have been quite limited, investigating only a single domain or NR-PKS pair. The work presented here represents the first time a systematic library of chimeric enzymes has been generated for the specific purpose of heterologous protein expression and purification. Indeed, the high performance of the library with respect towards soluble protein production validates the structure-guided approach to chimeric NR-PKS generation. Furthermore, the two-part chimeric NR-PKSs tested for functional activity were adherent to the established "rules" of engineered NR-PKSs (Figure 4.8).[5, 13f] For both M4P2 (CTB1/Pks4) and M4P6 (CTB1/Pks1), YWA1 (**2**) is the rational metabolite arising from the acetyl-primed hepta-β-ketone intermediate produced by the CTB1 SAT-KS-MAT[11] followed by Pks4[17] or Pks1[18] PT and TE directed regiospecific cyclizations.
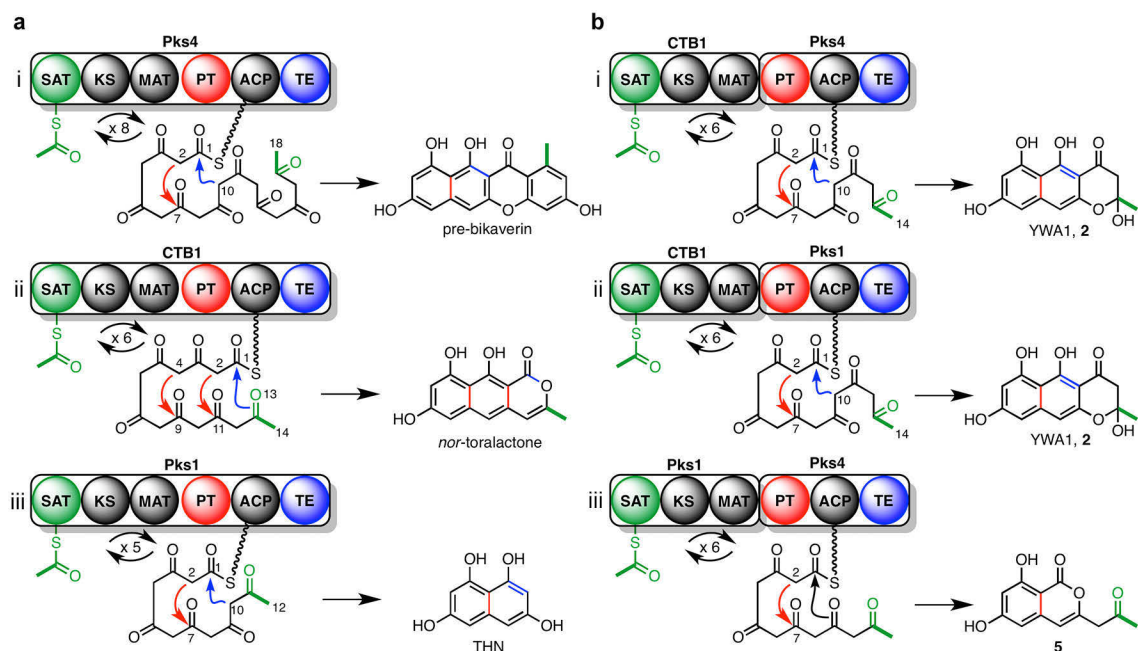
Figure 4.8│Mechanism for chimeric NR-PKS non-native product formation. (a) Reactions of native NR-PKSs: (i) Pks4, (ii) CTB1, and (iii) Pks1. (b) Reactions of chimeric NR-PKSs follow the biosynthetic logic proscribed by their constituent domains: (i) M4P2 (CTB1 SAT-KS-MAT/Pks4 PT-ACP-TE), (ii) M4P6 (CTB1 SAT-KS-MAT/Pks1 PT-ACP$_2$-TE), and (iii) M6P2 (Pks1 SAT-KS-MAT/Pks4 PT-ACP-TE).

The M6P2 (Pks1/Pks4) chimera stands in contrast to these successful enzymes. Although it did not produce its expected product—acetyl-THN (**6**)[5]—its clean production of isocoumarin **5** implies fidelity to the designed biosynthetic logic. Starter unit selection, chain length control, and cyclization pattern were all dictated by their associated parent domains—SAT, KS, and PT, respectively. Only the TE-catalyzed C10–C1 Claisen cyclization failed, even though this transformation showed limited success in the analogous deconstructed reaction. It should be noted that the deconstructed reaction contained 10 times the amount of protein than the chimeric reaction. It is likely that at this extreme of protein concentration, the TE domain was essentially saturating allowing for the successful redirection of chemistry. Because the domains are covalently attached in the chimeric enzyme, it is expected that the effective concentration of each domain is quite high. In fact, the significant enhancement of turnover for the chimeric enzymes

116

supports this notion. It was therefore surprising that the M6P2-TE domain failed. Of

course, the intermediate presented to the TE would be quite divergent from its native

substrate, perhaps contributing to poor catalysis. Under these conditions, it is not

unexpected that pyrone formation—the common off-loading derailment mechanism[4]—

could outcompete TE-directed Claisen cyclization. Nevertheless, the efficiency for all of

the chimeric enzymes over the deconstructed proteins augurs well, especially if chimeric

enzymes are to be used for the fermentation of new products.

Despite the promise of the domain-swapping approach, all attempts to rationally

engineer a topopyrone synthase failed. That is not to say that the method is without merit.

In fact, all of the engineered enzymes are proficient in poly-β-ketone extension implying

that many of the underlying design elements are sound. In particular, poly-β-ketone

extension would require correct global folding of the enzymes as well as successful

replication of crucial domain-domain interactions governing the early stages of NR-PKS

catalysis. We attribute the failure of these enzymes to flaws in the latter half of

biosynthesis—PT-catalyzed cyclization and TE-mediated release.

The designed topopyrone synthase that we explored in the most detail contained a

PT domain derived from PksA. The PksA PT domain is the only PT for which a structure

is known.[8] There are several features of this particular domain that make it unique. First,

the C4–C9, C2–C11 pattern of PT-catalyzed cyclization is quite rare. To date, only two

families of enzymes have been demonstrated to adopt this cyclization pattern: the

norsolorinic acid synthases[8]—the family containing PksA—and the *nor*-toralactone

synthases[11]—a family for which CTB1 is the only known member. The PT domains of

different cyclization families are distinct by sequence comparison;[13a] however, it

remains unclear if this divergence in primary structure translates to significant difference in tertiary structure. Homology models of different PT cyclization family members based on the PksA PT structure imply an alternatively shaped active site for these domains that could contribute to their regiospecificity.[8] Furthermore, the hexyl-binding region of the PksA PT domain is a unique feature that not only accommodates the unusual hexanoyl starter unit but also orders the binding of the linear poly-β-ketone in the PT active site. In fact, the PksA PT-G1491L mutant is nonproductive for the wild type enzyme.[8] Previously, in reactions of domain-swapped deconstructed NR-PKSs, all reactions containing the PksA PT domain failed if the linear poly-β-ketone presented to it was primed with an acetyl starter unit.[5] Taken together, these data suggest that the PksA PT domain is unique and perhaps cannot be relied upon to take up the expected linear nona-β-ketone intermediate of our designed topopyrone synthases. It was our hope that by mutating the critical Gly1491 some of these inherent challenges could be overcome. The results imply that—despite our efforts—the PT domain of the designed enzymes failed, thereby derailing the catalytic cycle.

The differential product outputs of the PT null and PT and TE null mutant enzymes suggest that the designed enzymes were further hampered by inherent TE editing. In addition to its biosynthetic role, the TE domain of NR-PKSs has a crucial editing function.[6b, 18] When improper intermediates occasionally are produced on the ACP, the TE editing mechanism intervenes to clear the ACP of the misformed intermediate through hydrolysis. In so doing, the TE resets the enzyme allowing it to enter a new catalytic cycle. Control of TE-mediated editing is a crucial feature in any domain-swapped paradigm, as previously described (see Chapter 3).[5] A careful balance

of directed biosynthesis and TE-mediated editing is of crucial importance for domain-swapped enzymes.[5, 6b] The overall level of production in the topopyrone synthase PT and TE null mutant was much lower than either the native chimera or the PT null mutant. This result implies that—in the case of an active TE—editing plays a central function, clearing the enzyme allowing from greater overall production. Without this editing function, production by the PT and TE null mutant was stalled as intermediates occupied the ACP. In this way, TE editing is a double-edged sword. In clearing the ACP, it allows for greater turnover, but limits directed biosynthesis by removing potentially productive intermediates prematurely. This feature of catalysis is of particular concern for a domain-swapped enzyme, as the structural features governing TE editing are unknown.

While our attempts to engineer a topopyrone synthase were unsuccessful, we maintain that our approach was sound and that given an expanded pool of domains, an engineered topopyrone synthase can be achieved. To this end, PT domains of the C4–C9, C2–C11 cyclization class must be identified that do not fall into the norsolorinic acid synthase family. Furthermore, a larger set of TE domains of the C14–C1 Claisen must be found. Of course, a better understanding of global NR-PKS structure is of critical importance for not only understanding catalysis but also for designing enzymes. To date, little is truly known of the global structure of these enzymes and the domain-domain interactions that govern their biochemistry. With a more sophisticated understanding of structure, we could better refine our domain-swapping approach. Finally, better expression conditions for NR-PKSs must be identified. The limited yields from *E. coli* BL21(DE3) expressions used for this study truly hampered our progress. If more reliable

conditions could be identified, a greater number of designed enzymes could be screened for activity.

The eventual goal for designed biosynthetic enzymes is to use these enzymes for metabolic engineering and fermentation of non-native products. Traditionally, the NR-PKSs have been considered as quite limiting due to their relatively poor expression in heterologous hosts. However, the advances in heterologous production of chimeric NR-PKSs in *E. coli* BL21(DE3) presented here promise a way forward. To that end, we tested the fermentation capabilities of *E. coli* BL21(DE3) with respect to a native NR-PKS. We chose Pks1 as our model system as its expression in *E. coli* BL21(DE3) is reliable and its *in vitro* product—flaviolin (**11**)—is easily distinguishable and quantifiable.[18]

Care must be taken with respect to culture conditions for high production of fungal polyketide products in *E. coli*. The cold shock conducted immediately prior to protein induction is of critical importance. Without this step, induction of NR-PKS expression does not occur. Furthermore, the composition of the culture medium must be considered. Cultures conducted in LB + 10% glycerol have been shown to facilitate a higher malonyl-CoA to acetyl-CoA ratio[19]—a crucial precursor metabolite in polyketide biosynthesis. In fact, cultures of the bacterial, iterative PKS CalE8—the PKS central to calicheamicin biosynthesis by *Micromonospora echinospora*[20]—in *E. coli* conducted in LB + 10% glycerol not only helped with enzymatic expression but also in identification of accumulated enzymatic products. However, such an approach is not feasible for the NR-PKSs, as cultures conducted in LB + 10% glycerol failed to induce protein expression. It remains unclear why the addition of glycerol would effect protein expression in this fashion.

Flaviolin titers were highest at the onset of the stationary phase. After this point, titers general began to fall, especially in cultures conducted in baffled flasks. It is likely that changing metabolic conditions commensurate with the onset of the stationary phase lead to catabolism of flaviolin, decreasing its titer. Of course, this effect is to be expected and, generally, fermentations are not conducted in the stationary phase. While the flaviolin titers were modest, they represent a respectable benchmark for further optimization. Yields could be improved by a variety of methods. In particular, increasing flux through the pathway could be achieved by altering the endogenous precursor pool. Several metabolic engineering strategies have been adopted to improve malonyl-CoA pools for the express purpose of increased polyketide production.[21] Such approaches have been reasonably successful and could be easily adapted for the production of fungal polyketide products in *E. coli*. Additionally, both continuous product extraction and continuous fermentation could help improve titers.

4.4. Conclusion

In summary, we find that "domain-swapped" NR-PKS chimeric enzymes are readably attainable and can be heterologously expressed as functional enzymes in *E. coli*. Three chimeric enzymes were tested for activity. They were found to reproduce—within limits—the activity of their analogous domain-swapped deconstructed NR-PKS but with much improved productivities and efficiencies. Together, the data suggest that the underlying logic of polyketide biosynthesis is maintained in chimeric enzymes. Attempts to design a topopyrone synthase through domain-swapping were unsuccessful. Two chassis were considered and PT active site manipulations were attempted. Unfortunately, none of the engineered enzymes produced topopyrone. They were, however, competent

in polyketide extension implying that the approach was sound if not successful. With a more diverse set of domains from which to draw, we are confident that an engineered topopyrone synthase can be attained. Furthermore, pilot studies into the fermentation of fungal polyketide products in *E. coli* were promising.

4.5. *Experimental Methods*

4.5.1. *Reagents and Biological Strains*

All reagents and primers were purchased from Sigma-Aldrich (St. Louis, MO, USA), unless otherwise stated. Magnesium chloride, sodium chloride, potassium phosphate monobasic, potassium phosphate dibasic, and deoxynucleotides (dNTPs) were purchased from Thermo Scientific (Waltham, MA, USA). Isopropyl-β-D-thiogalactoside (IPTG), kanamycin, and nickel (high density) agarose beads were purchased from Gold Biotechnology (St. Louis, MO, USA). Imidazole was purchased from Acros Organics (Geel, Belgium). Yeast extract and tryptone were purchased from Boston BioProducts (Ashland, MA, USA). Agar was purchased from bioWORLD (Dublin, OH, USA). Plasmids pET-24a(+) and pET-28a(+) were purchased from EMD Millipore (Darmstadt, Germany). All enzymes used for manipulation of DNA and 10 × HF Buffer were purchased from New England Biolabs (Ipswich, MA, USA). Bio-Rad protein assay dye for determining protein concentration by the Bradford assay was purchased from Bio-Rad laboratories (Hercules, CA, USA). Bacterial strain *Escherichia coli* BL21(DE3) was purchased from Sigma-Aldrich and made electrocompetent by standard protocols (see Appendix B).[22]

4.5.2. Preparation of Expression Constructs

DNA manipulations were carried out in accordance with established procedures.[23] Details of the expression plasmids used in this study are summarized in Tables E.6–8 (Appendix E). Gibson assembly was used to generate chimeric NR-PKS constructs (see Appendix B for general protocol). Touchdown PCR was used to generate DNA components for Gibson assembly (see Appendix B for general protocol). The composition of each PCR in this study is presented in Table E.1 AND E.3 (Appendix E). Primers used are presented in Table E.5 (Appendix E). The composition of each Gibson assembly reaction in this study is presented in Table E.2 AND E.4 (Appendix E). The plasmid DNA part for Gibson assembly was prepared by digesting pET-24(a)+ with NdeI and NotI. All expression plasmids were maintained in *E. coli* BL21(DE3) cells stored in 20% glycerol at −80 °C.

4.5.3. Protein Expression and Purification

Proteins were prepared by heterologous expression in *E. coli* BL21(DE3) and purified by nickel-affinity purification as delineated in Appendix B. Eluted protein fractions were exchanged into 100 mM potassium phosphate pH 7.0, 10% glycerol by serial dilution and passing protein fractions through an Amicon Ultracentrifuge tube, according to the manufacturer's instructions. Purified protein concentrations were determined by the Bradford assay using bovine serum albumin (New England Biolabs) as a standard as described in Appendix B.[24] ACP monodomain proteins were activated to their *holo* form through the action of the promiscuous phosphopantetheinyl transferase Svp, as previously described.[4] Excess protein was flash frozen in liquid nitrogen and stored at −80 °C until further use.

4.5.4. In Vitro Reactions

In vitro reactions were conducted using acyl-CoA substrates (acetyl-CoA and malonyl-CoA), as previously described.[4, 8] Reactions of chimeric NR-PKSs were conducted at a final concentration of 1 μM in 100 mM potassium phosphate pH 7.0, 1 mM *tris*(2-carboxyethyl)phosphine (TCEP), and 10% glycerol. Reactions of deconstructed NR-PKSs were conducted at a final concentration of 10 μM each protein in 100 mM potassium phosphate pH 7.0, 1 mM TCEP, and 10% glycerol. In each case, reactions were initiated with the simultaneous addition of 0.5 mM acetyl-CoA and 2 mM malonyl-CoA and were conducted for either 4 h or overnight at room temperature. Reactions were quenched by acidification with HCl, and products were extracted into ethyl acetate. Organic fractions were combined, dried under vacuum, and dissolved in 40% acetonitrile (*aq.*) at a final volume equivalent to the initial reaction volume.

4.5.5. Product Profile Analysis

Reaction mixtures were analyzed by reverse phase HPLC on an Agilent 1200 (Agilent Technologies, Santa Clara, CA, USA). Solvent A was water + 0.1% formic acid. Solvent B was acetonitrile + 0.1% formic acid. Purified reaction extracts were injected onto a linear gradient of 5% to 95% solvent B over 10.8 min at 1.25 mL/min on a Kinetex XB-C18 column (4.6 × 75 mm, 2.6 μm, Phenomenex, Torrence, CA, USA). Mass data were collected on a Waters Acquity/Xevo-G2 UPLC-MS system (Waters, Milford, MA, USA). MS and MS/MS spectra were collected in positive ion mode. UV data were determined from the diode array detector (DAD) signal at the maximum HPLC chromatogram signal. A complete collection of characterization data for each product is presented in Appendix E.

4.5.6. E. coli Fermentation of Polyketide Products

Pilot studies of polyketide fermentation were conducted in *E. coli* BL21(DE3) cells harboring pEPks1alt—a pET-24a(+) based plasmid harboring the Pks1 gene from *C. lagenaria*. Test media included Luria-Bertani (LB) broth, LB supplemented with 10% glycerol, and Terrific Broth (TB). Test vessels included shaking culture flasks with or without baffles. Starter cultures were prepared by inoculating 10 mL LB + 50 μg/mL kanamycin from a −80 °C 20% glycerol cell stock of *E. coli* BL21(DE3) cells harboring pEPks1alt. Starter cultures were incubated at 37 °C with shaking at 250 rpm, overnight. Test cultures were inoculated from starter cultures by diluting 1:100. Test cultures were incubated at 37 °C with shaking at 250 rpm. Upon achieving an OD_{600} of ~0.7, all cultures were incubated without shaking in an ice-water bath for 30 min. Immediately following cold shock, Pks1 expression was induced with the addition of IPTG (final concentration 1 mM). Cultures were incubated at 19 °C with shaking at 250 rpm for 4 days. Aliquots (20 mL) were collected at prescribed times after IPTG induction and centrifuged at $4100 \times g$, 15 min, at 4 °C. The supernatant was decanted and acidified by the addition of HCl, and then stored at −80 °C until further use.

Total metabolite extracts from the collected supernatants were prepared through extraction with an equal volume of ethyl acetate, thrice. Organic fractions were combined, washed with brine, dried with sodium sulfate, and evaporated under vacuum until dry. Total metabolite extracts were dissolved in methanol, filtered through 0.2 μm PTFE filters, and analyzed by HPLC according to the same protocol for NR-PKS *in vitro* reactions. Titers for flaviolin were calculated from the peak areas of 280 nm chromatograms. A concentration standard curve was prepared from HPLC analysis of an

authentic flaviolin sample (Figure E.7, Appendix E). Flaviolin was generously provided

by Dr. Eric A. Hill (Johns Hopkins University, Baltimore, MD, USA).

4.6. References

[1]   G. M. Cragg, D. J. Newman, *Biochim Biophys Acta* **2013**, *1830*, 3670-3695.
[2]   J. Staunton, K. J. Weissman, *Nat Prod Rep* **2001**, *18*, 380-416.
[3]   (a) R. J. Cox, T. J. Simpson, in *Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates, Vol. 459* (Ed.: D. A. Hopwood), Elsevier Academic Press Inc, San Diego, CA, **2009**, pp. 49-78; (b) J. M. Crawford, C. A. Townsend, *Nature Rev Microbiol* **2010**, *8*, 879-889; (c) T. Maier, M. Leibundgut, N. Ban, *Science* **2008**, *321*, 1315-1322.
[4]   J. M. Crawford, P. M. Thomas, J. R. Scheerer, A. L. Vagstad, N. L. Kelleher, C. A. Townsend, *Science* **2008**, *320*, 243-246.
[5]   A. G. Newman, A. L. Vagstad, P. A. Storm, C. A. Townsend, *J Am Chem Soc* **2014**, *136*, 7348-7362.
[6]   (a) W. J. Zhang, Y. R. Li, Y. Tang, *Proc Natl Acad Sci USA* **2008**, *105*, 20683-20688; (b) A. L. Vagstad, S. B. Bumpus, K. Belecki, N. L. Kelleher, C. A. Townsend, *J Am Chem Soc* **2012**, *134*, 6865-6877.
[7]   J. M. Crawford, B. C. R. Dancy, E. A. Hill, D. W. Udwary, C. A. Townsend, *Proc Natl Acad Sci USA* **2006**, *103*, 16728-16733.
[8]   J. M. Crawford, T. P. Korman, J. W. Labonte, A. L. Vagstad, E. A. Hill, O. Kamari-Bidkorpeh, S. C. Tsai, C. A. Townsend, *Nature* **2009**, *461*, 1139-U1243.
[9]   T. P. Korman, J. M. Crawford, J. W. Labonte, A. G. Newman, J. Wong, C. A. Townsend, S. C. Tsai, *Proc Natl Acad Sci USA* **2010**, *107*, 6246-6251.
[10]   H. Zhou, K. Qiao, Z. Gao, M. J. Meehan, J. W. Li, X. Zhao, P. C. Dorrestein, J. C. Vederas, Y. Tang, *J Am Chem Soc* **2010**, *132*, 4530-4531.
[11]   A. G. Newman, A. L. Vagstad, K. Belecki, J. R. Scheerer, C. A. Townsend, *Chem Commun* **2012**, *48*, 11772-11774.
[12]   (a) J. F. Sanchez, Y. M. Chiang, E. Szewczyk, A. D. Davidson, M. Ahuja, C. E. Oakley, J. W. Bok, N. Keller, B. R. Oakley, C. C. C. Wang, *Mol Biosyst* **2010**, *6*, 587-593; (b) T. Awakawa, K. Yokota, N. Funa, F. Doi, N. Mori, H. Watanabe, S. Horinouchi, *Chem Biol* **2009**, *16*, 613-623.
[13]   (a) Y. R. Li, W. Xu, Y. Tang, *J Biol Chem* **2010**, *285*, 22762-22771; (b) T. Liu, Y. M. Chiang, A. D. Somoza, B. R. Oakley, C. C. C. Wang, *J Am Chem Soc* **2011**, *133*, 13314-13316; (c) H. H. Yeh, S. L. Chang, Y. M. Chiang, K. S. Bruno, B. R. Oakley, T. K. Wu, C. C. C. Wang, *Org Lett* **2013**, *15*, 756-759; (d) Y. Xu, T. Zhou, S. Zhang, L.-J. Xuan, J. Zhan, I. Molnar, *J Am Chem Soc* **2013**, *135*, 10783-10791; (e) Y. Xu, T. Zhou, S. Zhang, P. Espinosa-Artiles, L. Wang, W. Zhang, M. Lin, A. A. Gunatilaka, J. Zhan, I. Molnar, *Proc Natl Acad Sci U S A* **2014**, *111*, 12354-12359; (f) A. L. Vagstad, A. G. Newman, P. A. Storm, K. Belecki, J. M. Crawford, C. A. Townsend, *Angew Chem Int Ed Engl* **2013**, *52*, 1718-1721.
[14]   C. R. Huitt-Roehl, E. A. Hill, M. M. Adams, A. L. Vagstad, J. W. Li, C. A. Townsend, *ACS Chem Biol* **2015**, *10*, 1443-1449.

[15] D. Ishiyama, Y. Kanai, H. Senda, W. Iwatani, H. Takahashi, H. Konno, S. Kanazawa, *J Antibiot* **2000**, *53*, 873-878.

[16] C. A. Townsend, *Nat Prod Rep* **2014**, *31*, 1260-1265.

[17] S. M. Ma, J. Zhan, K. Watanabe, X. Xie, W. Zhang, C. C. Wang, Y. Tang, *J Am Chem Soc* **2007**, *129*, 10642-10643.

[18] A. L. Vagstad, E. A. Hill, J. W. Labonte, C. A. Townsend, *Chem Biol* **2012**, *19*, 1525-1534.

[19] Y. Takamura, G. Nomura, *J Gen Microbiol* **1988**, *134*, 2249-2253.

[20] (a) K. Belecki, J. M. Crawford, C. A. Townsend, *J Am Chem Soc* **2009**, *131*, 12564-+; (b) K. Belecki, C. A. Townsend, *Angewandte Chemie-International Edition* **2012**, *51*, 11316-11319.

[21] J. W. Choi, N. A. Da Silva, *J Biotechnol* **2014**, *187*, 56-59.

[22] W. J. Dower, J. F. Miller, C. W. Ragsdale, *Nucleic Acids Res* **1988**, *16*, 6127-6145.

[23] J. Sambrook, D. W. Russell, *Molecular cloning: A laboratory manual*, Cold Spring Harbor Laboratory Press, Plainview, NY, **2001**.

[24] M. M. Bradford, *Anal Biochem* **1976**, *72*, 248-254.

Chapter 5: Molecular Characterization of the Cercosporin Biosynthetic Pathway in the Fungal Plant Pathogen *Cercospora nicotianae*

5.1. Introduction

The plant pathogenic *Cercospora* species are a widespread and destructive genus

of ascomycetous fungi characterized by their production of the phytotoxin cercosporin (**1**,

Figure 5.1b).[1] Cercosporin belongs to the perylenequinone natural product family. The

perylenequinone metabolites share a characteristic core architecture (**2**) that is essential to

Figure 5.1 │ The currently proposed cercosporin biosynthetic pathway. The cercosporin toxin biosynthetic (CTB) gene cluster has been identified in *C. nicotianae* (a). The proposed cercosporin biosynthesis hinges upon the formation of carboxylic acid **5** by the NR-PKS CTB1. The direct product of CTB1 is *nor*-toralactone (**6**), precluding the proposed biosynthetic scheme (b).
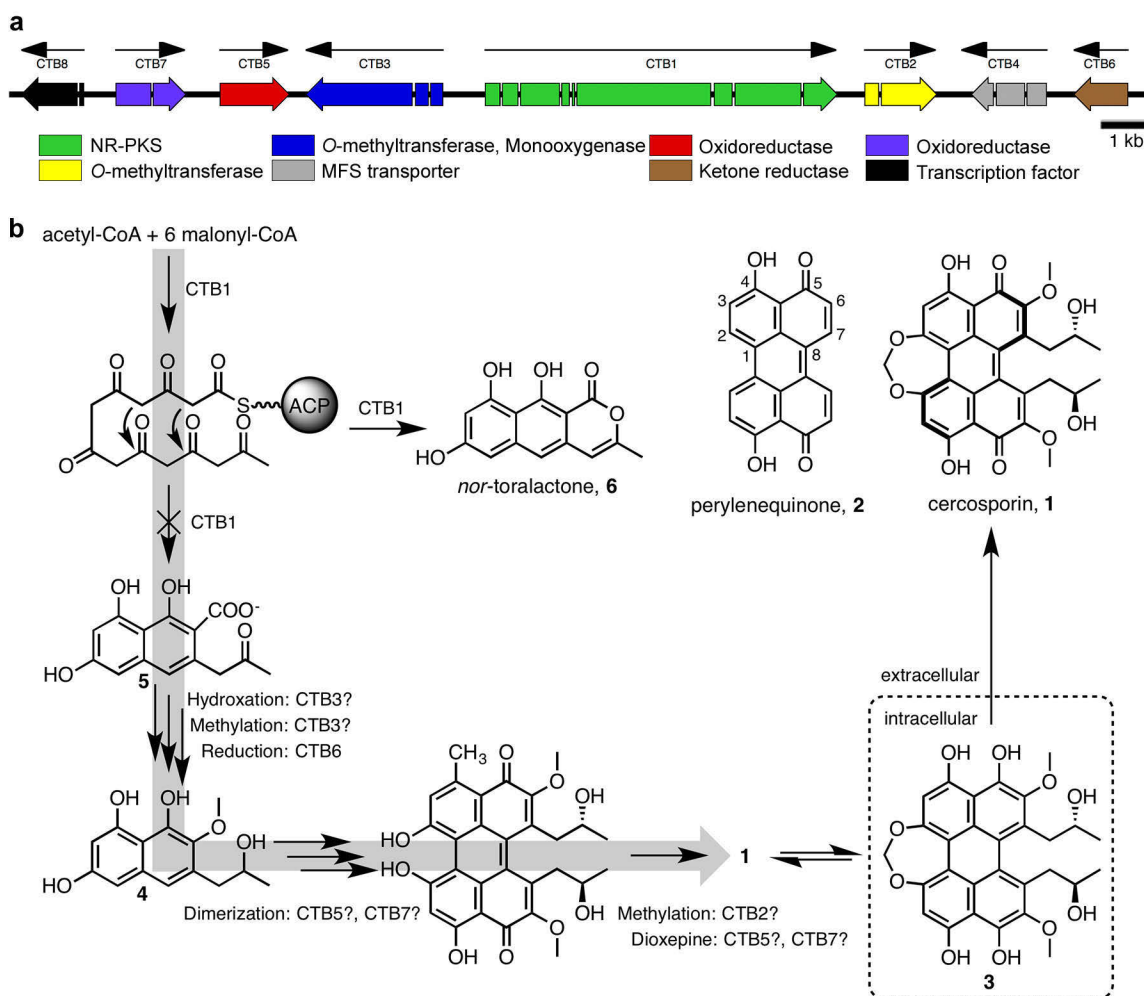
their toxicity.[2] Multiple perylenequinone natural products have been identified from fungal and aphidian sources.[3] Cercosporin was first isolated in 1957 from *Cercospora kikuchii* T. Matsu & Tomoyasu—a fungal pathogen of soybeans.[4] Perylenequinone metabolites contain the same highly oxidized conjugated pentacyclic core regardless of varied substituents—primarily at positions C7 and C7′—and are usually C_2-symmetric. Notably, most perylenequinone metabolites are helically chiral, demonstrating a preference for one particular atropisomer despite their conjugated, sp^2-hybridized carbon core.[5] Steric clashes between substituents in the mature perylenequinone metabolites account for the observed helical chirality. The biosynthetic origin of this property is unknown.

Cercosporin—like all perylenequinone metabolites—functions as a photosensitizing agent.[6] Upon absorption of visible light, cercosporin facilitates transfer of this energy to O_2, leading to the production of the potent reactive oxygen species (ROS) singlet oxygen ($^1O_2$) and superoxide radical (O_2^-). These ROS cause indiscriminate damage to a variety of cellular targets including cell membranes, nucleic acids, proteins, and lipids.[7] Peroxidation of the cell membrane is particularly pernicious and is the primary mode of toxicity, causing ion leakage from the host organism.[7a, b] A direct cellular target of cercosporin has not been found and it is believed that its toxicity is entirely attributed to the production of the damaging ROS. As a result, cercosporin exerts broad toxicity to bacteria, fungi, and mice.[8] Production of cercosporin by *Cercospora nicotianae*—a fungal pathogen of tobacco—is crucial to the fungus's pathogenicity and is concomitant with lesion formation on tobacco leaves.[9] Multiple putative resistance mechanisms have been proposed in *C. nicotianae*, but the primary

Table 5.1 | The cercosporin biosynthetic gene cluster from *C. nicotianae*.

Gene	Accession Number	Length (bp)	Intron	Amino Acids	Domains and Motifs	InterPro Annotation	Predicted Function
CTB1	AY649543	7036	8	2196	PKS β-ketoacyl synthase PKS acyl transferase Polyketide product template domain Acyl carrier protein-like PKS thioesterase domain	IPR020841 IPR020801 IPR030918 IPR009081 IPR020802	*nor*-toralactone synthase
CTB2	DQ991505	1439	1	461	O-methyltransferase, family 2	IPR001077	*O*-methyltransferase
CTB3	DQ355149	2734	2	871	O-methyltransferase, family 2 Monooxygenase, FAD-binding	IPR001077 IPR002938	*O*-methyltransferase FAD-dependent monooxygenase
CTB4	DQ991506	1696	3	512	Major facilitator superfamily domain	IPR020846	MFS transporter
CTB5	DQ991507	1380	0	459	FAD linked oxidase, N-terminal	IPR006094	FAD-dependent oxidoreductase
CTB6	DQ991508	1074	0	357	NAD(P)-binding domain	IPR016040	NAD(P)H-dependent ketone reductase
CTB7	DQ991509	1401	1	450	Monooxygenase, FAD-binding	IPR002938	FAD-dependent monooxygenase
CTB8	DQ991510	1245	1	397	Zn(2)-C6 fungal-type DNA binding domain Aflatoxin regulatory protein	IPR001138 IPR013700	Transcription factor

mode of resistance is through reversible reduction of the perylenequinone moiety to dihydrocercosporin (**3**)—a species with little photosensitizing activity. Dihydrocercopsorin is oxidized to cercosporin spontaneously upon export from the fungus, restoring its photoactivated toxicity.[10]

Cercosporin is classified as a polyketide natural product as determined from an early substrate-feeding assay in which the metabolite bore the characteristic alternating polyketide labeling pattern from acetyl and malonyl substrates.[11] Eventually, a gene from *C. nicotianae* encoding a fungal polyketide synthase—dubbed CTB1 (cercosporin toxin biosynthesis 1)—was identified through restriction enzyme mediated integration mutagenesis.[9a] CTB1 is absolutely necessary for cercosporin production and bears all the hallmarks of an iterative, non-reducing polyketide synthase (NR-PKS).[12] Using

CTB1 as a benchmark, the complete cercosporin biosynthetic gene cluster from *C. nicotianae* was determined (Figure 5.1a, Table 5.1).[13] The cluster comprises eight genes, six of which are believed to be responsible for cercosporin assembly (CTB1, 2, 3, 5, 6, and 7).[9a, c, 13-14] The zinc finger transcription factor CTB8 co-regulates expression of the cluster,[13] while the major facilitator superfamily (MFS) transporter CTB4 exports the final metabolite.[9b] In addition to the identified gene cluster, the $Zn(II)Cys_6$ transcription factor CRG1 is implicated in both cercosporin production and resistance.[15] The $Zn(II)Cys_6$ family of transcription factors is unique to fungi and regulates a diverse set of cellular processes.[16] Regulation by CRG1 is complex and poorly understood; however, its expression is implicated in regulation of chemical detoxification, multidrug membrane transport, and antioxidant biosynthesis pathways.[15c] Cercosporin production is completely dependent upon exposure to light and *C. kikuchii* grown in the dark will not accumulate cercosporin.[17] The regulatory mechanism governing light dependence is unknown.

On the basis of individual genetic homology of the *C. nicotianae* gene cluster and retrobiosynthetic analysis, Chen and coworkers proposed a biosynthetic pathway for cercosporin (Figure 5.1b).[13] The proposed pathway hinges upon the metabolite's C_2-symmetry and the authors argued that dimerization of two identical aromatic intermediates **4** would lead to the perylenequinone core. They postulated that CTB1 produces carboxylic acid **5** from which the aromatic intermediate **4** is derived. This pathway is reasonable and accounts for the putative activity of all the biosynthetic gene cluster members; nevertheless, it is likely incorrect. Using an enzyme-deconstruction approach, we previously characterized the *in vitro* activity of CTB1.[12] Surprisingly, the

naphthopyrone *nor*-toralactone (**6**) is the unambiguous *in vitro* product of CTB1 (Figure 5.1b). The identification of this intermediate is problematic, as there is no clear member of the gene cluster that could presumably open the pyrone moiety—an event that must happen to access the perylenequinone core architecture. Herein, we present an alternative biosynthetic pathway bolstered by metabolites accumulated in pathway-interrupted *C. nicotianae* knockouts. Furthermore, we characterize the *in vitro* activity of CTB3—an unusual didomain protein containing *O*-methyltransferase and flavin-dependent monooxygenase domains—and implicate it in pyrone opening of *nor*-toralactone.

5.2.   Results

5.2.1. Chemical Identification of Intermediates

C. nicotianae has been previously cultivated in liquid potato dextrose broth (PDB) and on solid potato dextrose agar (PDA). While functional cercosporin biosynthetic gene knockout strains of *C. nicotianae* grew in PDB, reproducible metabolite profiles could not be obtained. Therefore, metabolites were isolated from cultures grown on PDA under constant light, which provided reproducible metabolite profiles. Cultures of wild-type *C. nicotianae* resulted in a clean metabolic profile with the red-pigmented cercosporin (**1**) as the principal metabolite (Figure 5.2a). Cercosporin was absent in the metabolite profiles of all functional cercosporin biosynthetic gene knockout strains. Secondary metabolite accumulation was missing in the ΔCTB1 mutant—the knockout mutant of the central NR-PKS of cercosporin biosynthesis (Figure 5.2b).
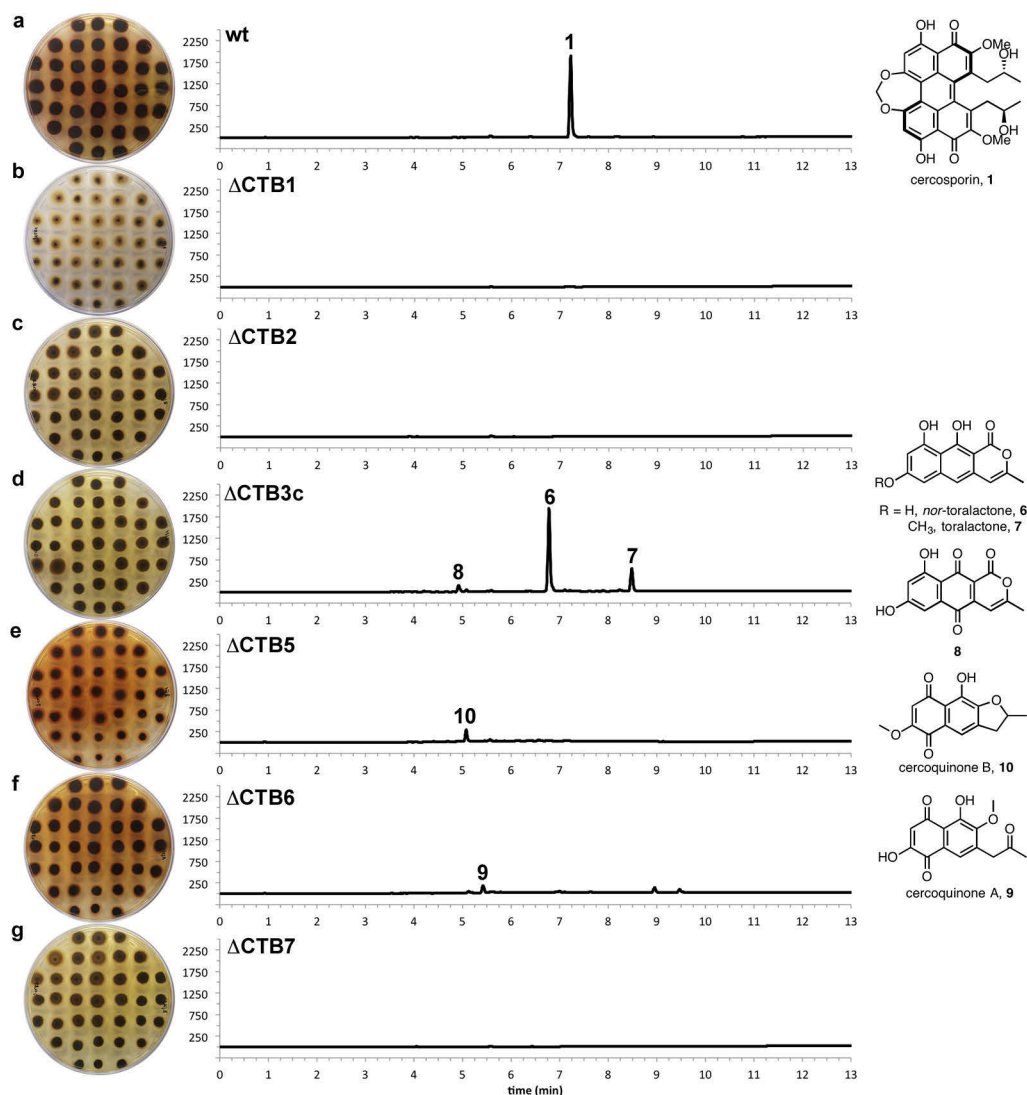
Figure 5.2│Metabolic profiles of CTB gene cluster mutant strains. Chromatograms at 250 nm of extracted metabolite profiles for (a) wild-type, (b) ΔCTB1, (c) ΔCTB2, (d) ΔCTB3c, (e) ΔCTB5, (f) ΔCTB6, and (g) ΔCTB7 strains displayed along with images of the mycelia for each strain. The metabolites were prepared at a concentration of 10 cm^2 colony surface area per mL in methanol. Identified cercosporin intermediate metabolites are displayed.

The structures of the major accumulated metabolites from *C. nicotianae* CTB gene cluster knockout strains were elucidated by characteristic UV spectra, exact masses, mass fragmentation patterns, and NMR spectra. We identified the previously characterized naphthopyrones *nor*-toralactone (**6**) and toralactone (**7**) both as metabolites of the ΔCTB3c mutant (Figure 5.2d). Also accumulated in the ΔCTB3c mutant was the oxidation product of *nor*-toralactone, naphthoquinone **8**. Two previously unobserved

133

naphthoquinones were isolated from the ΔCTB6 and ΔCTB5 mutants, which we called

cercoquinone A (**9**) and cercoquinone B (**10**), respectively (Figures 5.2f and 5.2e). No

major compounds were observed in the extracted metabolite profiles for the ΔCTB2 and

ΔCTB7 mutants (Figure 5.2c and 5.2g). The profiles for each of these mutants were

similar to those of the ΔCTB1 mutant.

Structural assignments for new compounds cercoquinones A and B were bolstered

by their ^{1}H NMR spectra, HRMS, mass fragmentation patterns, and UV spectra. The

aromatic CH resonances in either compound are not coupled, supporting the 1,4-quinone

oxidation pattern. Furthermore, each compound had fragmentation ions diagnostic for

1,4-naphthoquinones. Mass fragmentation patterns for naphthoquinone congeners have
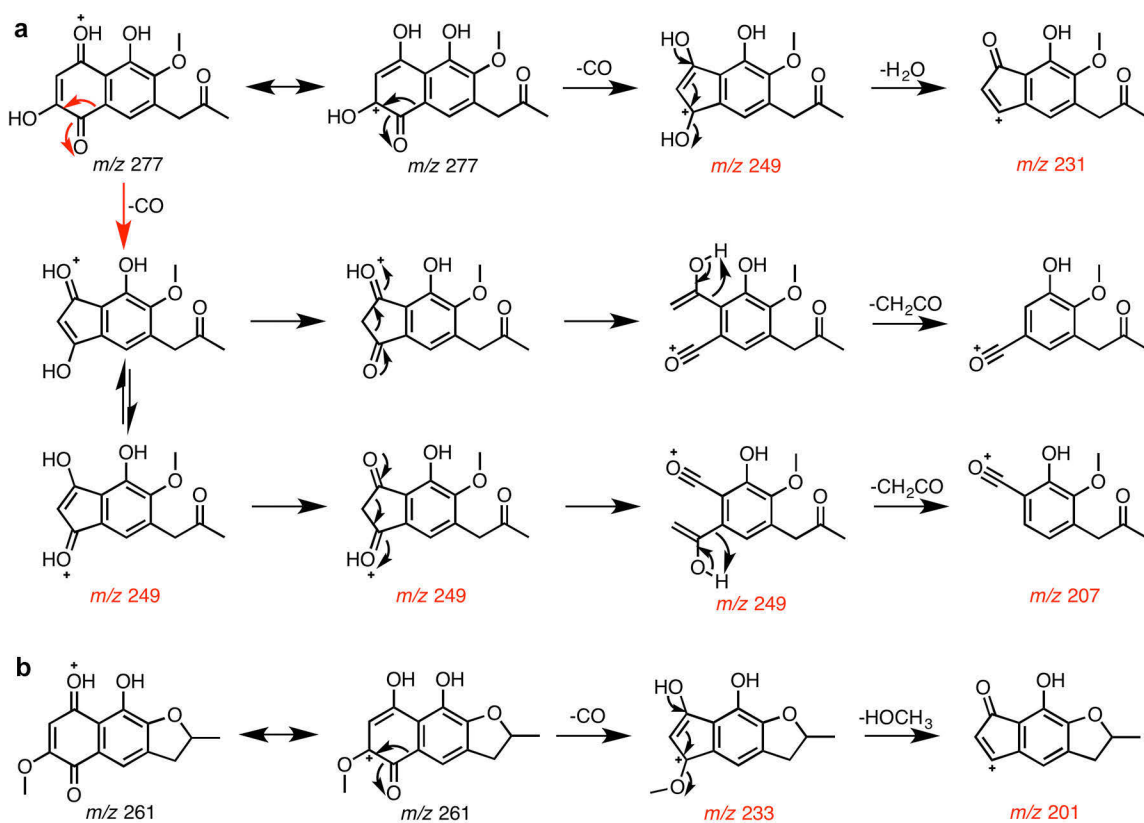
Figure 5.3 │ Mass fragmentation ions for cercoquinones A and B justify their structural assignment. Characteristic fragments of 1,4-napthoquinones are observed for (a) cercoquinone A and (b) cercoquinone B. Observed fragmentation ions are in red.

been previously described.[18] Protonated 2-hydroxyl-1,4-naphthoquiones have been shown to fragment along two predominant pathways—ring contraction with elimination of CO and water resulting in a carbocation or quinone ring-opening resulting in an oxonium ion. Fragment ions from both pathways were observed for cercoquinone A (*m/z* 207, 249, and 231, Figure 5.3a), while fragmentations arising from only the first pathway were observed for cercoquinone B (*m/z* 233 and 201, Figure 5.3b). The methoxyl substituent at the C2 position in cercoquinone B precluded the ring-opening mechanism. Thus, the ion fragments confirmed the sites of *O*-methylation in both cercoquinones A and B. Both cercoquinones A and B are similar to known compounds produced by various *Fusarium* species.[19]

5.2.2. *Phenotypic Analysis of Gene Knockout Strains*

The mycelia of wild-type *C. nicotianae* were blood red in color, with pigmentation occurring at about 4 days after inoculation (Figure 5.2a). The pigmented metabolites were primarily concentrated in the mycelia with a small amount exported into the agar surrounding individual colonies. The mycelia of each knockout strain displayed a different pigmentation from wild-type *C. nicotianae*, with pigmentation similarly occurring at about 4 days after inoculation. The ΔCTB1 mutant did not display any pigmentation (Figure 5.2b). The ΔCTB3c mutant adopted a dark yellow-brown coloration, with slight export of pigmented metabolites into the agar (Figure 5.2d). The mycelia of both ΔCTB6 and ΔCTB5 mutants turned a dark orange-red with significant export of colored compounds into the agar (Figures 5.2f and 5.2e). Although they did not accumulate any observed secondary metabolite, the mycelia of the ΔCTB2 and ΔCTB7
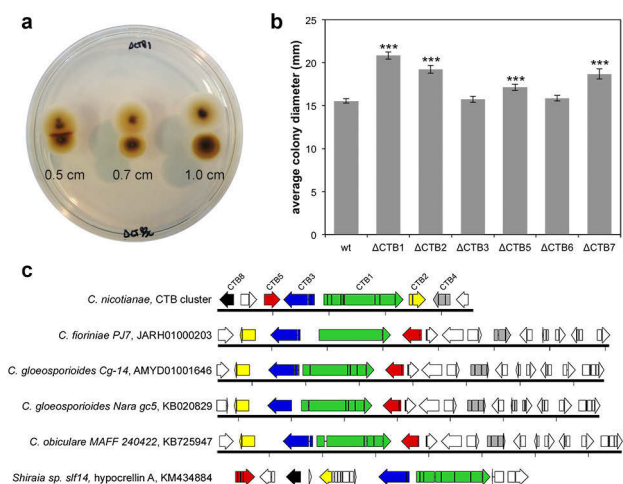
Figure 5.4│Phenotypic and genetic analysis of the CTB cluster. (a) Cercosporin biosynthesis was complemented at the colony-colony interface of the ΔCTB1/ΔCTB3c mutant pair (top/bottom, respectively). The numbers indicate the distance that colonies were inoculated apart from one another in cm. (b) Average colony diameters of CTB disruption mutants are displayed (n = 37). Error bars represent the 95% confidence interval. Stars indicate statistically significant difference from wild-type (p < 0.01). (c) Comparison of gene clusters similar to the CTB gene cluster of *C. nicotianae* (top).

mutants both adopted a yellow-brown color with some export of pigmented compounds into the agar (Figures 5.2c and 5.2g).

In addition to the pigmentation phenotypes, the average colony diameter for each individual strain was distinctive after 7 days of growth (Figure 5.4b). The average colony diameters for the wild type, ΔCTB3c, and ΔCTB6 strains were the smallest at 15.5 ± 0.8, 15.7 ± 1.1, and 15.9 ± 1.0 mm (average ± standard deviation, n = 37). The average colony diameter for the ΔCTB5 mutant was of intermediate size at 17.1 ± 1.1 mm, while the average colony diameters for the ΔCTB1, ΔCTB2, and ΔCTB7 mutants were the largest at 20.8 ± 1.3, 19.2 ± 1.4, and 18.7 ± 1.8 mm. The average colony diameter showed an inverse correlation with metabolite accumulation; the strains with identifiable cercosporin metabolites (wild-type, ΔCTB3c, ΔCTB6, and ΔCTB5) displayed smaller colony diameters while the strains without identifiable intermediates (ΔCTB1, ΔCTB2, and ΔCTB7) displayed larger colony diameters.

5.2.3. Cercosporin Complementation Assay

To test whether accumulated metabolites represented on-pathway intermediates we conducted a complementation assay in which pairs of CTB gene cluster knockout strains were grown adjacent to one another. Cercosporin biosynthetic complementation was indicated by red pigmentation at the colony-colony interface. The only pair that could successfully complement cercosporin biosynthesis was the ΔCTB1 and ΔCTB3c mutant pair (Figure 5.4a). Clear complementation was observed between 4 and 5 days after inoculation between colonies spotted approximately 0.5 cm apart. Pigmentation was isolated to the ΔCTB1 mutant with no clear accumulation of cercosporin in the ΔCTB3c mutant. We attempted to detect cercosporin extracted from agar plugs of the colony-colony interface using a previously described spectrophotometric assay[20]; however, the amount of accumulated cercosporin was below the assay's detection limit. No other mutant pair showed clear cercosporin complementation (Figure F.3, Appendix F).

5.2.4. Genetic Analysis of the Cercosporin Gene Cluster

The antiSMASH algorithm was used to determine gene clusters homologous to the CTB gene cluster.[21] The top five sequences all came from fungal plant pathogens (Figure 5.4c). Of these sequences, only one was associated with a known natural product, the hypocrellin A biosynthetic cluster from *Shiraia* sp. slf14 responsible for the formation of the perylenequinone hypocrellin A (**13**).[22] Five homologous CTB genes were shared among all identified clusters: homologs of CTB1, 3, 2, 5, and 4. The CTB1 homologs contained the canonical NR-PKS domain architecture as determined by a BLASTp search. All of the homologs of CTB3 displayed the same unique didomain architecture

with a predicted N-terminal *O*-methyltransferase and a C-terminal flavin-dependent

monooxygenase.

5.2.5. *In Vitro Analysis of CTB3 Activity*

Predicting that it followed CTB1 on the cercosporin pathway, we investigated the

in vitro activity of CTB3 towards *nor*-toralactone—the *in vitro* product of CTB1.

Because of its unique didomain architecture, we deconstructed CTB3 into its constituent

domains—the N-terminal *O*-methyltransferase (CTB3-MT) and the C-terminal flavin-

dependent monooxygenase (CTB3-MO)—for ease of expression and to investigate each

domain's activity individually (Figure F.1, Appendix F). Both CTB3 and CTB3-MO

were purified as *holo* enzymes with the FAD cofactor bound to the proteins. The

presence of FAD was confirmed by HPLC and spectrophotometric analysis (Figure F.2,

Appendix F). Cosubstrates SAM and NADH were included in reactions of CTB3 as

needed. SAM served as the methyl donor for CTB3-MT and NADH served as the

reductant for CTB3-MO—transferring a hydride equivalent to FAD at the initiation of a

catalytic cycle.

In vitro reactions were conducted with both *nor*-toralactone (**6**) and toralactone

(**7**) as substrates. Initial reactions containing 10 μM CTB3 clearly processed these

substrates; however, the product profiles were ambiguous. Furthermore, both potential

substrates were processed by the CTB3-MO domain with reactions of CTB3-MO and

nor-toralactone being converted to a host of potential products. While these species were

not fully characterized, MS and UV data were indicative of perylenequinone-like

products with the clear loss of a single carbon. In order to simplify the analysis of

enzymatic reactions, significant effort went into optimizing the experimental conditions.

We found that lower protein concentrations (1 µM) and simplified sample preparation (filtration only before HPLC analysis) were necessary for reproducible and robust analysis.

Reactions of CTB3-MT showed turnover of *nor*-toralactone to toralactone (Figure 5.5a). Similarly, reactions containing a low concentration of protein (0.1 µM) showed partial methylation of *nor*-toralactone to toralactone. Together, these results indicated that toralactone served as an intermediate of the CTB3 catalytic cycle and as the true substrate for the CTB3-MO domain despite its apparent activity towards *nor*-toralactone.
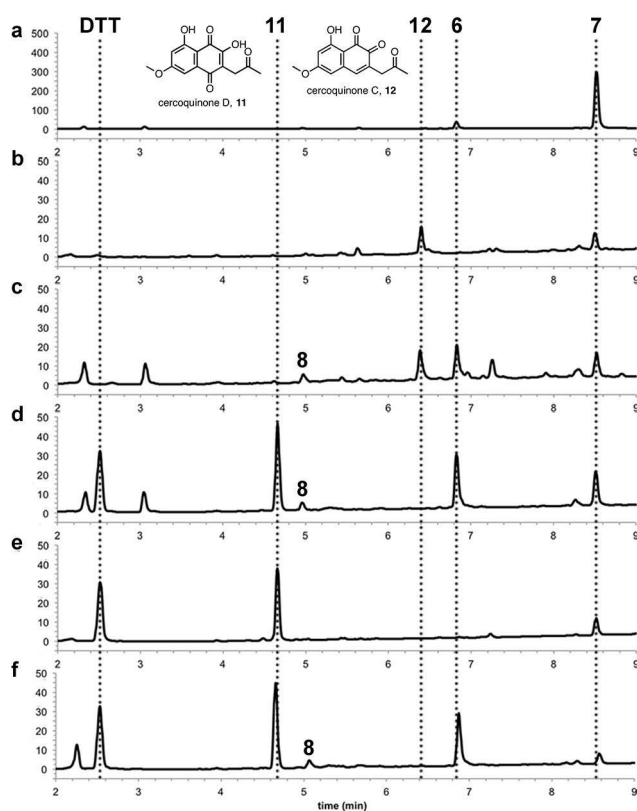
Figure 5.5│Product profiles of *in vitro* reactions of CTB3. The 280 nm chromatograms of the following reactions are displayed: (a) CTB3-MT with *nor*-toralactone, (b) CTB3-MO with toralactone, (c) CTB3-MT and CTB3-MO with *nor*-toralactone, (d) CTB3-MT and CTB3-MO with *nor*-toralactone under reductive conditions, (e) CTB3-MO with toralactone under reductive conditions, and (f) CTB3 with *nor*-toralactone under reductive conditions. Peaks for *nor*-toralactone (**6**) and toralactone (**7**) are indicated along with peaks for products cercoquinone C (**12**) and cercoquinone D (**11**). A peak for DTT and other cosubstrates are observed, as applicable.

Postulating that the programming of the catalytic cycle between the MT and MO domains was in part kinetically controlled, we conducted reactions that were allowed to proceed for 60 min before quenching—significantly longer than the initial 7–10 min reactions. Reactions of the CTB3-MO domain with toralactone serving as the substrate resulted in turnover to *o*-naphthoquinone **12**—cercoquinone C (Figure 5.5b). Cercoquinone C was also observed in reactions of deconstructed CTB3 with *nor*-toralactone as the substrate— albeit with a coterie of other uncharacterized products (Figure 5.5c). These additional products were absent in reactions with toralactone. Additionally, toralactone is an observed intermediate of this reaction along with the off-pathway oxidation product, naphthoquinone **8**.

The formation of cercoquinone C was unexpected. Presumably, this product is formed through its hydroquinone (**14**) even though that species was not directly observed. Assuming that oxidation was spontaneous, we postulated that conducting *in vitro* reactions under reductive conditions could have prevented the presumptive oxidation. Reactions under reductive conditions (1 mM DTT) resulted in the formation of a new *p*-naphthoquinone **11**—cercoquinone D (Figure 5.5d–f). Cercoquinone D was only observed when DTT was present. It was observed in either reactions of deconstructed CTB3 or intact CTB3 with *nor*-toralactone and reactions of the CTB3-MO domain with toralactone. Time-course experiments showed that the appearance of either cercoquinone C under nonreductive conditions and cercoquinone D under reductive conditions was coincident with the consumption of toralactone by CTB3-MO. Attempts to methylate the presumptive product **14** through the action of CTB2, an *O*-methyltransferase, were

unsuccessful under both normal and reductive conditions (Figure F.4, Appendix F).

CTB2 was also unsuccessful at methylating *nor*-toralactone.

Structural assignments of cercoquinones C and D were supported by HRMS, mass fragmentation, and UV spectra. Insufficient amounts of material made structural analysis by NMR for these products unfeasible. Structural analogs of both cercoquinones C and D containing the same quinone cores display similar UV spectra.[23] The fragmentation data were highly characteristic (Figures 5.6 and 5.7). As with cercoquinones A and B, fragmentation ions were observed that comport with the naphthoquinone core. It should be noted that 1,2-naphthoquinones and 1,4-naphthoquinones each fragment in distinct ways.[18c] Additionally, substituents at the C3 position—the 2-oxopropyl moieties in
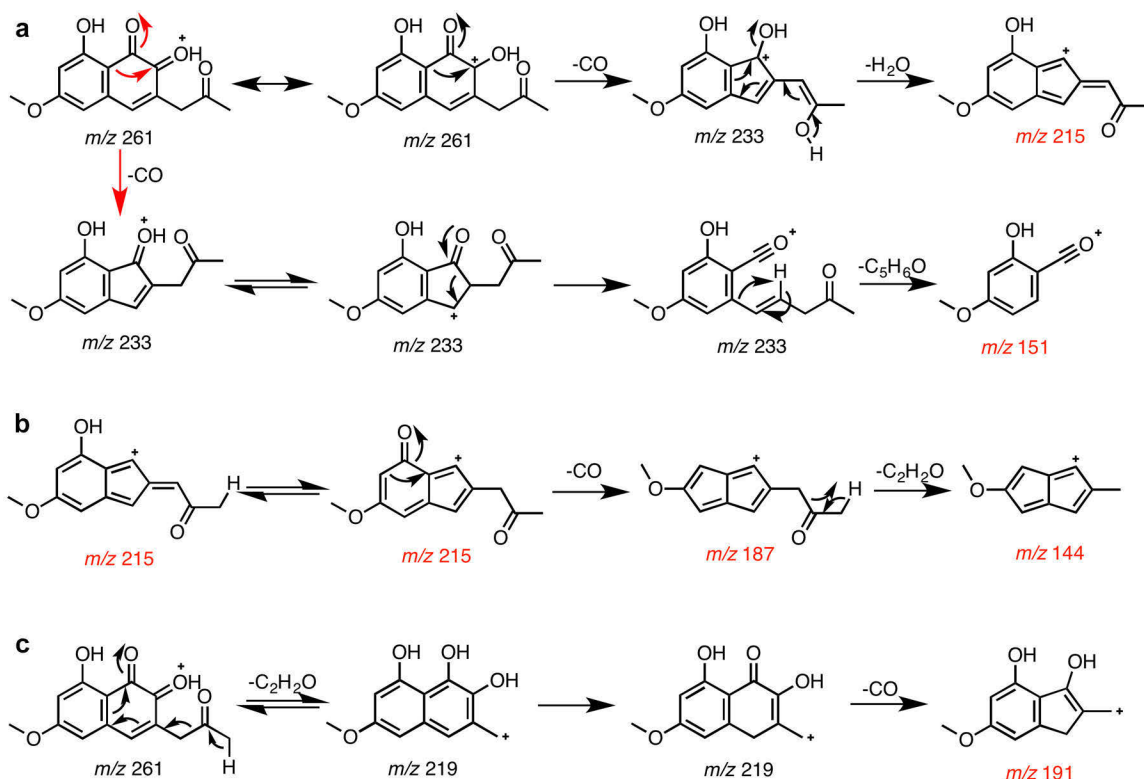
Figure 5.6│Mass fragmentation ions for cercoquinone C justify its structural assignment. (a) Characteristic fragments of 1,2-napthoquinones are observed. (b) The 2-oxopropyl fragmentation ions derived from the *m/z* 215 ion. (c) The *m/z* 191 ion derived from the initial 2-oxopropyl fragment ion. Observed fragmentation ions are in red.

cercoquinones C and D—can influence the fragmentation of the quinone ring.[18]

The cercoquinone C m/z 151 fragmentation ion corresponded to the product of the known fragmentation pattern for 1,2-napthoquinones (Figure 5.6a). Interestingly, this fragmentation pathway established an intermediate that was primed for dehydration resulting in the m/z 215 ion (Figure 5.6b). The m/z 215 ion underwent further fragmentation producing a pair of ions—m/z 187 and 144. The parent cercoquinone C ion was similarly primed for the elimination of the 2-oxopropyl substituent as ketene,
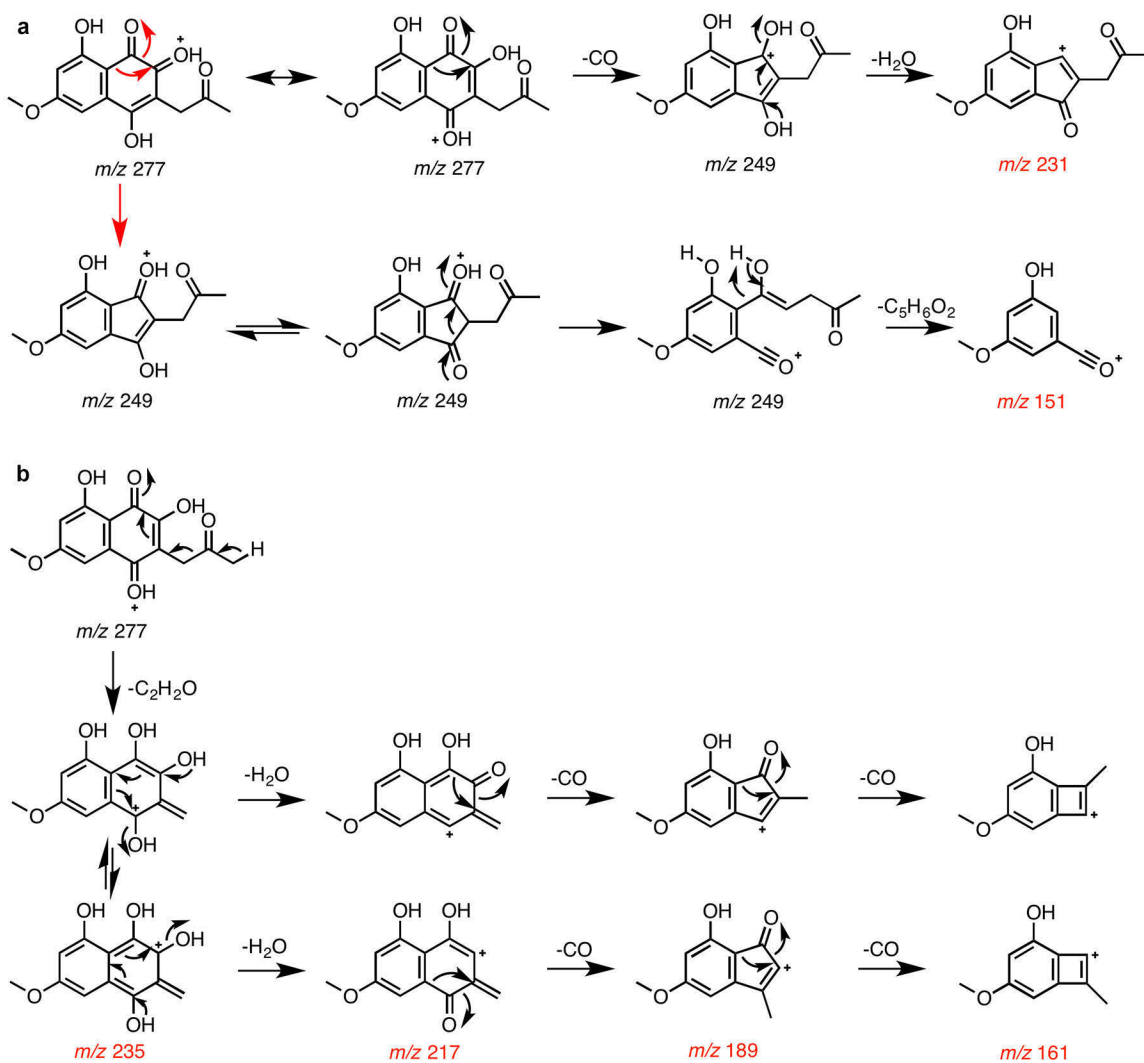
Figure 5.7 | Mass fragmentation ions for cercoquinone D justify its structural assignment. (a) Characteristic fragments of 2-hydroxyl-1,4-napthoquinones are observed. (b) The fragmentation ions derived from the initial 2-oxopropyl fragment ion. Observed fragmentation ions are in red.

142

yielding a resonance-stabilized benzylic cation (*m/z* 219, Figure 5.6c). Although this ion was not observed, its ring-contracted CO-elimination ion was present (*m/z* 191). Additionally, a predominant sodiated ion for cercoquinone C was observed in the HRMS—a common feature of 1,2-napthoquinones.[18c]

The cercoquinone D fragmentation pattern was characteristic for a 2-hydroxyl-1,4-naphthoquinone compound. The distinguishing ring-contracted ion (*m/z* 231) and the oxonium ion (*m/z* 151)—similar to those observed for cercoquinones A—were both evident (Figure 5.7a). As with cercoquinone C, the 2-oxopropyl substituent was fragmented and eliminated as ketene with the participation of the quinone moiety (Figure 5.7b). The resulting resonance-stabilized ion underwent a characteristic series of dehydrations and CO eliminations with ring-contractions generating a collection of distinctive ions (*m/z* 235, 217, 189, and 161).

5.3. *Discussion*

The results presented herein resolve the current ambiguity of cercosporin biosynthesis and suggest a common biosynthetic pathway for the perylenequinone natural products.[12-13] We propose a revised biosynthetic scheme for cercosporin based upon the accumulated metabolites of functional biosynthetic gene knockout strains (Figure 5.8). First, CTB1 acts according to its established *in vitro* chemistry producing *nor*-toralactone (**6**).[12] The bifunctional enzyme CTB3 methylates *nor*-toralactone to toralactone (**7**) before conducting an unusual oxidative aromatic ring opening producing metabolite **14**. The *O*-methyltransferase CTB2 is proposed to methylate the nascent OH-6 of intermediate **14**—blocking further oxidation at this site and yielding compound **15**— before the reductase CTB6 reduces the 2-oxopropyl ketone at position C7, giving
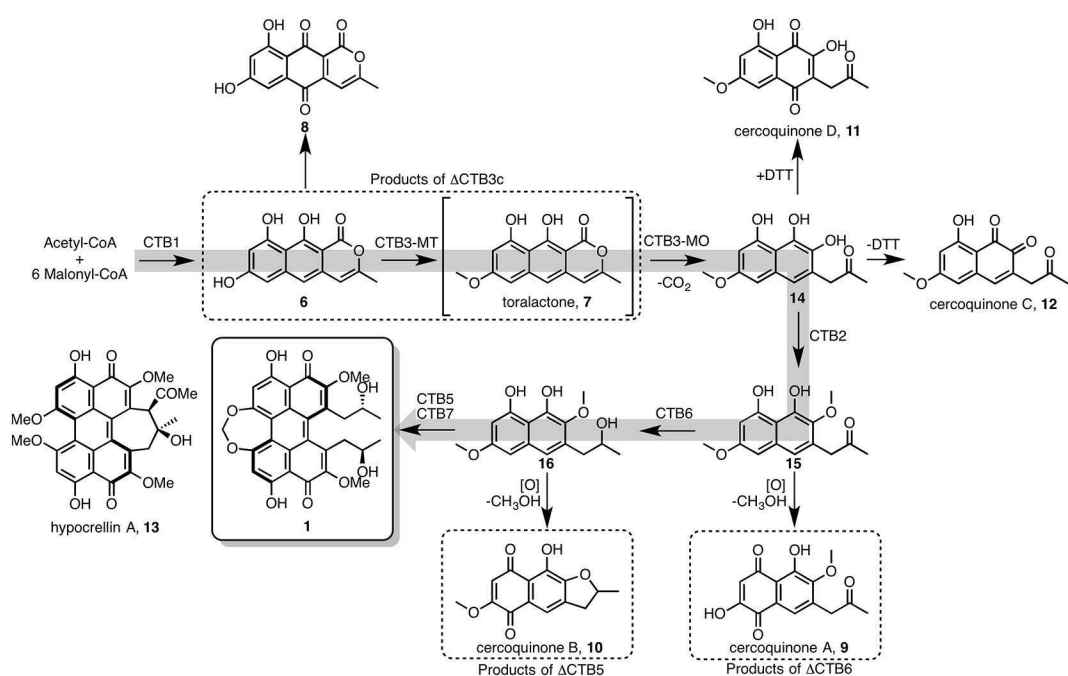
Figure 5.8│Proposed cercosporin biosynthetic pathway. On the strength of observed pathway intermediates, *in vitro* chemistry, phenotypic, genetic, and pairwise complementation a new biosynthetic scheme for cercosporin is presented.

naphthalene **16**. CTB5 is thought to be responsible for homodimerization of intermediate **16** with CTB7 installing the dioxepine moiety, finally producing cercosporin (**1**).

We suggest that several of the accumulated metabolites correspond to oxidation products of true on-pathway intermediates. The metabolites cercoquinone A (**9**) and cercoquinone B (**10**) from the ΔCTB6 and ΔCTB5 mutants, respectively, fall into this category. Both quinone products derive from the oxidation of electron-rich naphthalene derivatives. The oxidation of electron-rich aromatic metabolites is common and spontaneous—often leading to stabilized products.[24] ^{1}H NMR and mass fragmentation data comport with these structural assignments, especially with respect to the sites of oxidation and methylation. In fact, the observed fragmentation ions agreed with established fragmentation mechanisms for naphthoquinone products. The inability of

either ΔCTB6 or ΔCTB5 mutants to complement cercosporin activity as secretor-converter pairs bolsters the argument that cercoquinones A and B are rapidly formed off-pathway byproducts.

Interestingly, cercoquinone A displays the loss of a methyl group installed earlier through the activity of the *O*-methyltransferase domain of CTB3. The loss of the methyl at this position is unambiguous given the metabolite's characteristic mass fragmentation pattern; in particular, the retention of the methoxyl in both the ring contraction carbocation and the ring-opened oxonium ion confirms its structural assignment. We envision two possible scenarios that could account for the elimination of this moiety (Figure 5.9). In the first possibility, it is simply eliminated as methanol following oxidation to the *p*-quinone. Several structural features would favor addition of water at the methoxy position. The site of attack is vinylogous to the ketone of the quinone, which is—in turn—stabilized by a hydrogen bond to the *peri*-hydroxyl group. Collapse of the enolate intermediate would favor elimination of methanol resulting in cercoquinone A. In the second possibility, the methyl position is removed enzymatically. The methoxy at this position is eventually incorporated into the dioxepine functionality of the final product through the activity of CTB7. While the mechanism of this transformation is unknown,
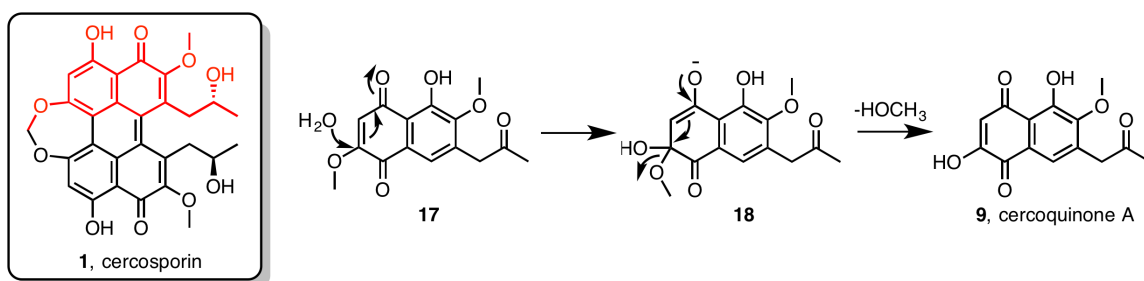
Figure 5.9 │ Proposed formation of cercoquinone A. Two mechanisms were considered. Addition of water followed by elimination of methanol (pictured) or enzymatic demethylation. Cercosporin is shown with a monomeric unit shown in red. CTB7 is postulated to form the dioxepine ring, a transformation that would require the elimination of a methyl group at the position of demethylation in cercoquinone A.

by analogy to studies of methylenedioxy formations,[25] it would likely involve oxidative loss. Given the structural similarity of cercoquinone A to the monomeric unit of cercosporin, we suggest that CTB7 could use this metabolite as an alternative substrate and catalyze the elimination of the methyl in question. Cercoquinone B stands in opposition to both of these mechanisms. Cercoquinone B possesses the same chemical features that favor the addition-elimination mechanism; yet, it retains the methoxyl at the comparable position. Similarly, it bears structural resemblance to the cercosporin monomeric unit and could be envisioned as an alternative CTB7 substrate. Unique to cercoquinone B is a dihydrofuran moiety on the opposite side of the molecule. Perhaps the presence of this functionality precludes it from serving as a substrate for CTB7—a claim that would seem to bolster the enzymatic elimination mechanism for cercoquinone A. We argue that the dihydrofuran moiety of cercoquinone B arises due to intramolecular nucleophilic aromatic substitution with elimination of methanol after oxidation to the quinone.

The absence of any cercosporin pathway metabolites in the ΔCTB2 and ΔCTB7 mutants is curious given that they are absolutely necessary to sustain cercosporin biosynthesis. Although we do not have direct evidence for degradation, we believe the presumptive metabolites of these knockout mutants are directed towards divergent pathways. Naphthoquinone secondary metabolites in fungi have been shown to inhibit fungal growth.[26] Both ΔCTB2 and ΔCTB7 mutant strains have more robust growth than either wild-type or mutant strains that accumulate significant amounts of naphthoquinone byproducts, suggesting that, indeed, the presumptive cercosporin pathway intermediates of these mutant strains are consumed. Despite their limited metabolic profiles, both

ΔCTB2 and ΔCTB7 mutant strains appear pigmented while the ΔCTB1 mutant strain is colorless. The putative cercosporin pathway intermediates produced in these mutant strains would likely be electron-rich, hydroxylated naphthalene analogs similar to the components of melanin.[27] Melanin—a crucial biological pigment—is produced in fungi by the polymerization of polyketide-derived 1,8-dihydroxynaphthalene monomers. It is a critical metabolite in plant pathogens as it provides mechanical strength to cellular structures responsible for host invasion.[28] Furthermore, melanin is highly insoluble and would not contribute to the extracted metabolite profiles analyzed in the current study. Together, the data suggest that the metabolites of the ΔCTB2 and ΔCTB7 mutants are likely directed towards a melanin-like biosynthetic pathway resulting in incorporation into insoluble, pigmented polymers.

As has been shown previously, the *in vitro* product of CTB1—an NR-PKS—is naphthopyrone *nor*-toralactone (**6**).[12] The evolution of *nor*-toralactone by the NR-PKS presents an immediate problem. In order to synthesize cercosporin from *nor*-toralactone, the pyrone ring must be opened; however, there is no apparent enzyme in the CTB cluster that could carry out this transformation through conventional lactonase-like hydrolysis— the common pathway for this type of reaction.[29] The accumulation of *nor*-toralactone in the ΔCTB3c mutant simultaneously confirms the importance of this metabolite while implicating a potential candidate in CTB3 for pyrone ring opening. Furthermore, the successful complementation of cercosporin biosynthesis in the ΔCTB1/ΔCTB3c mutant pair corroborates that *nor*-toralactone is an on-pathway intermediate. Nevertheless, it remains unclear why the pathway would proceed through a naphthopyrone intermediate at all. Given that downstream intermediates of the pathway are both prone to oxidation

and putatively redirection towards other biosynthetic pathways. Perhaps production of the relative stable intermediate *nor*-toralactone ensures sufficient flux into the cercosporin pathway at the outset.

CTB3 is an unusual didomain enzyme with an *O*-methyltransferase domain and a flavin-dependent monooxygenase domain. The *in vitro* enzymatic reactions presented herein confirm that both domains work in tandem to transform the polyketide product *nor*-toralactone to the on-pathway naphthalene analog **14**. Although naphthalene **14** is not directly observed, it is the implied shared precursor to both cercoquinone C (**12**) and cercoquinone D (**11**). The mass fragmentation and spectral data are in accord with the structural assignments for cercoquinones C and D. We observe characteristic mass fragments for both *ortho* and *para* quinones that confirm the positions of oxidation. The observed ring-contracted carbocations and ring-opened oxonium ions for both metabolites are the expected products of established fragmentation pathways.

The direct methylation of *nor*-toralactone to toralactone by CTB3-MT, as well as the appearance of toralactone in reactions of CTB3 with *nor*-toralactone are strong indicators that toralactone is a true intermediate of the CTB3 catalytic cycle. Furthermore, its appearance in the metabolite profile of the ΔCTB3c mutant suggests its biological importance. The ΔCTB3c mutant strain was produced by inserting the knockout cassette into the 5′-terminus of the CTB3 gene corresponding to the CTB3-MO domain. The sequence of the CTB3-MT domain is largely unaltered suggesting the possibility of a functional *O*-methyltransferase. We suspect that the ΔCTB3c mutant strain could retain some endogenous CTB3-MT activity resulting in the partial methylation of *nor*-toralactone in the disruption mutant strain. Alternatively, another *O*-

methyltransferase could conceivably convert *nor*-toralactone to toralactone; however, this

hypothetical *O*-methyltransferase could not be CTB2—the only other methyltransferase

in the CTB cluster—as it is incapable of this transformation *in vitro*.

The CTB3-MO domain shares strong primary sequence homology to the

canonical *p*-hydroxybenzoate (**17**) hydrolase family of flavin-dependent

monooxygenases.[30] This is a well-studied family of enzymes with a shared catalytic

mechanism and preference for electron-rich aromatic substrates. They are responsible for

the hydroxylation of simple phenolic systems with a preference for the *ortho* and *para*

positions. The reaction cycle proceeds in two halves: reductive and oxidative (Figure

5.10a).[31] The reductive half cycle is initiated with the formation of a ternary complex of

enzyme containing oxidized FAD (FAD_{ox}, **18**), NAD(P)H, and substrate. Upon formation
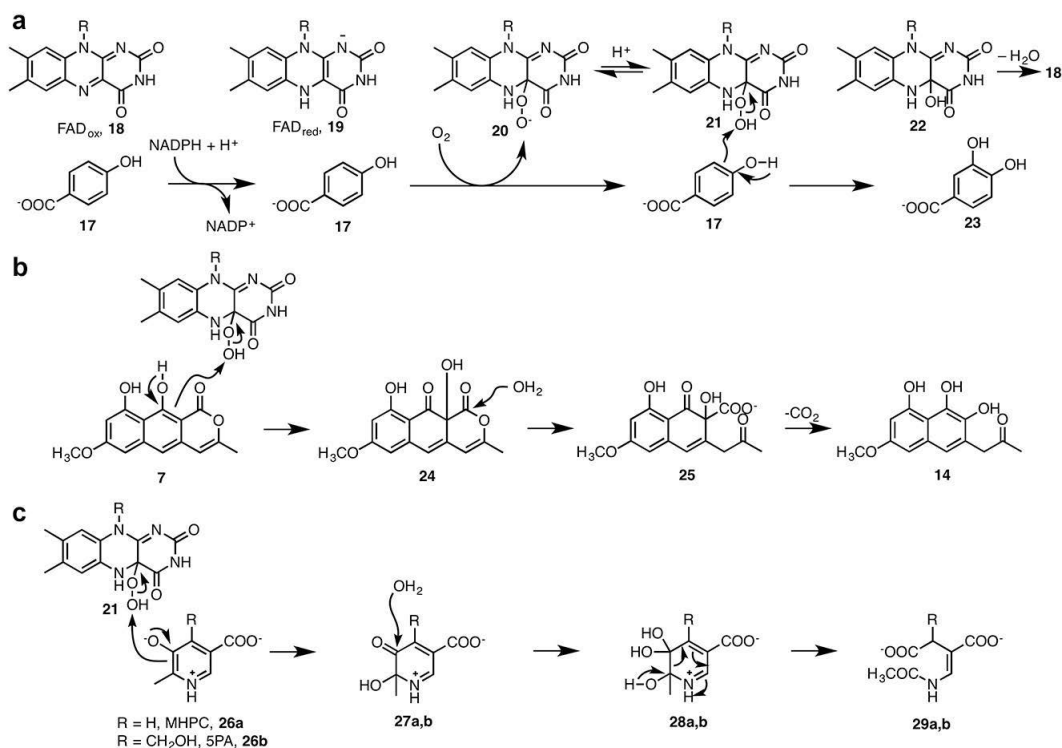
Figure 5.10 │ Proposed mechanism for CTB3 flavin-dependent monooxygenase domain. (a) General mechanism of *p*-hydroxybenzoate (**17**) hydroxylase family members. (b) Proposed mechanism for CTB3 catalyzed oxidative aromatic ring cleavage of toralactone (**7**). (c) Mechanism of MHPCO and 5PAO oxidative aromatic ring cleavage.

of the ternary complex, rapid reduction of FAD by NAD(P)H occurs (FAD$_{red}$, **19**). Reduction is contingent on substrate binding or—in some cases—analogs of the native substrate. NAD(P)$^+$ dissociation occurs at a rate similar to that of reduction. The resulting reduced enzyme and substrate complex has significant kinetic stability. In the oxidative half cycle, oxygen reacts with FAD$_{red}$, forming the characteristic flavin-C4a-hydroperoxide reactive intermediate (**21**). The deprotonated flavin-C4a-peroxy species (**20**) has never been observed in these enzymes. Hydroxylation of the substrate occurs through electrophilic aromatic substitution resulting in an enzyme bound product and flavin-C4a-hydroxide (**22**) complex. Water is eliminated from the flavin species resulting in FAD$_{ox}$ and the product is finally released.

We argue that the same mechanism is at play in the CTB3-MO domain with the enzyme installing hydroxide at the bridgehead position of toralactone (**7**, Figure 5.10b). The hydroxylated intermediate **24** would be far more susceptible to hydrolysis due to the loss of conjugation and aromaticity and the favorable enolate leaving group. Hydrolysis—enzyme catalyzed or spontaneous—would result in acid **25**. Loss of carbon dioxide from acid **25** to generate the *o*-hydroquinone **14** would likely be spontaneous and rapid with the reestablishment of aromaticity providing a large thermodynamic driving force. Alternatively, one could also consider a Baeyer-Villiger-like oxidation, with a flavin-C4a-peroxy nucleophile attacking the pyrone carbonyl. Ring expansion followed by hydrolysis and loss of carbon dioxide would generate the same product **14**. We consider this mechanism to be unlikely. The *p*-hydroxybenzoate monooxygenase enzyme family invariably proceeds through an electrophilic flavin-C4a-hydroperoxy species

while flavoenzyme catalyzed Baeyer-Villiger oxidations rely on the nucleophilic flavin-C4a-peroxy species.[32]

There is precedence for *p*-hydroxybenzoate monooxygenase family members participating in oxidative aromatic ring cleavage. The enzymes 2-methyl-3-hydroxypyridine-5-carboxylic acid (MHPC, **26a**) oxygenase (MHPCO) and 5-pyridoxic acid (5PA, **26b**) oxygenase (5PAO) are flavoenzymes of the *p*-hydroxybenzoate monooxygenase family catalyzing aromatic hydroxylation and subsequent aromatic ring cleavage reactions (Figure 5.10c).[33] Like conventional flavin monooxygenases, these enzymes use the flavin-C4a-hydroperoxy reactive intermediate (**21**) in their initial electrophilic aromatic oxidation; however, they undergo a subsequent ring-opening reaction. The mechanism of ring opening has not been definitively established, but it is known that ring opening is enzyme catalyzed in MHPCO and involves the addition of water to the carbonyl intermediate (**27a**).[34]

Although compound **14** is not observed *in vitro*, it is the implied common intermediate of both cercosporin C and D (**12** and **11**, respectively). The transfer of a hydride equivalent from compound **14** to FAD_{ox}, while not anticipated, is not entirely unreasonable. In the absence of a reductive environment, one could envision this oxidation is the dominating pathway towards cercoquinone C. Furthermore, the observed consumption of *nor*-toralactone by CTB3-MO implies an *ortho*-quinone intermediate. *Ortho*-quinone analogs of cercoquinone C have been used previously to access the perylenequinone core through simple coupling under acid conditions.[35] Assuming *nor*-toralactone serves as an alternative substrate for CTB3-MO, the observed but uncharacterized dimeric products could likely arise through analogous couplings.
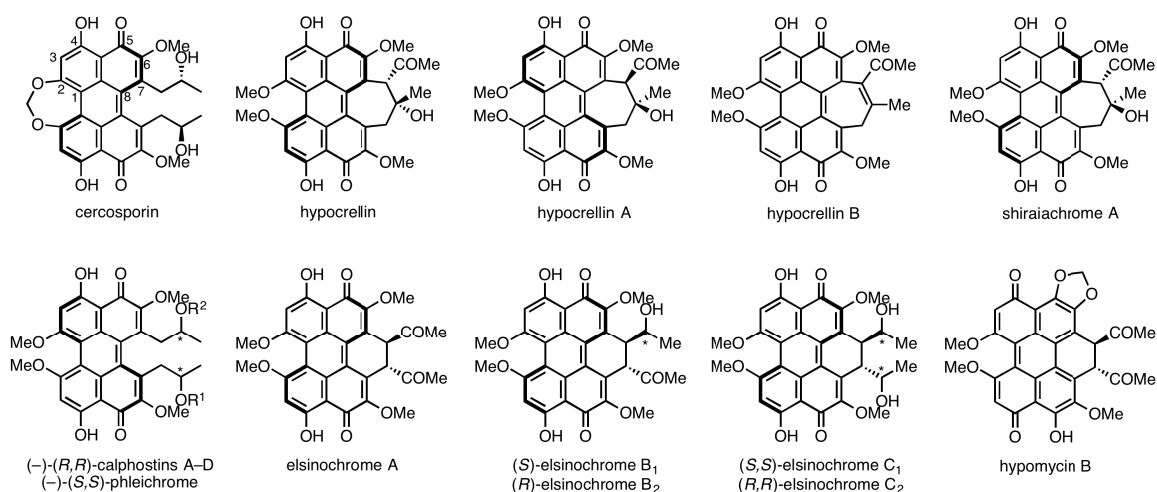
Figure 5.11 │ Structures of known fungal perylenequinone natural products. Common architectural features are observed in the family. The occurrence of methyoxy at 2, 2′ and 6, 6′ positions are invariable. As is the 2-oxypropyl derivatives at 7, 7′. The formation of the perylenequinone core also appears to be conserved.

Under reductive conditions, cercoquinone C—upon release—would be rapidly reduced back to species **14**. Electron rich naphthalene species such as compound **14** are susceptible to spontaneous oxidation. One could envision the oxidation of compound **14** to cercoquinone D as being rapid and complete, even under reductive conditions. The apparent instability of intermediate **14** could explain the observed preference for oxidation of *ortho*-quinone cercosporin C. In the quinone oxidation state, this species is protected from subsequent oxidation, ensuring biosynthetic fidelity.

Structural analysis of known fungal perylenequinone natural products reveals several common features among this family of metabolites (Figure 5.11).[35c] First, the 7, 7′ moieties are derived from 2-oxypropyl side chains. Second, the 2, 2′ substituents are always methoxyl or—in the case of cercosporin—a derivative thereof. Third, the 6, 6′ substituents are always methyoxy or derivatives thereof. Fourth, the perylenequinone core architecture remains unaltered. These core structural features suggest a common biosynthetic pathway for the perylenequinone metabolites. Indeed, the activities of

152

CTB1, CTB3, CTB2, and CTB5 delineated here would account for each of these features with CTB1 producing the common intermediate *nor*-toralactone, CTB3 methylating this intermediate at the OH-2 and opening the pyrone ring thereby installing a C6-OH, CTB2 methylating the nascent C6-OH, and CTB5 catalyzing dimerization yielding the perylenequinone carbon core. Interestingly, homologs of each of these enzymes are found in known gene clusters most similar to the CTB cluster. Of these clusters, only one is linked to an observed product, hypocrellin A (**13**).[22] Hypocrellin A differs from cercosporin in three key features: the 7, 7′ substituents are retained in the 2-oxypropyl oxidation state and are linked through an intramolecular aldol, the 2, 2′ substituents are methoxyl, and the compound adopts the alternative atropisomeric conformation. Comparing the CTB and hypocrellin A clusters, CTB6 and CTB7 appear unique to the CTB cluster. Given that these enzymes presumptively reduce the 2-oxypropyl substituents and install the dioxepine moiety, respectively, it makes sense that they would be missing in the hypocrellin A cluster, where these features are not present. Altogether, these data imply that perylenequinone natural products proceed through a common biosynthetic pathway in which *nor*-toralactone is initially produced and further processed by a CTB3 didomain homolog. Given its unique activity and rarity, CTB3 and its homologs make appealing targets for antifungal agents.

5.4. *Conclusion*

Here we resolve the current ambiguity of cercosporin biosynthesis and propose our own biosynthetic pathway. We demonstrate that the naphthoquinone *nor*-toralactone is a biologically relevant intermediate of the cercosporin pathway, further corroborating the previously observed *in vitro* activity of the NR-PKS CTB1. We also characterize the

activity of the unusual didomain protein CTB3, showing that its flavin-dependent monooxygenase domain is responsible for a unique oxidative aromatic cleavage. Together, these findings further not only our understanding of perylenequinone biosynthesis but also expand the known biochemical repertoire of flavin-dependent monooxygenases—an important class of enzymes in both primary and secondary metabolism. While addressing the initial steps of cercosporin biosynthesis, the data presented do not fully resolve the later portion of the biosynthetic pathway. The crucial perylenequinone dimerization step remains unexamined biochemically and enzymatically. Given the importance of cercosporin for pathogenicity, a complete understanding of its biosynthesis is of utmost importance. Indeed, the rare activity of CTB3 and its apparent conservation among perylenequinone biosynthetic gene clusters makes this enzyme an attractive target for *Cercospora* specific antifungal agents.

5.5. Experimental Methods

5.5.1. Reagents and Biological Strains

All reagents and primers were purchased from Sigma-Aldrich (St. Louis, MO, USA), unless otherwise specified. Magnesium chloride, sodium chloride, potassium phosphate monobasic, potassium phosphate dibasic, and deoxynucleotides (dNTPs) were purchased from ThermoFisher Scientific (Waltham, MA, USA). Isopropyl-β-D-thiogalactoside (IPTG), kanamycin, and nickel (high density) agarose beads were purchased from Gold Biotechnology (St. Louis, MO, USA). Imidazole was purchased from Acros Organics (Geel, Belgium). Difco™ potato dextrose agar (PDA) and Difco™ potato dextrose broth (PDB) were purchased from Beckton, Dickson, and Company (Franklin Lakes, NJ, USA). Yeast extract and tryptone were purchased from Boston

BioProducts (Ashland, MA, USA). Agar was purchased from bioWORLD (Dublin, OH, USA). Plasmids pET-24a(+) and pET-28a(+) were purchased from EMD Millipore (Darmstadt, Germany). All enzymes used for manipulation of DNA and 10 × HF Buffer were purchased from New England Biolabs (Ipswich, MA, USA). Bio-Rad protein assay dye for determining protein concentration by the Bradford assay was purchased from Bio-Rad laboratories (Hercules, CA, USA). Bacterial strain *Escherichia coli* BL21(DE3) was purchased from Sigma-Aldrich and made electrocompetent by standard protocols (Appendix B). DNA purified from *C. nicotianae* (ATCC[©] 18366[™]) was provided by Dr. Kuang-Ren Chung (currently of National Chung Hsing University, Taichung, Taiwan). *C. nicotianae* (ATCC[©] 18366[™]) wild type strain and cercosporin gene cluster knockout strains previously prepared were also generously provided by Dr. Chung.[9, 13-14]

5.5.2. Culture Conditions

C. nicotianae (ATCC[©] 18366[™]) and knockout strains were maintained on PDA. PDA plates were strictly maintained at a thickness of 3 mm. Starter cultures were prepared by spotting a colony from a stock plate on PDA with a sterile toothpick. The starter culture was allowed to grow, inverted, at 28 °C for 7 days under constant fluorescent light. Stock plates were prepared from fresh starter cultures that were wrapped in parafilm and stored at 4 °C in the dark until further use. Production cultures were prepared by spotting colonies from a fresh starter culture on PDA with a sterile toothpick. Colonies were spotted in a 2 cm grid. Mycelia were incubated, inverted, under constant fluorescent light at 28 °C for 7 days.

5.5.3. *Metabolite Extraction and Purification*

The total mass (mycelia and agar) from production cultures were collected, frozen in liquid nitrogen, and ground to a powder in a mortar and pestle. The powder was lyophilized for 12–16 h then suspended in water acidified with concentrated HCl. Metabolites were extracted from this slurry thrice with ethyl acetate at 4 °C. The ethyl acetate fractions were combined, washed with brine, dried with sodium sulfate, and evaporated under vacuum. The solid residue was stored at 4 °C until further use.

The extracted metabolites were dissolved in methanol and filtered through 0.4 μm PTFE filters. Metabolites were separated on a Sephadex LH-20 column (GE Healthcare, Little Chalfont, UK) using methanol as a mobile phase with a flow rate of 1.3 mL/min and injecting 1.2 mL aliquots. Fractions were collected every 5 min. Fractions were screened for metabolites of interest by reverse phase HPLC on an Agilent 1200 (Agilent Technologies, Santa Clara, CA, USA). Solvent A was water + 0.1% formic acid. Solvent B was acetonitrile + 0.1% formic acid. Fractions were injected onto a linear gradient of 5–95% solvent B over 10.8 min at 1.25 mL/min on a Kinetex XB-C18 column (4.6 mm × 75 mm, 2.6 μm, Phenomenex, Torrence, CA, USA). Typical injection volumes were 10–15 μL. Chromatograms at 250 and 280 nm were recorded.

Fractions containing metabolites of interest were combined, concentrated under vacuum, and purified by reverse phase HPLC on an Agilent 1100 instrument (Agilent Technologies). Solvent A was water + 0.1% formic acid. Solvent B was acetonitrile + 0.1% formic acid. Fractions were injected onto a linear gradient of 5–95% solvent B over 45 min at 4.726 mL/min on a Prodigy 5u ODS3 column (10 mm × 250 mm, 5 μm, Phenomenex). Injection volumes were typically 500–1000 μL. Chromatograms at the

expected λ_{max} for the metabolite of interest were recorded. Metabolites were manually collected as they eluted and the fractions from serial injections were combined. Acetonitrile and formic acid were removed from these samples by evaporation under vacuum in a rotary evaporator. The remaining aqueous solution was frozen in liquid nitrogen and lyophilized until dry. Purified samples were stored at 4 °C until further use.

5.5.4. Complementation Assays

Cercosporin biosynthesis functional knockouts were plated adjacent to one another to assay for cercosporin production complementation. Colonies were spotted with a sterile toothpick from a fresh starter culture on 3 mm thick PDA plates. The complete set of pairwise combinations of cercosporin biosynthetic functional knockouts was assayed. Colonies from different strains were spotted 0.5, 0.7, 1.0, 1.5, and 2.0 cm apart from one another. The cultures were incubated, inverted, under constant fluorescent light at 28 °C for 7 days. The appearance of red pigmentation at the boundary between two mycelia was determined to represent cercosporin production complementation. Evidence of complementation was monitored daily.

5.5.5. Preparation of Expression Constructs

DNA manipulations were carried out in accordance with established procedures.[36] Protocols for PCR and overlap extension PCR are found in Appendix B. Details of the expression plasmids used in this study are summarized in Table F.2 (Appendix F). Cut sites for protein deconstruction were guided by a variety of bioinformatics analyses including multiple sequence alignment, secondary structure prediction and the UMA algorithm for predicting interdomain regions.[37] Primer sequences used for cloning new constructs used in this study are presented in Tables F.1

(Appendix F). All expression plasmids were maintained in *E. coli* BL21(DE3) cells stored in 20% glycerol at −80 °C.

The complete CTB3 ORF was cloned into a pCR®-Blunt vector (ThermoFisher Scientific) for easier genetic manipulation. The insert was prepared by PCR using *C. nicotianae* (ATCC© 18366™) genomic DNA template with primers CTB3-5 and CTB3-3. The product was ligated into pCR®-Blunt using the Zero Blunt® PCR Cloning Kit (ThermoFisher Scientific) according to the manufacturer's instructions, yielding pCR-CTB3.

Two full-length CTB3 expression constructs were prepared. The first had a C-terminal 6 × His tag (pECTB3) and the second had an N-terminal 6 × His tag (p28CTB3). The CTB3 ORF contains three exons that were synthetically spliced using overlap extension PCR. Overlap extension PCR DNA fragments were prepared by PCR from pCR-CTB3 template. The exon 1 fragment was prepared with primers CTB3-AseI-5 and CTB3-ex1-3. The exon 2 fragment was prepared with primers CTB3-ex2-5 and CTB3-ex2-3. Two exon 3 fragments were prepared with primers CTB3-ex3-5 and CTB3-NotI-3 or CTB3-Stop-3, for fragments without and with a stop codon, respectively. The resulting PCR products were spliced together using overlap extension PCR with outside primers CTB3-AseI-5 and CTB3-NotI-3 or CTB3-Stop-3 for inserts without and with a stop codon, respectively. The inserts were digested with AseI and NotI-HF. The insert without a stop codon was ligated in NdeI and NotI-HF digested pET-24a(+) yielding pECTB3. The insert with a stop codon was ligated in NdeI and NotI-HF digested pET-28a(+) yielding p28CTB3.

Two CTB3 *O*-methyltransferase monodomain expression constructs were prepared by PCR. The first had a C-terminal 6 × His tag (pECTB3-MT1) and the second had an N-terminal 6 × His tag (p28CTB3-MT1). The inserts were prepared by PCR using pECTB3 template with primers CTB3-AseI-5 and CTB3-MT1-3 or CTB3-MT1-Stop-3 for inserts without or with a stop codon, respectively. The inserts were digested with AseI and NotI-HF. The insert without a stop codon was ligated into NdeI and NotI-HF digested pET-24a(+) yielding pECTB3-MT1. The insert with a stop codon was ligated into NdeI and NotI-HF digested pET-28a(+) yielding p28CTB3-MT1.

Two CTB3 flavin-dependent monooxygenase monodomain expression constructs were prepared by PCR. The first had a C-terminal 6 × His tag (pECTB3-MO3) and the second had an N-terminal 6 × His tag (p28CTB3-MO3). The inserts were prepared by PCR using pECTB3 template with primers CTB3-MO3-5 and CTB3-NotI-3 or CTB3-Stop-3 for inserts without or with a stop codon, respectively. The inserts were digested with NdeI and NotI-HF. The insert without a stop codon was ligated into NdeI and NotI-HF digested pET-24a(+) yielding pECTB3-MO3. The insert with a stop codon was ligated into NdeI and NotI-HF digested pET-28a(+) yielding p28CTB3-MO3.

The complete CTB2 ORF was cloned into a pCR®-Blunt vector for easier genetic manipulation. The insert was prepared by PCR using *C. nicotianae* (ATCC© 18366™) genomic DNA template with primers CTB2-5 and CTB2-3. The product was ligated into pCR®-Blunt using the Zero Blunt® PCR Cloning Kit according to the manufacturer's instructions, yielding pCR-CTB2.

Two CTB2 expression constructs were prepared. The first had a C-terminal 6 × His tag (pECTB2) and the second had an N-terminal 6 × His tag (p28CTB2). The CTB2

ORF contains two exons that were synthetically spliced using overlap extension PCR. Overlap extension PCR DNA fragments were prepared by PCR from pCR-CTB2 template. The exon 1 fragment was prepared with primers CTB2-NdeI-5 and CTB2-ex1-3. Two exon 2 fragments were prepared with primers CTB2-ex2-5 and CTB2-NotI-3 or CTB2-Stop-3, for fragments without and with a stop codon, respectively. The resulting PCR products were spliced together using overlap extension PCR with outside primers CTB2-NdeI-5 and CTB2-NotI-3 or CTB2-Stop-3 for inserts without and with a stop codon, respectively. The inserts were digested with NdeI and NotI-HF. The insert without a stop codon was ligated in NdeI and NotI-HF digested pET-24a(+) yielding pECTB2. The insert with a stop codon was ligated in NdeI and NotI-HF digested pET-28a(+) yielding p28CTB2.

5.5.6. *Protein Expression and Purification*

Proteins were prepared by heterologous expression in *E. coli* BL21(DE3) and purified by nickel-affinity purification as delineated in Appendix B. Cultures used for the expression of protein constructs containing the CTB3 flavin-dependent monooxygenase domain were supplemented with 1 μM riboflavin added immediately prior to inoculation. Eluted protein fractions were dialyzed at 4 °C overnight with three buffer switches into 100 mM potassium phosphate pH 7.0, 10% glycerol. Purified protein concentrations were determined by the Bradford assay using bovine serum albumin (New England Biolabs) as a standard as described in Appendix B.[38] Excess protein was flash frozen in liquid nitrogen and stored at −80 °C until further use.

The CTB3-MO monodomain was assayed for flavin identity and content by cofactor release through heat denaturation.[39] A 10 μM solution of CTB3-MO was

incubated in a boiling water bath for 10 min. The heat-denatured sample was cooled on ice, centrifuged (14500 × g at 4 °C for 10 min), and the supernatant was retained (containing free cofactor). The supernatant was also analyzed by HPLC on an Agilent 1200 fitted with a Prodigy 5u ODS3 column (4.6 × 250 mm, 5 μm, Phenomenex) using a linear gradient of 85–25% solvent A over 20 min at a flow rate of 1.0 mL/min. Solvent A was 5 mM ammonium acetate pH 6.5. Solvent B was methanol. The elution profile of released cofactor was compared to that of an authentic standard of FAD.

5.5.7. In Vitro Reactions of CTB3

Nor-toralactone and toralactone purified from the ΔCTB3c mutant strain were used as substrates for *in vitro* reactions of CTB3. Stock solutions of the substrates (2 mM) were prepared in 90% DMSO [*aq.*] and stored at −80 °C until use. NADH stock solutions (50 mM) were prepared immediately before use in 100 mM potassium phosphate pH 7.0, 10% glycerol. DTT and SAM stock solutions (100 mM and 50 mM, respectively) were prepared in 100 mM potassium phosphate pH 7.0, 10% glycerol and stored at −20 °C until use. Reactions were prepared in potassium phosphate pH 7.0, 10% glycerol. Reactions were prepared by diluting CTB3 (intact, CTB3-MT, CTB3-MO, or CTB2, 1.0 μM final concentration) in buffer followed by the addition of cosubstrates (NADH or SAM, as needed; 1 mM final concentration), followed by the addition of DTT, as needed (1 mM final concentration). Reactions were initiated with the addition of naphthopyrone substrate (0.1 mM final concentration, 4.5% DMSO final concentration) and immediately mixed by gentle agitation. Reactions were allowed to proceed at room temperature for 1 h, and then were quenched with the addition of 5 *N* HCl (5 μL to 250 μL reaction). The reaction mixtures were filtered through 0.2 μm PTFE filters and

analyzed by HPLC. HPLC was conducted on an Agilent 1200 system fitted with a diode array detector (DAD) and a Kinetex XB-C18 column (4.6 × 75 mm, 2.6 μm). The gradient method was as follows: 5–95% acetonitrile + 0.1% formic acid, over 10.8 min, at a flow rate of 1.25 mL/min.

5.5.8. *Chemical Characterization of Cercosporin Pathway Products*

^{1}H NMR data were collected on a 400 MHz Bruker Avance spectrometer (Bruker, Billerica, MA, USA). HRMS data were collected on a Waters Acquity/Xevo-G2 UPLC-MS system (Waters, Milford, MA, USA). MS and MS/MS spectra were collected in positive ion mode. HPLC was conducted on an Agilent 1200 system fitted with a DAD and a Kinetex XB-C18 column (4.6 × 75 mm, 2.6 μm). UV data were determined from the DAD signal at the maximum HPLC chromatograph signal. Spectra are presented in Appendix F (Figures F.5–18).

5,12-dihydroxy-8,9-bis((*R*)-2-hydroxypropyl)-7,10-dimethoxyperyleno[1,12-*def*][1,3]dioxepine-6,11-dione (cercosporin, 1)

Cercosporin was purified from wild type *C. nicotianae* (ATCC$^©$ 18366$^™$) as a red powder. ^{1}H NMR (400 MHz, CCl$_3$D, δ): 7.07 (s, 2H), 5.74 (s, 2H), 4.21 (s, 6H), 3.59 (dd, *J* = 12.8, 6.8 Hz, 2H), 3.39 (m, 2H), 2.90 (dd, *J* = 12.8, 6.0 Hz, 2H), 0.65 (d, *J* = 6.0 Hz, 6H); HRMS–ESI/ACPI-TOF (*m/z*): [MH$^+$] calcd for C$_{29}$H$_{27}$O$_{10}$, 535.1604; found, 535.1599 (100); UV (acetonitrile [*aq.*]) λ$_{max,}$ nm: 224, 260, 272, 324, 470, 566; HPLC (Kinetex XB-C18, 1.25 mL/min, 5–95% acetonitrile + 0.1% formic acid, 10.8 min) *t*$_R$ = 7.22 min. These data are in agreement with literature values for cercosporin.[5b]

7,9,10-trihydroxy-3-methyl-1*H*-benzo[*g*]isochromen-1-one (*nor*-toralactone, 6)

Nor-toralactone was purified from *C. nicotianae* ΔCTB3c as a yellow powder. ^{1}H

NMR (400 MHz, CCl$_3$D, δ): 13.61 (s, 1H), 9.50 (s, 1H), 6.92 (s, 1H), 6.61 (d, *J* = 2.4 Hz,

1H), 6.48 (d, *J* = 2.4 Hz, 1H), 6.23 (s, 1H), 2.27 (s, 3H); HRMS–ESI/ACPI-TOF (*m/z*):

[MH$^+$] calcd for C$_{14}$H$_{11}$O$_5$, 259.0607; found, 259.0606 (100); UV (acetonitrile [*aq.*]) λ$_{max}$,

nm: 268, 278, 292, 390; HPLC (Kinetex XB-C18, 1.25 mL/min, 5–95% acetonitrile +

0.1% formic acid, 10.8 min) t_R = 6.83 min. These data are in agreement with literature

values for *nor*-toralactone.[12]

9,10-dihydroxy-7-methoxy-3-methyl-1*H*-benzo[*g*]isochromen-1-one (toralactone, 7)

Toralactone was purified from *C. nicotianae* ΔCTB3c as a yellow powder. ^{1}H

NMR (400 MHz, CCl$_3$D, δ): 13.55 (s, 1H), 9.42 (s, 1H), 6.98 (s, 1H), 6.62 (d, *J* = 2.4 Hz,

1H), 6.54 (d, *J* = 2.4 Hz, 1H), 6.24 (s, 1H), 3.90 (s, 3H), 2.27 (s, 3H); HRMS–ESI/ACPI-

TOF (*m/z*): [MH$^+$] calcd for C$_{15}$H$_{13}$O$_5$, 273.0763; found, 273.0764 (100); UV

(acetonitrile [*aq.*]) λ$_{max}$, nm: 268, 278, 292, 390; HPLC (Kinetex XB-C18, 1.25 mL/min,

5–95% acetonitrile + 0.1% formic acid, 10.8 min) t_R = 8.51 min. These data are in

agreement with literature values for toralactone.[40]

**7,9-dihydroxy-3-methyl-5-(λ^3-oxidanylidene)-1*H*-benzo[*g*]isochromene-1,10(5*H*)-
dione (8)**

Product **8** was purified from *C. nicotianae* ΔCTB3c as a red powder. This product

was also observed as an oxidation product of *nor*-toralactone in *in vitro* CTB3 reactions.

HRMS–ESI/ACPI-TOF (*m/z*): [MH$^+$] calcd for C$_{14}$H$_9$O$_6$, 273.0399; found, 273.0388

(100); UV (acetonitrile [*aq.*]) λ$_{max}$, nm: 248, 270sh, 296sh, 474; HPLC (Kinetex XB-

C18, 1.25 mL/min, 5–95% acetonitrile + 0.1% formic acid, 10.8 min) t_R = 4.92 min.

2,5-dihydroxy-6-methoxy-7-(2-oxopropyl)naphthalene-1,4-dione (cercoquinone A, 9)

Metabolite **9** was purified from *C. nicotianae* ΔCTB6 as an orange powder. ^{1}H

NMR (400 MHz, CCl$_3$D, δ): 12.76 (s, 1H), 7.49 (s, 1H), 6.26 (s, 1H), 4.05 (s, 3H), 3.80

(s, 2H), 2.27 (s, 3H); HRMS–ESI/ACPI-TOF (*m/z*): [MH$^+$] calcd for C$_{14}$H$_{13}$O$_6$,

277.0712; found, 277.0711 (100); HRMS/MS–ESI/ACPI-TOF (*m/z*): [MH$^+$ –CO] calcd

for C$_{13}$H$_{13}$O$_5$, 249.0763; found, 249.0756 (15); [MH$^+$ –CO –H$_2$O] calcd for C$_{13}$H$_{11}$O$_4$,

231.0657; found, 231.0649 (17); [MH$^+$ –C$_3$H$_5$O] calcd for C$_{11}$H$_8$O$_5$, 220.0372; found,

220.0369 (54); [MH$^+$ –C$_2$H$_5$O$_2$] calcd for C$_{12}$H$_8$O$_4$, 216.0423; found, 216.0415 (30);

[MH$^+$ –CO –H$_2$C=C=O] calcd for C$_{11}$H$_{11}$O$_4$, 207.0657; found, 207.0655 (100); [MH$^+$ –

C$_3$H$_4$O$_2$] calcd for C$_{11}$H$_9$O$_4$, 205.0501; found, 205.0494 (42); [MH$^+$ –C$_4$H$_5$O$_2$] calcd for

C$_{10}$H$_8$O$_4$, 192.0423; found, 192.0414 (57); [MH$^+$ –C$_4$H$_6$O$_2$] calcd for C$_{10}$H$_7$O$_4$, 191.0344;

found, 191.0343 (43); [MH$^+$ –C$_3$H$_4$O$_3$] calcd for C$_{11}$H$_9$O$_3$, 189.0552; found, 189.0546

(63); [MH$^+$ –C$_4$H$_4$O$_3$] calcd for C$_{10}$H$_9$O$_3$, 177.0552; found, 177.0543 (35); [MH$^+$ –

C$_4$H$_6$O$_3$] calcd for C$_{10}$H$_7$O$_3$, 175.0395; found, 175.0390 (79); [MH$^+$ –C$_4$H$_4$O$_4$] calcd for

C$_{10}$H$_9$O$_2$, 161.0603; found, 161.0594 (47); [MH$^+$ –C$_6$H$_6$O$_3$] calcd for C$_8$H$_7$O$_3$, 151.0395;

found, 151.0388 (21); UV (acetonitrile [*aq.*]) λ$_{max}$, nm: 248, 300, 422; HPLC (Kinetex

XB-C18, 1.25 mL/min, 5–95% acetonitrile + 0.1% formic acid, 10.8 min) t_R = 5.42 min.

9-hydroxy-6-methoxy-2-methyl-2,3-dihydronaphtho[2,3-*b*]furan-5,8-dione

(cercoquinone B, 10)

Metabolite **10** was purified from *C. nicotianae* ΔCTB5 as a red-orange powder.

^{1}H NMR (400 MHz, CD$_3$CN, δ): 7.55 (s, 1H), 6.19 (s, 1H), 3.97 (s, 3H), 2.78 (m, 2H),

1.16 (d, *J* = 6.4 Hz, 3H); HRMS–ESI/ACPI-TOF (*m/z*): [MH$^+$] calcd for C$_{14}$H$_{13}$O$_5$,

261.0763; found, 261.0761 (100); HRMS/MS–ESI/ACPI-TOF (*m/z*): [MH$^+$ –CH$_3$] calcd

for $C_{13}H_{10}O_5$, 246.0528; found, 246.0529 (36); [MH$^+$ –CO] calcd for $C_{13}H_{13}O_4$, 233.0814; found, 233.0812 (37); [MH$^+$ –2CH$_3$] calcd for $C_{12}H_7O_5$, 231.0293; found, 231.0295 (37); [MH$^+$ –CH$_3$ –H$_2$O] calcd for $C_{13}H_8O_4$, 228.0423; found, 228.0419 (22); [MH$^+$ –C$_3$H$_4$O] calcd for $C_{11}H_9O_4$, 205.0501; found, 205.0499 (46); [MH$^+$ –C$_3$H$_6$O] calcd for $C_{11}H_7O_4$, 203.0344; found, 203.0345 (100); [MH$^+$ –CO –CH$_3$OH] calcd for $C_{12}H_9O_3$, 201.0551; found, 201.0531 (26); [MH$^+$ –C$_2$H$_5$O$_2$] calcd for $C_{12}H_8O_3$, 200.0473; found, 200.0474 (68); [MH$^+$ –C$_3$H$_5$O$_3$] calcd for $C_{11}H_8O_2$, 172.0534; found, 172.0523 (25); UV (acetonitrile [*aq.*]) λ_{max}, nm: 250, 302, 426; HPLC (Kinetex XB-C18, 1.25 mL/min, 5–95% acetonitrile + 0.1% formic acid, 10.8 min) t_R = 5.07 min.

3,5-dihydroxy-7-methoxy-2-(2-oxopropyl)naphthalene-1,4-dione (cercoquinone D, 11)

1,4-Naphthoquinone **11** was analyzed from *in vitro* reactions of CTB3 against *nor*-toralactone or toralactone as a pale yellow residue. Limited yields hindered ^{1}H NMR analysis of this product. HRMS–ESI/ACPI-TOF (*m/z*): [MH$^+$] calcd for $C_{14}H_{13}O_6$, 277.0712; found, 277.0717 (100); HRMS/MS–ESI/ACPI-TOF (*m/z*): [MH$^+$ –H$_2$C=C=O] calcd $C_{12}H_{11}O_5$, 235.0607; found, 235.0615 (100); [MH$^+$ –CO –H$_2$O] calcd $C_{13}H_{11}O_4$, 231.0657; found, 231.0656 (24); [MH$^+$ –H$_2$C=C=O –H$_2$O] calcd $C_{12}H_9O_4$, 217.0501; found, 217.0502 (65); [MH$^+$ –H$_2$C=C=O –H$_2$O –CO] calcd $C_{11}H_9O_3$, 189.0551; found, 189.0558 (51); [MH$^+$ –H$_2$C=C=O –H$_2$O –2CO] calcd $C_{10}H_9O_2$, 161.0602; found, 161.0611 (42); [MH$^+$ –CO –C$_5$H$_6$O$_2$] calcd $C_8H_7O_3$, 151.0395; found, 151.0401 (46); UV (acetonitrile [*aq.*]) λ_{max}, nm: 260, 320sh, 380sh; HPLC (Kinetex XB-C18, 1.25 mL/min, 5–95% acetonitrile + 0.1% formic acid, 10.8 min) t_R = 4.65 min.

8-hydroxy-6-methoxy-3-(2-oxopropyl)naphthalene-1,2-dione (cercoquinone C, 12)

1,2-Naphthoquinone **12** was analyzed from reactions of CTB3 under non-reductive conditions. Limited yields hindered ^{1}H NMR analysis of this product. HRMS–ESI/ACPI-TOF (*m/z*): [MH[+]] calcd for $C_{14}H_{13}O_5$, 261.0763; found, 261.0767 (100); [MNa[+]] calcd for $C_{14}H_{12}O_5Na$, 283.0582; found, 283.0582 (76); HRMS/MS–ESI/ACPI-TOF (*m/z*): [MH[+] –CO –H_2O] calcd $C_{13}H_{11}O_3$, 215.0708; found, 215.0709 (100); [MH[+] –H_2C=C=O –CO] calcd $C_{11}H_{11}O_3$, 191.0708; found, 191.0708 (91); [MH[+] –2CO –H_2O] calcd $C_{12}H_{11}O_2$, 187.0759; found, 187.0757 (92); [MH[+] –2CO –H_2O –CH_3] calcd $C_{11}H_8O_2$, 172.0524; found, 172.0522 (56); [MH[+] –H_2C=C=O –2CO] calcd $C_{10}H_{11}O_2$, 163.0759; found, 163.0758 (31); [MH[+] –CO –C_5H_6O] calcd $C_8H_7O_3$, 151.0395; found, 151.0399 (28); [MH[+] –H_2C=C=O –2CO –H_2O] calcd $C_{10}H_9O$, 144.0653; found, 144.0576 (77); UV (acetonitrile [*aq.*]) λ_{max}, nm: 216, 286, 306sh, 386, 504; HPLC (Kinetex XB-C18, 1.25 mL/min, 5–95% acetonitrile + 0.1% formic acid, 10.8 min) t_R = 6.39 min.

5.6. References

[1] M. E. Daub, *Phytopathology* **1981**, *71*, 213-213.
[2] (a) B. J. Morgan, S. Dey, S. W. Johnson, M. C. Kozlowski, *J Am Chem Soc* **2009**, *131*, 9413-9425; (b) B. J. Morgan, C. A. Mulrooney, M. C. Kozlowski, *J Org Chem* **2010**, *75*, 44-56.
[3] M. E. Daub, S. Herrero, K. R. Chung, *FEMS Microbiol Lett* **2005**, *252*, 197-206.
[4] (a) S. Kuyama, T. Tamura, *J Am Chem Soc* **1957**, *79*, 5725-5726; (b) S. Kuyama, T. Tamura, *J Am Chem Soc* **1957**, *79*, 5726-5729.
[5] (a) R. J. J. C. Lousberg, U. Weiss, C. A. Salemink, A. Arnone, L. Merlini, G. Nasini, *J Chem Soc D* **1971**, 1463-1464; (b) S. Yamazaki, T. Ogawa, *Agric Biol Chem* **1972**, *36*, 1707-&; (c) G. Nasini, L. Merlini, G. D. Andreetti, G. Bocelli, P. Sgarabotto, *Tetrahedron* **1982**, *38*, 2787-2796.
[6] M. E. Daub, R. P. Hangarter, *Plant Physiol* **1983**, *73*, 855-857.
[7] (a) M. E. Daub, *Plant Physiol* **1982**, *69*, 1361-1364; (b) M. E. Daub, S. P. Briggs, *Plant Physiol* **1983**, *71*, 763-766; (c) M. E. Daub, M. Ehrenshaft, *Annu Rev Phytopathol* **2000**, *38*, 461-490.

[8]   (a) S. Yamazaki, A. Okubo, Y. Akiyama, K. Fuwa, *Agric Biol Chem* **1975**, *39*, 287-288; (b) A. O. Fajola, *Physiol Plant Pathol* **1978**, *13*, 157-164; (c) M. E. Daub, *Phytopathology* **1982**, *72*, 370-374.

[9]   (a) M. Choquer, K. L. Dekkers, H. Q. Chen, L. H. Cao, P. P. Ueng, M. E. Daub, K. R. Chung, *Mol Plant Microbe Interact* **2005**, *18*, 468-476; (b) M. Choquer, M. H. Lee, H. J. Bau, K. R. Chung, *FEBS Lett* **2007**, *581*, 489-494; (c) K. L. Dekkers, B. J. You, V. S. Gowda, H. L. Liao, M. H. Lee, H. H. Bau, P. P. Ueng, K. R. Chung, *Fungal Genet Biol* **2007**, *44*, 444-454.

[10]   (a) M. E. Daub, G. B. Leisman, R. A. Clark, E. F. Bowden, *Proc Natl Acad Sci USA* **1992**, *89*, 9588-9592; (b) G. B. Leisman, M. E. Daub, *Photochem Photobiol* **1992**, *55*, 373-379; (c) C. C. Sollod, A. E. Jenns, M. E. Daub, *Appl Environ Microbiol* **1992**, *58*, 444-449.

[11]   A. Okubo, S. Yamazaki, K. Fuwa, *Agric Biol Chem* **1975**, *39*, 1173-1175.

[12]   A. G. Newman, A. L. Vagstad, K. Belecki, J. R. Scheerer, C. A. Townsend, *Chem Commun* **2012**, *48*, 11772-11774.

[13]   H. Q. Chen, M. H. Lee, M. E. Daub, K. R. Chung, *Mol Microbiol* **2007**, *64*, 755-770.

[14]   H. Q. Chen, M. H. Lee, K. R. Chung, *Microbiology* **2007**, *153*, 2781-2790.

[15]   (a) K. R. Chung, A. E. Jenns, M. Ehrenshaft, M. E. Daub, *Mol Gen Genet* **1999**, *262*, 382-389; (b) K. R. Chung, M. E. Daub, K. Kuchler, C. Schuller, *Biochem Biophys Res Commun* **2003**, *302*, 302-310; (c) S. Herrero, A. Amnuaykanjanasin, M. E. Daub, *FEMS Microbiol Lett* **2007**, *275*, 326-337.

[16]   B. Akache, K. Wu, B. Turcotte, *Nucleic Acids Res* **2001**, *29*, 2181-2190.

[17]   M. Ehrenshaft, R. G. Upchurch, *Appl Environ Microbiol* **1991**, *57*, 2671-2676.

[18]   (a) R. Vessecchi, F. S. Emery, S. E. Galembeck, N. P. Lopes, *Rapid Commun Mass Spectrom* **2010**, *24*, 2101-2108; (b) R. Vessecchi, F. S. Emery, S. E. Galembeck, N. P. Lopes, *J Mass Spectrom* **2012**, *47*, 1648-1659; (c) R. Vessecchi, F. S. Emery, N. P. Lopes, S. E. Galembeck, *Rapid Commun Mass Spectrom* **2013**, *27*, 816-824.

[19]   (a) J. H. Tatum, R. A. Baker, R. E. Berry, *Phytochemistry* **1987**, *26*, 795-798; (b) L. Studt, P. Wiemann, K. Kleigrewe, H. U. Humpf, B. Tudzynski, *Appl Environ Microbiol* **2012**, *78*, 4468-4480.

[20]   A. E. Jenns, M. E. Daub, *Phytopathology* **1995**, *85*, 906-912.

[21]   T. Weber, K. Blin, S. Duddela, D. Krug, H. U. Kim, R. Bruccoleri, S. Y. Lee, M. A. Fischbach, R. Muller, W. Wohlleben, R. Breitling, E. Takano, M. H. Medema, *Nucleic Acids Res* **2015**, *43*, W237-243.

[22]   H. Yang, Y. Wang, Z. Zhang, R. Yan, D. Zhu, *Genome Announc* **2014**, *2*.

[23]   A. Ebnother, T. M. Meijer, H. Schmid, *Helv Chim Acta* **1952**, *35*, 910-928.

[24]   M. H. Wheeler, M. A. Klich, *Pestic Biochem Physiol* **1995**, *52*, 125-136.

[25]   (a) W. Bauer, M. H. Zenk, *Tetrahedron Lett* **1989**, *30*, 5257-5260; (b) M. Rueffer, M. H. Zenk, *Phytochemistry* **1994**, *36*, 1219-1223.

[26]   A. G. Medentsev, V. K. Akimenko, *Phytochemistry* **1998**, *47*, 935-959.

[27]   H. C. Eisenman, A. Casadevall, *Appl Microbiol Biotechnol* **2012**, *93*, 931-940.

[28]   Z. Chen, M. A. Nunes, M. C. Silva, C. J. Rodrigues, Jr., *Mycologia* **2004**, *96*, 1199-1208.

[29]   P. W. Thomas, E. M. Stone, A. L. Costello, D. L. Tierney, W. Fast, *Biochemistry* **2005**, *44*, 7559-7569.

[30]   M. M. E. Huijbers, S. Montersino, A. H. Westphal, D. Tischler, W. J. H. van Berkel, *Arch Biochem Biophys* **2014**, *544*, 2-17.

[31]   B. Entsch, W. J. van Berkel, *The FASEB Journal* **1995**, *9*, 476-483.

[32]   I. A. Mirza, B. J. Yachnin, S. Wang, S. Grosse, H. Bergeron, A. Imura, H. Iwaki, Y. Hasegawa, P. C. Lau, A. M. Berghuis, *J Am Chem Soc* **2009**, *131*, 8848-8854.

[33]   P. Chaiyen, *Arch Biochem Biophys* **2010**, *493*, 62-70.

[34]   P. Chaiyen, P. Brissette, D. P. Ballou, V. Massey, *Biochemistry* **1997**, *36*, 8060-8070.

[35]   (a) F. M. Hauser, D. Sengupta, S. A. Corlett, *J Org Chem* **1994**, *59*, 1967-1969; (b) C. A. Merlic, C. C. Aldrich, J. Albaneze-Walker, A. Saghatelian, J. Mammen, *J Org Chem* **2001**, *66*, 1297-1309; (c) C. A. Mulrooney, B. J. Morgan, X. Li, M. C. Kozlowski, *J Org Chem* **2010**, *75*, 16-29.

[36]   J. Sambrook, D. W. Russell, *Molecular cloning: A laboratory manual*, Cold Spring Harbor Laboratory Press, Plainview, NY, **2001**.

[37]   D. W. Udwary, M. Merski, C. A. Townsend, *J Mol Biol* **2002**, *323*, 585-598.

[38]   M. M. Bradford, *Anal Biochem* **1976**, *72*, 248-254.

[39]   A. Aliverti, B. Curti, M. A. Vanoni, *Flavoprotein Protocols* **1999**, *131*, 9-23.

[40]   Z. J. Zhang, B. Yu, *J Org Chem* **2003**, *68*, 6309-6313.

Appendix A: NR-PKS Sequences

A.1. Native NR-PKS

Six native fungal NR-PKSs were investigated in this dissertation. Expression

constructs of all of these proteins were prepared in pET-24a(+) vectors for IPTG

inducible expression in *E. coli*. The resulting native proteins contained C-terminal 6 ×

His tags. In some cases, the NR-PKSs were codon optimized in portions of the genes—

typically the N-terminus—to enhance expression in *E. coli*. The sequences of these genes

and their resulting proteins are presented in this section. Underlined sequences are codon

optimized. Bold sequences are from the 6 × His tags.

Appendix Table A.1 | Details of full-length NR-PKS expression constructs.

Protein	Source Organism	Accessesion	Plasmid	Vector	Tag	MW (kDa)
PksA	*Aspergillus parasiticus*	Q12053	pEPksA	pET-24a(+)	C-His	232.0
Pks4	*Giberella fujikouroi*	CAB92399	pEPks4	pET-24a(+)	C-His	222.7
ACAS	*Aspergillus terreus*	EAU31624	pEACAS	pET-24a(+)	C-His	195.4
CTB1	*Cercospora nicotianae*	AAT69682	pECTB1	pET-24a(+)	C-His	238.0
wA	*Aspergillus nidulans*	CAA46695	pEwA	pET-24a(+)	C-His	236.9
Pks1	*Colletotichum lagenaria*	BAA18956	pEPks1alt	pET-24a(+)	C-His	238.1

A.1.1. PksA

Expression Construct (pEPksA) Gene Sequence:
ATGGCTCAATCAAGGCAACTCTTTCTCTTCGGCGATCAGACAGCGGATTTTGTTCCCAAGCTCCGCAGTTT
ACTATCCGTCCAGGACAGCCCTATTCTAGCCGCCTTTCTTGACCAGTCCCACTATGTCGTGCGAGCCCAGA
TGCTGCAGAGCATGAACACGGTTGATCACAAGTTGGCTCGAACCGCTGACCTGCGCCAAATGGTCCAGAAG
TATGTCGACGGCAAACTGACCCCTGCATTTCGAACCGCTCTAGTGTGCCTCTGCCAGTTGGGATGCTTCAT
CCGGGAATATGAGGAATCTGGCAACATGTACCCACAGCCCAGTGACAGCTACGTGCTGGGATTCTGCATGG
GTTCCTTGGCCGCTGTGGCGGTAAGCTGCAGTCGCTCCCTGTCAGAGCTGCTGCCTATCGCTGTACAAACT
GTGTTGATTGCCTTCCGCCTCGGTCTTTGCGCCCTGGAGATGCGGGATCGGGTGGATGGGTGTAGCGATGA
TCGAGGTGACCCTTGGTCTACCATTGTTTGGGGTCTGGATCCCCAGCAAGCTCGTGATCAGATTGAAGTGT
TCTGTCGGACCACAAACGTACCTCAGACAAGGCGTCCGTGGATCAGCTGCATCTCTAAGAATGCCATCACC
CTAAGTGGCAGTCCATCCACTTTGAGGGCGTTCTGTGCGATGCCTCAGATGGCCCAGCACCGGACTGCCCC
AATTCCCATCTGTTTACCGGCCCACAATGGCGCCCTCTTCACGCAGGCAGATATCACTACCATACTAGACA
CGACGCCTACCACTCCTTGGGAGCAACTGCCCGGCCAAATACCTTATATTTCCCATGTCACGGGGAATGTA
GTCCAGACTTCCAACTACCGGGACCTTATAGAGGTAGCCCTGTCTGAGACTCTCTTGGAGCAAGTGCGACT
AGACTTGGTTGAGACTGGACTGCCACGCCTTTTGCAATCTCGTCAGGTCAAGAGCGTCACCATCGTACCAT
TCTTGACTCGCATGAATGAGACAATGAGCAACATTCTCCCAGACAGCTTTATCAGTACAGAGACAAGGACT
GACACCGGACGAGCCATCCCAGCTTCAGGTCGACCAGGCGCAGGCAAGTGCAAGCTGGCTATTGTGTCCAT

```
GTCGGGGAGGTTCCCTGAATCACCGACCACCGAAAGCTTTTGGGACCTTCTATACAAAGGGTTGGATGTTT
GTAAAGAGGTTCCCCGTCGACGGTGGGACATCAACACGCATGTGGATCCCAGCGGGAAAGCACGAAACAAA
GGGGCTACCAAATGGGGCTGCTGGCTAGATTTCTCAGGCGATTTTGATCCCCGATTCTTTGGGATCTCGCC
CAAAGAGGCGCCACAGATGGATCCAGCTCAGCGCATGGCCTTGATGTCTACTTACGAGGCAATGGAGCGGG
CTGGTTTGGTTCCCGACACCACGCCGTCGACCCAGCGAGACCGCATTGGGGTCTTCCACGGAGTCACCAGT
AACGACTGGATGGAGACCAATACAGCCCAGAACATTGACACATACTTCATCACCGGTGGAAATCGGGGGTT
TATTCCCGGGCGCATTAACTTCTGTTTCGAATTTGCCGGACCCAGCTATACCAATGACACGGCCTGTTCAT
CCAGTCTAGCTGCCATCCACCTGGCCTGCAATTCTCTCTGGCGGGGCGACTGTGACACGGCGGTGGCAGGA
GGAACTAACATGATCTATACTCCTGATGGTCACACAGGATTGGACAAAGGGTTCTTTCTTTCCCGGACTGG
CAACTGCAAACCCTACGACGACAAGGCCGATGGTTACTGCCGAGCTGAGGGGGTCGGGACGGTGTTCATCA
AACGGCTGGAAGATGCTCTGGCAGATAATGACCCCATCCTTGGCGTTATTCTAGATGCTAAAACTAATCAC
TCAGCCATGTCGGAGTCGATGACTCGGCCGCACGTGGGCGCCCAAATCGATAACATGACGGCGGCGCTGAA
TACCACTGGACTCCATCCCAATGACTTTAGCTACATTGAGATGCATGGCACTGGCACCCAGGTAGGGGATG
CGGTGGAGATGGAGTCAGTCCTGTCGGTGTTTGCGCCGTCCGAAACCGCCAGAAAGGCGGATCAGCCACTA
TTTGTCGGCTCAGCCAAGGCCAACGTAGGACATGGAGAGGGAGTGTCTGGGGTTACGAGCCTTATTAAGGT
TCTGATGATGATGCAGCACGATACCATACCTCCTCACTGCGGCATCAAACCGGGCAGCAAATCAACCGCA
ACTTCCCTGATCTTGGAGCTCGCAATGTGCACATCGCCTTTGAACCCAAGCCCTGGCCACGAACACACACT
CCGCGCAGGGTGCTTATCAACAACTTCAGTGCCGCGGGAGGGAATACTGCCTTGATAGTGGAAGACGCTCC
GGAGCGTCACTGGCCGACAGAGAAGGATCCGCGCTCTAGTCATATCGTCGCCCTGTCTGCGCATGTGGGGG
CTTCCATGAAAACCAACCTCGAACGACTGCATCAGTATCTCCTGAAAAACCCCCACACTGATCTCGCGCAG
CTGTCATATACTACTACTGCGCGTCGATGGCATTATCTACACCGAGTGAGCGTCACTGGCGCGTCTGTTGA
AGAAGTGACTCGCAAGCTAGAGATGGCCATACAGAACGGGGACGGAGTCAGTCGACCCAAAAGCAAGCCGA
AGATTCTCTTTGCTTTCACGGGACAAGGGTCTCAATATGCAACTATGGGTAAGCAGGTGTACGATGCGTAT
CCATCTTTCAGAGAGGACCTGGAGAAGTTTGATCGGTTGGCGCAAAGTCATGGCTTCCCTAGCTTTCTTCA
CGTCTGTACTTCACCTAAAGGGGATGTGGAAGAGATGGCTCCCGTTGTGGTGCAACTGGCTATCACTTGTC
TCCAAATGGCCCTTACTAACCTCATGACCTCCTTCGGGATCCGTCCCGATGTGACAGTGGGGCATAGTTTG
GGTGAATTTGCAGCCCTGTATGCGGCGGGAGTTCTTTCGGCCTCAGACGTCGTTTACCTTGTTGGTCAAAG
AGCGGAGCTACTCCAGGAGCGCTGCCAACGCGGGACGCATGCCATGCTGGCTGTGAAAGCTACCCCTGAAG
CGTTGTCCCAATGGATCCAGGATCATGACTGTGAGGTGGCCTGTATTAATGGCCCTGAAGATACCGTTCTC
AGTGGCACCACTAAGAATGTTGCCGAGGTTCAACGCGCTATGACGGACAACGGGATCAAATGCACGCTGTT
GAAACTGCCGTTTGCCTTCCATTCTGCCCAGGTGCAACCTATTCTGGACGACTTTGAGGCCCTGGCTCAGG
GAGCGACATTTGCCAAGCCTCAACTACTAATTCTCTCTCCCTTGCTGCGGACAGAAATCCACGAACAAGGC
GTCGTGACTCCATCATATGTCGCGCAACATTGTCGTCACACCGTAGATATGGCCCAAGCTTTGAGATCTGC
TCGAGAAAAGGGACTCATCGACGACAAAACCCTCGTCATTGAGCTGGGACCGAAGCCATTAATCTCGGGCA
TGGTGAAAATGACACTGGGAGACAAAATTAGCACCTTACCCACTCTAGCACCTAACAAGGCCATTTGGCCC
AGCCTGCAGAAGATTCTCACCTCGGTCTACACGGGTGGGTGGGATATTAATTGGAAGAAATATCACGCCCC
TTTCGCCTCCTCCCAGAAGGTGGTGGATCTGCCGAGCTACGGCTGGGATTTGAAGGACTACTACATCCCGT
ATCAGGGTGACTGGTGTCTGCATCGCCACCAGCAGGATTGTAAGTGCGCCGCTCCTGGCCACGAAATCAAA
ACGGCCGACTACCAAGTGCCTCCTGAGTCGACGCCTCACCGTCCATCCAAGCTGGACCCTAGCAAGGAGGC
CTTCCCCGAAATCAAGACCACCACGACACTCCATCGAGTGGTGGAAGAGACGACCAAACCTTTGGGCGCCA
CCCTAGTTGTGGAGACAGACATATCTCGGAAGGATGTCAACGGCCTCGCTCGAGGGCACCTTGTCGATGGG
ATCCCTTTGTGTACCCCTTCCTTTTATGCTGACATCGCCATGCAAGTGGGCCAATACAGTATGCAACGGCT
CCGTGCGGGACATCCGGGGGCCGGTGCCATAGATGGCCTTGTGGACGTGTCCGACATGGTGGTGGACAAAG
CGTTGGTTCCCCATGGGAAGGGACCTCAATTGCTGCGCACGACGCTTACCATGGAGTGGCCGCCCAAGGCT
GCTGCTACTACGCGAAGCGCCAAAGTCAAATTCGCCACCTATTTTGCCGATGGGAAGCTCGATACGGAGCA
TGCCAGCTGTACTGTCAGATTCACAAGCGATGCACAGTTGAAATCTCTACGCCGGTCTGTGTCCGAGTACA
AGACCCACATTCGTCAGTTACATGATGGCCATGCTAAGGGACAGTTCATGCGATACAATAGGAAGACCGGG
TACAAGCTCATGAGCAGCATGGCTCGGTTTAATCCCGACTACATGCTCCTAGATTATCTGGTGCTGAACGA
AGCAGAGAACGAGGCAGCAAGTGGTGTAGACTTCTCGTTGGGATCGTCGGAAGGCACCTTCGCAGCTCACC
CAGCTCACGTCGATGCCATCACTCAGGTGGCCGGCTTTGCTATGAATGCCAATGACAATGTCGACATTGAG
AAACAGGTCTACGTTAATCACGGTTGGGACTCGTTCCAGATCTACCAACCGCTGGATAATAGCAAGTCTTA
CCAGGTCTACACCAAGATGGGTCAAGCGAAGGAGAATGATTTGGTGCATGGCGATGTGGTAGTTCTGGACG
GAGAACAAATCGTTGCTTTCTTCCGCGGCCTTACGCTGCGATCAGTTCCTCGTGGTGCACTGCGTGTCGTC
CTGCAGACTACAGTGAAAAAGGCCGATCGCCAACTAGGATTTAAGACAATGCCGTCGCCGCCGCCCCCGAC
AACGACAATGCCAATATCGCCTTATAAACCAGCTAATACTCAGGTTTCCAGCCAAGCTATTCCAGCAGAGG
CCACTCATTCTCACACCCCGCCACAGCCAAAGCATTCCCCGGTACCGGAAACTGCCGGAAGCGCTCCAGCG
GCAAAAGGAGTAGGTGTCAGTAACGAAAAGTTAGATGCTGTAATGCGAGTCGTTTCGGAGGAGAGTGGAAT
TGCCCTCGAGGAGCTCACCGATGACAGCAACTTTGCTGACATGGGCATCGACTCTCTGAGTTCAATGGTCA
```

```
TTGGGAGCCGCTTCAGAGAGGACCTGGGGCTGGACCTGGGGCCTGAGTTTTCTCTTTTCATTGACTGCACT
ACCGTGCGTGCCTTGAAAGACTTCATGTTGGGAAGCGGGGATGCTGGCAGTGGCTCCAATGTAGAAGATCC
TCCCCCATCAGCTACTCCCGGCATCAACCCCGAAACCGATTGGTCTAGCAGTGCCTCTGATAGTATTTTCG
CCAGCGAAGACCACGGTCATTCGAGTGAGTCCGGCGCCGACACCGGAAGTCCGCCTGCACTTGATCTGAAG
CCCTACTGCCGCCCCTCAACTTCTGTCGTCCTACAAGGTCTACCTATGGTGGCGCGGAAAACTCTGTTTAT
GCTCCCTGATGGCGGGGGGTCTGCGTTCTCCTACGCCTCCCTGCCGCGCCTCAAATCAGATACTGCCGTTG
TGGGCCTGAATTGCCCTTATGCTCGGGATCCCGAGAACATGAACTGCACACATGGAGCTATGATTGAGAGC
TTTTGCAATGAGATCCGGCGGCGACAGCCACGGGGCCCCTATCACCTGGGCGGCTGGTCGTCCGGTGGTGC
ATTCGCTTACGTCGTGGCCGAGGCACTTGTTAACCAAGGCGAGGAGGTGCATTCGTTAATCATCATTGATG
CGCCTATTCCCCAAGCCATGGAACAACTTCCCCGAGCATTTTACGAGCACTGCAATAGCATTGGATTGTTC
GCTACCCAGCCGGGGGCTAGTCCGGACGGCTCGACTGAGCCTCCATCCTACTTAATCCCACACTTTACCGC
TGTGGTGGATGTGATGCTGGATTACAAGCTGGCCCCGTTGCATGCGCGCCGGATGCCCAAGGTCGGCATCG
TCTGGGCGGCAGATACAGTCATGGACGAGCGGGACGCTCCCAAGATGAAAGGAATGCATTTTATGATTCAG
AAGCGGACGGAATTTGGTCCCGATGGGTGGGATACGATCATGCCCGGGGCCTCGTTTGACATTGTCCGAGC
AGACGGTGCTAATCATTTTACGTTGATGCAAAAGGAACATGTCTCTATAATTAGCGATCTGATCGACCGGG
TCATGGCT**GCGGCCGCACTCGAGCACCACCACCACCACCACTGA**
```

Expression Construct (pEPksA) Protein Sequence:
```
MAQSRQLFLFGDQTADFVPKLRSLLSVQDSPILAAFLDQSHYVVRAQMLQSMNTVDHKLARTADLRQMVQK
YVDGKLTPAFRTALVCLCQLGCFIREYEESGNMYPQPSDSYVLGFCMGSLAAVAVSCSRSLSELLPIAVQT
VLIAFRLGLCALEMRDRVDGCSDDRGDPWSTIVWGLDPQQARDQIEVFCRTTNVPQTRRPWISCISKNAIT
LSGSPSTLRAFCAMPQMAQHRTAPIPICLPAHNGALFTQADITTILDTTPTTPWEQLPGQIPYISHVTGNV
VQTSNYRDLIEVALSETLLEQVRLDLVETGLPRLLQSRQVKSVTIVPFLTRMNETMSNILPDSFISTETRT
DTGRAIPASGRPGAGKCKLAIVSMSGRFPESPTTESFWDLLYKGLDVCKEVPRRRWDINTHVDPSGKARNK
GATKWGCWLDFSGDFDPRFFGISPKEAPQMDPAQRMALMSTYEAMERAGLVPDTTPSTQRDRIGVFHGVTS
NDWMETNTAQNIDTYFITGGNRGFIPGRINFCFEFAGPSYTNDTACSSSLAAIHLACNSLWRGDCDTAVAG
GTNMIYTPDGHTGLDKGFFLSRTGNCKPYDDKADGYCRAEGVGTVFIKRLEDALADNDPILGVILDAKTNH
SAMSESMTRPHVGAQIDNMTAALNTTGLHPNDFSYIEMHGTGTQVGDAVEMESVLSVFAPSETARKADQPL
FVGSAKANVGHGEGVSGVTSLIKVLMMQHDTIPPHCGIKPGSKINRNFPDLGARNVHIAFEPKPWPRTHT
PRRVLINNFSAAGGNTALIVEDAPERHWPTEKDPRSSHIVALSAHVGASMKTNLERLHQYLLKNPHTDLAQ
LSYTTTARRWHYLHRVSVTGASVEEVTRKLEMAIQNGDGVSRPKSKPKILFAFTGQGSQYATMGKQVYDAY
PSFREDLEKFDRLAQSHGFPSFLHVCTSPKGDVEEMAPVVVQLAITCLQMALTNLMTSFGIRPDVTVGHSL
GEFAALYAAGVLSASDVVYLVGQRAELLQERCQRGTHAMLAVKATPEALSQWIQDHDCEVACINGPEDTVL
SGTTKNVAEVQRAMTDNGIKCTLLKLPFAFHSAQVQPILDDFEALAQGATFAKPQLLILSPLLRTEIHEQG
VVTPSYVAQHCRHTVDMAQALRSAREKGLIDDKTLVIELGPKPLISGMVKMTLGDKISTLPTLAPNKAIWP
SLQKILTSVYTGGWDINWKKYHAPFASSQKVVDLPSYGWDLKDYYIPYQGDWCLHRHQQDCKCAAPGHEIK
TADYQVPPESTPHRPSKLDPSKEAFPEIKTTTTLHRVVEETTKPLGATLVVETDISRKDVNGLARGHLVDG
IPLCTPSFYADIAMQVGQYSMQRLRAGHPGAGAIDGLVDVSDMVVDKALVPHGKGPQLLRTTLTMEWPPKA
AATTRSAKVKFATYFADGKLDTEHASCTVRFTSDAQLKSLRRSVSEYKTHIRQLHDGHAKGQFMRYNRKTG
YKLMSSMARFNPDYMLLDYLVLNEAENEAASGVDFSLGSSEGTFAAHPAHVDAITQVAGFAMNANDNVDIE
KQVYVNHGWDSFQIYQPLDNSKSYQVYTKMGQAKENDLVHGDVVVLDGEQIVAFFRGLTLRSVPRGALRVV
LQTTVKKADRQLGFKTMPSPPPPTTTMPISPYKPANTQVSSQAIPAEATHSHTPPQPKHSPVPETAGSAPA
AKGVGVSNEKLDAVMRVVSEESGIALEELTDDSNFADMGIDSLSSMVIGSRFREDLGLDLGPEFSLFIDCT
TVRALKDFMLGSGDAGSGSNVEDPPPSATPGINPETDWSSSASDSIFASEDHGHSSESGADTGSPPALDLK
PYCRPSTSVVLQGLPMVARKTLFMLPDGGGSAFSYASLPRLKSDTAVVGLNCPYARDPENMNCTHGAMIES
FCNEIRRRQPRGPYHLGGWSSGGAFAYVVAEALVNQGEEVHSLIIIDAPIPQAMEQLPRAFYEHCNSIGLF
ATQPGASPDGSTEPPSYLIPHFTAVVDVMLDYKLAPLHARRMPKVGIVWAADTVMDERDAPKMKGMHFMIQ
KRTEFGPDGWDTIMPGASFDIVRADGANHFTLMQKEHVSIISDLIDRVMA**AAALEHHHHHH**
```

A.1.2. Pks4

Expression Construct (pEPks4) Gene Sequence:
```
ATGGCGAGCAGCGCCGACGTGTATGTGTTCGGCGATCAGAGCACCCCAGTACTGGACAAATTACAGGCATT
GGTTCGCGTGAAGGATAACGCTTTACTGACATCGTTTTTGGGCGAAGCCTTTTTGGCCGTTCGCCGTGAAA
TCGTGTCCCTCTCGTCTCTGGAGCGTAAGAGTATTCCGGAAGCGGAGTCTTTAAGCCTCCTGCTTGAAGGC
GTACGGCGCTCCGAGCCCCACGCAGCTCTGGATAGCGCGTTTGTCTGTATTTACGAAATTGGTTATTACAT
CGACTACCTGGCTCGTTCCGACAAACAGCATCCACCGGCAGCGCCTTCATTGCTGTTAGGTATTTGTACTG
```

171

GCTCGATCGCCGCCGCGGCAGTCAGCTGCGCAAAAGATGTTTTTGAAATTAGTCGCCTGGGAGTTGAGGCT
GCCACGGTCGCGTTCCGTCTGGGTATGCATGTCCGTCGCCGCGCAGAAATCTTGGGTATTCGACTCCTTC
ATCCTGGAGTATGATCCTGTCATCTAACCAAGAGGAACTGGTGAGCGAAGCGCTGAAAGAATTCTCAAAGG
AAAAAAATCTTACATACAGTTCGCGTCCCTACATCAGCGCCACTGGTCCTGGTTTCACAACAATCAGCGGT
CCCCCATCTATCCTCGAGTCAGTGAAGTCTTGCGATACTTTCTCTGGAAAGAGGTTATATCCAGCTCCCAT
CTATGGTCCCTACCATAACTCTTCATCATACTCCGAATCCAGCCTGGAACATGGCCTTGCATCAATCCTTG
AAGACGTTGGGTTCCTAGAGAACGAGATGCTCATTCCCATCATCTCCTGCGCCTCCGGATCTCGCTTAGAC
CAGTTGTCTTTCGGGAACCTCCTCAAGAACGTTTTGAGCAGCGCTCTGTCTCAGCAAATTCGTATGGATCT
CGTCACCGACGCTCTAGTCGAAACCGTTTCTGGCACTGAAGCTACCTTGATACCGGTCAATGCCCAAACAA
CTGTTTGCAGTCTCGCTGATTGGTTGGCTAAGAGAGGAGCTACAACTCGCATTGGGCCAACCTTGGAGAGC
CTGACCAAAGACCGAGCAGAACCCAATCTTGCTCCTGGCGATGAGAACAAGATCGCCATCATCGGTTTCAG
CGGTCGATTCCCAGAAGCCGACAACCTCGACGAGTTCTGGGATCTTCTAATCCGTGGTCTCGACGTCCATA
AACCTGTCCCCGAAGAGCGCTTTGCTCGTGACCACTACGATCCTACTGGCCAGCGCAAGAACACCAGCCAA
GTCCAGTACGGCTGCTGGTTAAAGTCTGCAGGTTACTTCGACACCCAGTTCTTCCACATGTCGCCCAAGGA
AGCCATGCAGACTGATCCAGCTCAGCGACTCGCCCTGCTCACGGCTTACGAAGCCCTCGAGATGGCTGGTG
TCGTGCCGGACCGAACTCCTTCCACGCAGCGGAACCGCGTGGGAGTTTACTACGGTACTACGAGTAACGAT
TGGGGCGAGGTGAACTCCTCACAGGATGTTGATACTTACTACATTCCTGGAGCCAATCGTGCCTTCATTCC
TGGAAGGGTCAATTACTTTTTCAAGTTCACGGGCCCGAGTATCGCTGTTGACACTGCTTGCTCTTCGAGTC
TTGCTGCGATCAATCTTGCGATCACGTCGTTGAAGAACAGGGATTGTAACACTGCGATTGCTGGCGGAACC
AATGTCATGACGAACCCGGACAACTTTGCAGGCCTCGATCGCGGGCATTTCCTTTCCCGTACGGGAAACTG
CAAGGCTTTTAACGACGGTGCGGATGGTTACTGCCGCGCTGATGGAATCGGCACCTTGATTCTCAAACGCC
TTCCAGATGCTATCGCAGACAGTGATCCCATCTTTGGAGTCATCCTGGGAGCACACACCAACCACTCAGCA
GAGTCCGTCTCCATCACCCGACCTCTCGCAGACGCCCAGGAGTATCTCTTCAAGAAGCTGTTGAACGAAAC
CGGTATCCATCCTCACGATGTTAGCTACGTCGAGATGCACGGAACGGGGACCCAGGCTGGCGATGCAGTTG
AGATGCGATCCGTCCTCAACTCATTCGCCTTCGATCACAGCCGTCCACGCGATAAGAGCTTGTATCTCGGC
TCAGTCAAAGCAAACGTCGGACACGCTGAATCAGCCTCTGGAGTTCTCGCCATCATCAAGGTGCTGCTGAT
GATGCAGAAGAACACGATTCCGCCTCACTGCGGTATCAAGACCAAGATCAACCAGGGTTTTCCTAAGGACT
TGGATCATCGTGGTGTGCGCATTGCTCTGAAGGATAGCGTGGATTGGTCTCGGCCTGAGGGCGGTAAGCGC
AGGGTGTTGGTGAACAACTTCTCTGCTGCTGGTGGAAACACTTCATTGCTGCTTGAGGATGGTCCCGCTGT
TCATCCAGCACGACAGCATCAAGATGGGGATGCTCGTACTGAGCATGTCGTGGCTGTCTCGGCGCGGTCTA
CCAAGGCTTTGGAGGAGAACCTGAAGGCCCTTGAAGCTTACATCGCCAACTCTTGGGCTCCGGAGGGTGAG
CTTCTCTCTCAGCTTTCATACACGACTACTGCTCGACGTGTACACCACAGTCGACGTGTTGCGTTTGTCAC
CAACGGTCTTGACGACTTGAGGAAGTCTCTTCTCAAGGCTGCTACCGATGCTGGTCAAGTCAAGGGCATTC
CCGCAGTGTCACCCAAGGTCGGCTTCCTCTTCACTGGACAGGGAGCGCAGGAGACTGCCATGGCCATTGGG
TACTACAAGAGCTTCTCATCCTTCCGATCAGACATTCACCAGCTTGACTCCATCGCAACTCTTCAGGGACT
CCCATCTGTTCTCCCACTCATTCACGGCACAACACCCGTTGAAGATCTCTCTGCAGTTGTCGTGCAGCTCG
GAACGTGCATTATCCAGATTTCACTCGCCCGCTTCTGGATCAGCCTCGGTATCACTCCCCAGTACGTCATC
GGACATAGCCTGGGTGAGTATGCTGCTCTTCAGATTGCTGGTGTTCTCAGCGTCAACGATGCCATCTTTCT
CTGTGGTCACAGGGCAGCGCGCTTCTTGACAAGAAGTGCACTGCCTACACGCACGGCATGGTCGCTGTCAAAG
CGGCGGCTGATGATTTGAGGCAACACATCTCATCTGACTTGAAGGTCGAGATTGCTTGTGTCAACGGCGCT
GAAGACACTGTCCTTAGTGGTCCCAACGCGGACATCGAGTCGCTCTGTGGGAAGTTGACTCAGGCTGGCTA
CAAGCTTCACAAACTAGAGATTCCGTTTGCCTTCCACTCTTCTCAGGTTGATCCCATTCTAGATGACCTCG
AAGAACTGGCGTCCCAAGTTGGGTTCCACGAACCCAAACTTCCTATCGTCTCGCCTCTTCTGCGCACGCTA
CTCACTGGTGATACCCTTGGTCCTCAGTACATCCGACGTCATTGCCGTGAGACAGTTGACTTCCTCGGTGC
TATCAAAATGGCTGAATCGCAAGGCATCATGGACCGAAGCGGCATGTGTATCGAAATCGGAGCCCATCCCA
TCCTCACGCGAATGGTCAAGTCAATCATCGGCCAAGACTTCCGGTGCCTTGCGTCTCTCCGCCGCAAAGAG
GACCACTTCAAGACTCTCGCAGATTCCCTTTGCGCTTTGCATCTGGCTGGCTTCTCCGTGAACTGGGACGA
GTACCATCGTGACTTTGCATCTTCTCGCAATGTTCTTCAGCTCCCCAAGTATAGTTGGCAGCTTGCCAACT
ACTGGATGCAGTACAAGTACAGCTGGTGTCTTACCAAGGGCGACGCTCCCGTTGAGAATGGCCCGGTCGGT
GCAGTTGTACAAGCGCGCGCTTTACGTCTTTCTGATAGCGTTCACAACGTGATTGAACAAGTGCATGGCGA
CAAACGTTCCAGTATCACGGTCGAGTCAGATATGCATGATCCGTCGTTGCTGGCCATTGCGCAGAACCATC
GCGTAAATGGTCTCACCATGGCACCGAGCACTCTCTTCGCCGACATCGCCTTCACCCTCGCCAAGCACCTC
ATCCAGAACCACGGCCTCGATACGCACACGAACCTACCTTCAATCAACAACATGGCTGTCGAAAAGGCTCT
CATCGTGGGAGAGACTGGCCCTCAGCTGTTCCGAGCATCGCTCGATATGGACTGGACCACCATGCGCGGTT
CAGTTCGCATCTTCAGCGTCGGTGCAAACGGAAAGCAGACCACTCTCCATGCAGTCTGCGATGTAGCGGTC
GAGAATCCAAGTTCTCACCGTGAAAGCTGGCAGAGTAACGCGTACCTCATCCAGCGAGGCATCAAACAGCT
CGTCCAAGGCGCCAGTGATGGTTCTGCTCACATGATGCGTCGTGGTCTCTTGTACAAGATATTCTCCAACT
CGGTCCAGTACGGGTCTGCGTTCCAAGGCATCGAGCAGGTCTGGTTTGACAGTGAAGGACTCGAAGGAACT

```
GGCAAGGTCTTTATGCCCTCAGGCAAAGACACTTTCGCGCTCAACCCGTACTGCTGTGACAGTCTGGGTCA
CATCACTGGCTTCATCATGAACTGCAGTGATAGCTTGGATCTTGATGATCATGTCTACATCAATCATGGCT
GGCGCACGCTTCGCTTGGTTGAACCTTACCAGTGTGACGTGCAGTATCAGACATACGTCAAGATGCAGGCG
GTAGGCTCTGATGATTCGACTTATAGTGGCGATGTCCACGTTCTTCGTGATGGTAAGATTATCGGTATCTG
TGGTGGTGTCACGTTCAAGAAGGTCGCTCGCAAGGTACTCGAGATGCTTCTACCCAAGCCTTCCGGAGCCA
AAGCCAAGCATGGCGTTGTTAAGCACGTTGCTCCCGAGCCCGTCAAACACGTTGTTCTCACTCCCCCGTCT
ACCACTAGCCATAGCGTCGGCACAACATCACCTCCCGAGCCAACAGAGTCGCCTGTCGGCTCTGCTTCTGG
CCTTATTCAGAAGGCTCTCGAGATCATTGCCGATGAGATTGGAGTCGACATCTCTCAGCTCACTGACACTA
CACTCTTGGCTGACCTGGGTGTTGATTCGCTCATGTCACTGACCATTCTGGGCAACTTCCGAGAGGAACTC
GACTTGGACATTCCAGCAGCCCAGTTCTACGAGTTCTCAACAGTTCAGGATCTCAAGTCCTTCCTTGGAGC
CAATGACCAAGACTTCTCAAGCTCAAACTCTGAGGCGGAGAGTTCTGCGTCTAGCGCTGCTTCAACATCCC
CTAGTGATCACGGTGATGATGTCGTTGAGGAGGTCAAGCCTGTAGTGGCTGAGATACCTCGGTCAACCTCT
ACGATCTTGCAGGGCACCAAGCACTGTTCTCAAACACTCTTTCTGTTCCCCGACGGTGCTGGTTCAGCAAC
GTCTTATGTCACTCTCCCTTCTATTTCGTCCGACATGAGGGTCATTGGATTGAACAGTCCTTACCTTACCA
AGCCGCATGAGTTCAACTGTGCTCTCCAGGACATCACAGGCTCTTACTTGAACGAAGTGCGTCGACGTCAA
CCTCAAGGCCCTTACCATCTTGCAGGTTGGTCAGCAGGAGGCGTCTCGGCCTTCGACGCCGCTCGTCAGCT
TGTATCAGAGGGAGAGGTAGTCGAGAGCCTCATCCTGATTGACTCCCCCAACCCGGTTGGTCTTGGCAAGC
TGCCAAAGCGTATGTACGACTTCCTCGAAAAGTCGGGTATCTTCGGCGCTTTCGAGATGGGAGAGGAAGCT
CAAGCGCCGCCTGACTGGCTCTTCCAGCACTTCTGTGTCTTCATCGAGGCTTTGGATAGGTACGTGCCTGA
ACCGTTTGAGCATGGTATGGCGCCTAAGACGACCATCATCTGGGCTGCAGACGGCGTGTGCAAGAACCCTG
ATGATCCCCGTCCTGAGGCACAGCCTGACGACCCTCGGGGAATGAACTGGCTCTTGAATAACAGGGAGGAC
TTTGGGCCGAATGGTTGGGATGAGTTCATTGGCGCGGGAAACATCAGCACAATGGCGATTGAGAATGCCAA
CCACTTTACTATGATGAGGGAGCCAATTGCGTCGGCTCTCTGTGCGAAGATCCGGGAGACAATGGGCGTAA
AT**GCGGCCGCACTCGAGCACCACCACCACCACCACTGA**
```

Expression Construct (pEPks4) Protein Sequence:

```
MASSADVYVFGDQSTPVLDKLQALVRVKDNALLTSFLGEAFLAVRREIVSLSSLERKSIPEAESLSLLLEG
VRRSEPHAALDSAFVCIYEIGYYIDYLARSDKQHPPAAPSLLLGICTGSIAAAAVSCAKDVFEISRLGVEA
ATVAFRLGMHVRRRAENLGYSTPSSWSMILSSNQEELVSEALKEFSKEKNLTYSSRPYISATGPGFTTISG
PPSILESVKSCDTFSGKRLYPAPIYGPYHNSSSYSESSLEHGLASILEDVGFLENEMLIPIISCASGSRLD
QLSFGNLLKNVLSSALSQQIRMDLVTDALVETVSGTEATLIPVNAQTTVCSLADWLAKRGATTRIGPTLES
LTKDRAEPNLAPGDENKIAIIGFSGRFPEADNLDEFWDLLIRGLDVHKPVPEERFARDHYDPTGQRKNTSQ
VQYGCWLKSAGYFDTQFFHMSPKEAMQTDPAQRLALLTAYEALEMAGVVPDRTPSTQRNRVGVYYGTTSND
WGEVNSSQDVDTYYIPGANRAFIPGRVNYFFKFTGPSIAVDTACSSSLAAINLAITSLKNRDCNTAIAGGT
NVMTNPDNFAGLDRGHFLSRTGNCKAFNDGADGYCRADGIGTLILKRLPDAIADSDPIFGVILGAHTNHSA
ESVSITRPLADAQEYLFKKLLNETGIHPHDVSYVEMHGTGTQAGDAVEMRSVLNSFAFDHSRPRDKSLYLG
SVKANVGHAESASGVLAIIKVLLMMQKNTIPPHCGIKTKINQGFPKDLDHRGVRIALKDSVDWSRPEGGKR
RVLVNNFSAAGGNTSLLLEDGPAVHPARQHQDGDARTEHVVAVSARSTKALEENLKALEAYIANSWAPEGE
LLSQLSYTTTARRVHHSRRVAFVTNGLDDLRKSLLKAATDAGQVKGIPAVSPKVGFLFTGQGAQETAMAIG
YYKSFSSFRSDIHQLDSIATLQGLPSVLPLIHGTTPVEDLSAVVVQLGTCIIQISLARFWISLGITPQYVI
GHSLGEYAALQIAGVLSVNDAIFLCGHRAALLDKKCTAYTHGMVAVKAAADDLRQHISSDLKVEIACVNGA
EDTVLSGPNADIESLCGKLTQAGYKLHKLEIPFAFHSSQVDPILDDLEELASQVGFHEPKLPIVSPLLRTL
LTGDTLGPQYIRRHCRETVDFLGAIKMAESQGIMDRSGMCIEIGAHPILTRMVKSIIGQDFRCLASLRRKE
DHFKTLADSLCALHLAGFSVNWDEYHRDFASSRNVLQLPKYSWQLANYWMQYKYSWCLTKGDAPVENGPVG
AVVQARALRLSDSVHNVIEQVHGDKRSSITVESDMHDPSLLAIAQNHRVNGLTMAPSTLFADIAFTLAKHL
IQNHGLDTHTNLPSINNMAVEKALIVGETGPQLFRASLDMDWTTMRGSVRIFSVGANGKQTTLHAVCDVAV
ENPSSHRESWQSNAYLIQRGIKQLVQGASDGSAHMMRRGLLYKIFSNSVQYGSAFQGIEQVWFDSEGLEGT
GKVFMPSGKDTFALNPYCCDSLGHITGFIMNCSDSLDLDDHVYINHGWRTLRLVEPYQCDVQYQTYVKMQA
VGSDDSTYSGDVHVLRDGKIIGICGGVTFKKVARKVLEMLLPKPSGAKAKHGVVKHVAPEPVKHVVLTPPS
TTSHSVGTTSPPEPTESPVGSASGLIQKALEIIADEIGVDISQLTDTTLLADLGVDSLMSLTILGNFREEL
DLDIPAAQFYEFSTVQDLKSFLGANDQDFSSSNSEAESSASSAASTSPSDHGDDVVEEVKPVVAEIPRSTS
TILQGTKHCSQTLFLFPDGAGSATSYVTLPSISSDMRVIGLNSPYLTKPHEFNCALQDITGSYLNEVRRRQ
PQGPYHLAGWSAGGVSAFDAARQLVSEGEVVESLILIDSPNPVGLGKLPKRMYDFLEKSGIFGAFEMGEEA
QAPPDWLFQHFCVFIEALDRYVPEPFEHGMAPKTTIIWAADGVCKNPDDPRPEAQPDDPRGMNWLLNNRED
FGPNGWDEFIGAGNISTMAIENANHFTMMREPIASALCAKIRETMGVN**AAALEHHHHHH**
```

A.1.3. ACAS

Expression Construct (pEACAS) Gene Sequence:
ATGGATTTCACCCCGTCGACAGGCACTCCGGGGGGAGTTCCGGAGGATGAAGCTCACATATTTCACCAATGA
ATTTCCCTCAGATGACCTCCCCGGTCTTGCTCGTCAACTGCATCTTCACAGCAAAGATAGACGTCATCATA
TCCTGGCCCGTTTCCTACAAGATGCAACTCTGGCTATCCGCGAGGAGGTTGCGCAGCTGCCACCGGCGTTG
AAAGACTTGCTACCCCCCTTCGAATCCGTTCTGACTTTTGTCGAATACCCAGAGATGCGCAAGGGTCCTTT
GTGTGGATCGATAGATGGAGTGTTGCTCTCCACAGTTGAACTGGCCACATTCATTGGCTACTTTGAAGAGT
TCCCTGAAACATATGACTTCGAGTCCGCCCACACATACCTGGCCGGGCTGGGAATCGGGCTTTTATCAGCC
GCAGCAGTCTCACTTTCGCGGACCTTAGCAGACATTGCCTATGTCGGTTCCGAAGTTGTACGCATGGCCTT
CCGTCTCGGGGTTCTGGTCGACAACGTGTCTGAGCAGTTGCAACCTCGGGAACTGGCCAGTCATGGAACCC
CCGATAGTTGGGCGTATGTATTGCCGAACGTGACGCGAGAGGCGGTTCAGCAAGAGCTTGATGTCATCCAC
TCGGGAGAGGGAATACCCGAGACAGGCAAAATCTTCGTCAGTGCCTTCAGCCAGAGCTCTGTGACTGTGAG
CGGCCCTCCGTCGAGACTGAAAGATCTCTTCCGCACGTCCGACTTTTTCCGGGATCGAAGGTTCTTCTCAC
TCCCCGTATTCGGTGGATTATGCCATGCAAAGCACATCTACACAGAGACGCACGTACAGCAAGTCGTACGG
ACGAAGCCGATGGATATGTTAAGTGCCAGAGTGCTACCTCGAATTCCCATCTTTTCGCCCAGCAGCGGGAG
CCCTTTTCCTGCAGCCACCGCTACGGAGTTGTTTGAGGGGATCATCTCCGAAATCCTCACCCAAGTCATCC
AATGGGACAATGTGATCCAAGGTGCGCTTGACCAGATCAACATTCTCTCTCCGTCCGAGTTCCAAGTATTG
GTCTTCCGCATCTCGCTCCCTATTCACGATCTGATGGCTGCGGTAAACACGGAACTGAAGGGATTCCAAGC
CACTACGAAGGAAATCATGCCCTGGGTTTCGCACACGGCCAAAGACAGGATCCCCAGAGAACCATCGCAAT
CGAAGATTGCCATTGTCGGCATGTCCTGCCGGCTGCCCGGCGGAGCCACCAATACCGAGAAATTTTGGGAT
GTTCTGGAGCAAGGCTTGGACGTGTATCGGACGATCCCACCGGACCGTTTTGACGTTAATACTCATTACGA
TCCGGCGGGCAAAAGGGTTAACGCTAGTCACACTCCGTATGGTTGCTTTATTGAGGAGCCAGGCCTCTTCG
ATGCCCCGTTCTTCAACATGTCCCCTCGTGAAGCGCAACAGACCGACCCCATGCAGAGGTTAGCTTTGGTT
ACCGCGTATGAAGCCCTGGAACGAGCAGGCTATGTCCCGAACCGTACACCGGCGACCAACAAGCACCGCAT
TGGAACTTTCTATGGCCAAGCGAGTGATGACTACCGCGAAGTGAATACGGCGCAAGACGTCGATACATATT
TCATCACCGGCGGATGCCGCGCGTTTGGTCCCGGACGGATCAACTATTTCTTCAAGTTCTGGGGCCCGAGT
TACAGTATTGACACCGCCTGCTCGTCCAGCCTAGCAACCGTGGAGGCCGCTTGCACTTCGCTTTGGAACGG
CTCCACTGACACCGCAGTGGTAGGGGGCGTAAACGTCCTGACGAACTCGGACGCATTTGCGGGCTTAAGCC
GTGGTCACTTTTTGTCCAAGATCCCCGGGGCGTGCAAGACCTGGGACTGTAATGCCGACGGCTACTGCCGT
GCCGACGGGGTGATCTCGTTGGTAATGAAACGACTGGAGGATGCACAGGCTGATAATGACAACATCCTCGG
TGTTATCCTGGGAGCTGCGACGAATCACTCGGCCGACGCCGTTTCCATTACTCATCCCCACGCTGGTGCAC
AAGCACACCTGTTCCGGGATGTTCTGCGGAATGCCGGCGTCGATTCCCATGATGTCAGCTACGTTGAGCTG
CACGGTACCGGGACCCAGGCGGGCGACTTTGAGGAGATGAAATCCGTCACAGATGTGTTTGCCCCCCTCAC
GAAGCGCCGAAGCCCTAACCAGCCTCTCTATGTGGGCGCTGTCAAGGCAAATGTAGGCCACGGAGAGGCGG
TGGCTGGAGTCACTGCACTGCTGAAGGTTCTCTTGATGCTCCAGAAGAGCGTCATCCCTCCACATGTTGGC
ATTAAGAACAGCATTAACCCGCAGATACCCAAAGACCTGGACAAGCGCAACCTCCACATTCCGTACGAGAA
ACAGTCATGGAAGAGCACCCCGGGCAAGTCCCGGATTGCTGTTGTGAATAACTTTAGCGCGGCAGGTGGCA
ATACATCCGTTGTGCTTGAAGAGGGTCCAGTCACAGAGCTCACGGGAGTCGACCCCCGTCCATCCCACGTG
GTGGCCATCTCGGCCAAGAGTAAGGTTTCCTTGAAAGGAAACCTCGAGCGTTTCGCTGCCTATATCGATAC
CAATCCGGGCGTCTCCTTGTCTCACCTCTCCTACACGACTATGGCCCGACGCCATCACCATAACCATCGAC
TGGCTGCGGCGGTGTCCGACGCCGAGCAGTTAAAGAAGCAGCTGACCTCATGGATGCAGTCGGTCGAGTCG
CATAGACCTATCTCTGCTACTGGGCCTCCACCGGTGGCGTTTGCATTTACCGGGCAGGGGGCCTCATACAA
GTCCATGAACGTGGAGCTGTTCCATACCCTGCCAAGTTTCCGCGAGCAAATGATGCATTTGGACGCTCTTG
CTCAACAACAAGGGTTTCCCTCCTTCATACCCGCTATTGATGGAAGCCACCCTCAGGACCACGCCCACTCG
CAAGTCGTCACGCAGCTCGCTCTGACAGGCACTCAAATTGCGCTTGCTAAGTACTGGATGTCCTTGGGCGT
CCGGCCGGAGGTGGTCGTTGGCCATAGTCTAGGGGAATTTGCCGCACTACATATCGCCGGCGTCCTCTCGG
CCGGGGACACCCTCTTCCTGGTTGGTCGACGGGCCCAACTCCTTGAACAGCACTGCGTACAGGGCAGCTAT
CAGATGTTAGCGGTCCGCGCATCCGTGTCCCAAATCGAAGAGATTGCGGACGGCCGGCTGTATGAAGTTGC
TTGTATCAACGGACCGAAGGAGACCGTGCTTAGCGGGACCCGGCAGGAAATTCGCAACATCGCAGAGCATC
TCATGACCAAAGGGTACAAGTGCACCGCTCTAGAAGTTGCCTTTGCCGGTCATTCGGCCCAACTGGACCCT
ATTCTGGACACGTATGAACAGATTGCCACTAAAGGGGCGATATTCCACCCTCCCAACCTGCCGATTATCTC
GCCCTTGCTCGGAAAGGTCATCTTTGATGACAAGACGGTCAATGCAACCTACATGCGACGGGCGTCGCGTG
AGACAGTGAATTTCCACGCGGCTTTGGAGACGGCGCAAAGAATCTCCACTGTTGACGACACAACTGCGTGG
GTAGAGATCGGGCCCCATCCCGTTTGCATGGGCTTTATCCGGTCCACTCTGCAGTCCACGGCGCTGACAGT
CCCTTCCCTGCGTCGCGGGGAGGAAAGCTGGGTAACAATCACTCGGAGTCTCAGCTCCCTGCACTGTGCCG
GAGTGGAGGTTCATTGGAACGAGTTTCACCGGCCATTTGAGCAGGCGCTGAGACTTCTGGACCTGCCTACC
TACAGCTGGAACGACAAGAACTACTGGATTCAATATAATGGAGACTGGGCGCTCACGAAGGGCAACACCTT

```
CTATAGCTCTCAACAACAGAACTCCGCCGCCGTGGATGAATTGCCATCCGGCCCGAGGACTTCAACAGTTC
AGAAGATCGTGGAGGAGAGCTTCGACGGACGTGCAGCCAGGGTGGTCATGCAATCGGACCTGATGCAATCC
GATCTATTGGAGGCTGCATATGGTCATAAGATGAATGGGTGCGGTGTCGTGACATCGTCAATCCACGCTGA
TGTTGGATTTACCCTCGGTCAATACGTCTACAAGAAACTGAACCCCAATACCAAGGTTCCCGCGATGAACA
TGGCGAGCTTGGAGGTTCTGAAGGGACTGGTGGCTAACAAGAACACCGACAAGCCCCAGCGATTCCAAGTG
ACGGTTACTACCACCGACATCAATTCTAGGATCCTTCAGCTCGAATGGCGTAACGTCCATGCTCACGGCCC
TGCAGAGGAACCTTTTGCCAGCGCCAAGATCTACTACTGCGACGCTTCCGAATGGCTGCTGTCTTGGCGGC
CGACCCTCCATCTGGTCCAGGGACGCATCCAGGCTCTCGAGCGCCTGGCGGAAGCCGGTATTGCCAACCGG
TTCTCCGGACGCATGGCTTACAACCTTTTTGCCAACAGTCTCGTTGACTACGCGGGGAAATACCGCGGGAT
GCAGTCCGTGGTCATGCATGAACTCGAGGCGTTCGCGGATGTGACGCTGACGGTGGAGAAGGCGGGTACCT
GGACAGTCCCGCCGTACTTTATTGATAGCGTGGCGCATTTAGCCGGCTTCATCATGAATGTATCCGACGCC
ATCGACAACCAGAAGAACTTCTGTGTGACGCCCGGGTGGAATTCAATGCGGTTCGCCGCTCCCCTTGTTGC
CGGTGGCAGATACCGCTCCTACGTCAAGATGATCCCCACCGTGGAGGATGACAGTGTCTACCTGGGGGACG
TCTACATACTCCAAAACGACATGATTATCGGCATGGTGGGCGGGATCAAATTCCGGCGCTACCCCCCGACTC
CTTCTGAACCGTTTTTTCTCCCCCAGTGATGACAGCACCGCCAAGACAGCCGCTGGGGAGACCCCGCCCGC
ACCCACCACCACTGCTGCCACTGCTATCACTGCTGCCACTAGCACCACTAGCACCACTAGCACCGCATCCA
CAGGCCAGCCGCCCAAAGTGGATGAGACATCTCCGGTCGACTCTAATAGTACGGCAGCAAGAGCACTGGCG
TTGGTGGCAAAGGAGGCGGGGATGGAGGTCACCGATCTTCAGGATGACGCTATTTTCGCCAACCTGGGCGT
GGACAGTCTGATGAGCCTCGTAATTGCCGAAAAATTTCGTGAGGAACTTGGGGTCGTCGTGGCGGGGAGTT
TATTCTTGGAATATCCCACCGTGGGCGACCTGAAGAGCTGGCTTCTAGAATATTACAGC**GCGGCCGCACTC**
**GAGCACCACCACCACCACCACTGA**
```

Expression Construct (pEACAS) Protein Sequence:
```
MDFTPSTGTPGEFRRMKLTYFTNEFPSDDLPGLARQLHLHSKDRRHHILARFLQDATLAIREEVAQLPPAL
KDLLPPFESVLTFVEYPEMRKGPLCGSIDGVLLSTVELATFIGYFEEFPETYDFESAHTYLAGLGIGLLSA
AAVSLSRTLADIAYVGSEVVRMAFRLGVLVDNVSEQLQPRELASHGTPDSWAYVLPNVTREAVQQELDVIH
SGEGIPETGKIFVSAFSQSSVTVSGPPSRLKDLFRTSDFFRDRRFFSLPVFGGLCHAKHIYTETHVQQVVR
TKPMDMLSARVLPRIPIFSPSSGSPFPAATATELFEGIISEILTQVIQWDNVIQGALDQINILSPSEFQVL
VFRISLPIHDLMAAVNTELKGFQATTKEIMPWVSHTAKDRIPREPSQSKIAIVGMSCRLPGGATNTEKFWD
VLEQGLDVYRTIPPDRFDVNTHYDPAGKRVNASHTPYGCFIEEPGLFDAPFFNMSPREAQQTDPMQRLALV
TAYEALERAGYVPNRTPATNKHRIGTFYGQASDDYREVNTAQDVDTYFITGGCRAFGPGRINYFFKFWGPS
YSIDTACSSSLATVEAACTSLWNGSTDTAVVGGVNVLTNSDAFAGLSRGHFLSKIPGACKTWDCNADGYCR
ADGVISLVMKRLEDAQADNDNILGVILGAATNHSADAVSITHPHAGAQAHLFRDVLRNAGVDSHDVSYVEL
HGTGTQAGDFEEMKSVTDVFAPLTKRRSPNQPLYVGAVKANVGHGEAVAGVTALLKVLLMLQKSVIPPHVG
IKNSINPQIPKDLDKRNLHIPYEKQSWKSTPGKSRIAVVNNFSAAGGNTSVVLEEGPVTELTGVDPRPSHV
VAISAKSKVSLKGNLERFAAYIDTNPGVSLSHLSYTTMARRHHHNHRLAAAVSDAEQLKKQLTSWMQSVES
HRPISATGPPPVAFAFTGQGASYKSMNVELFHTLPSFREQMMHLDALAQQQGFPSFIPAIDGSHPQDHAHS
QVVTQLALTGTQIALAKYWMSLGVRPEVVVGHSLGEFAALHIAGVLSAGDTLFLVGRRAQLLEQHCVQGSY
QMLAVRASVSQIEEIADGRLYEVACINGPKETVLSGTRQEIRNIAEHLMTKGYKCTALEVAFAGHSAQLDP
ILDTYEQIATKGAIFHPPNLPIISPLLGKVIFDDKTVNATYMRRASRETVNFHAALETAQRISTVDDTTAW
VEIGPHPVCMGFIRSTLQSTALTVPSLRRGEESWVTITRSLSSLHCAGVEVHWNEFHRPFEQALRLLDLPT
YSWNDKNYWIQYNGDWALTKGNTFYSSQQQNSAAVDELPSGPRTSTVQKIVEESFDGRAARVVMQSDLMQS
DLLEAAYGHKMNGCGVVTSSIHADVGFTLGQYVYKKLNPNTKVPAMNMASLEVLKGLVANKNTDKPQRFQV
TVTTTDINSRILQLEWRNVHAHGPAEEPFASAKIYYCDASEWLLSWRPTLHLVQGRIQALERLAEAGIANR
FSGRMAYNLFANSLVDYAGKYRGMQSVVMHELEAFADVTLTVEKAGTWTVPPYFIDSVAHLAGFIMNVSDA
IDNQKNFCVTPGWNSMRFAAPLVAGGRYRSYVKMIPTVEDDSVYLGDVYILQNDMIIGMVGGIKFRRYPRL
LLNRFFSPSDDSTAKTAAGETPPAPTTTAATAITAATSTTSTTSTASTGQPPKVDETSPVDSNSTAARALA
LVAKEAGMEVTDLQDDAIFANLGVDSLMSLVIAEKFREELGVVVAGSLFLEYPTVGDLKSWLLEYYS**AAAL**
**EHHHHHH**
```

A.1.4. CTB1

Expression Construct (pECTB1) Gene Sequence:
```
ATGGAAGATGGCGCCCAAATGCGCGTGGTAGCCTTCGGGGACCAAACCTATGACTGCTCAGAAGCGGTGAG
CCAGCTGTTGCGCGTACGTGACGATGCGATCGTGGTCGATTTTTTAGAACGTGCACCGGCAGTTCTTAAAG
CGGAACTGGCGCGCTTGTCCAGCGAGCAGCAGGAAGAGACTCCGCGCTTTGCAACCCTGGCCGAGTTAGTC
CCTCGTTATCGGGCCGGCACGCTGAATCCGGCCGTGAGTCAGGCTCTCACTTGTATTGCTCAACTGGGTCT
```

```
GTTCATCCGCCAGCATTCGAGCGGCCAAGAAGCGTATCCCACCGCCCACGATTCTTGCATTACCGGCGTTT
GTACCGGCGCATTGACGGCGGTTGCGGTCGGCTCCGCAAGCTCAGTCACAGCTCTGGTTCCATTAGCGCTG
CATACGGTGGCAGTTGCGGTACGTCTGGGAGCTCGTGCTTGGGAGATTGGTAGCTGCCTCGCTGATGCCCG
TCGGGGTGCGAACGGGCGTTACGCATCGTGGACAAGTGCCGTGGGTGGAATTTCTCCGCAGGATCTTCAGG
ACCGCATCTCTGCGTACACCGCCGAACAGGCGCTTGCATCCGTGTCCGTTCCATATCTCTCGGCAGCGGTC
GGCCCGGGCCAATCAAGTGTGTCGGCTGCTCCCGTCATCCTCGACGCGTTCCTCAGCACACTTCTCCGACC
GCTGACAACAACGCGACTTCCCATTACCGCTCCGTATCACGCACCTCACCTCTTCACTGCGAAAGATGTGC
AGCATGTCACGGACTGTCTGCCTCCCAGCGAGGCATGGCCGACAGTGCGAATTCCCATCATCAGCTTCTCC
CGCGATGAGGCAGTGTCGCGTGGcTGCAAGCTTCCCGGCCGCCATGAGTGAGGCCGTGCGAGACTGCTTGA
TCCGCCCCATTGCCCTCGACCGCATGGCAGTGAGCATTGCCAATCATGCTCGAGATCTTGGCAAAGACTCG
GTGCTTCCGTCACCCATCGCTCTGTCATTCAGCGACAAGCTGGGCCCACAAGTAAACAGCCACTTGCCAGG
CGCGAAGGCACCCACGCCTGAGCTGACATCAAAATCTATCCCGTCTGCTATAGGAGCGGAACAGCAGCCGA
TGGCAAAATCCCCCATTGCTATCCTTGCTGCCTCTGGTCGTTTCCCACAATCTTCATCCATGGACCAGTTC
TGGGATGTACTTATTAACGGCGTCGATACGCACGAACTTGTACCTCCAACCCGTTGGAATGCAGCTACTCA
TGTCTCCGAAGATCCTAAAGCAAAGAATGTGTCTGGCACAGGTTTCGGTTGCTGGCTTCACGAAGCCGGTG
AATTTGATGCCGCTTACTTCAACATGAGTCCGCGTGAAGCGCCGCAGGTTGATCCAGCACAGCGTCTGGCC
CTGCTGACCGCAACTGAAGCTCTGGAACAGGCAGGTGTTGTTCCGAACCGTACCTCTAGCACCCAGAAGAA
CCGTGTAGGCGTGTGGTACGGCGCTACTAGCAACGACTGGATGGAGACTAACTCTGCGCAGAACGTGGATA
CCTACTTTATCCCGGGTGGCAATCGCGCATTCATCCCAGGAAGAGTGAACTACTTCCACAAGTTTAGTGGC
CCATCTTACACGATCGATACAGCCTGCAGCTCCAGTTTGGCAGCATTGCACATGGCGTGCAACGCACTTTG
GCGAGGCGAAGTCGATACAGCCATCGTAGGTGGGACCAACGTCCTCACGAATCCAGACATGACGGCGGGAC
TCGACGCGGGTCACTTCTTGTCAAGGTCTGGAAATTGTAAGACGTTCGACGACGAAGCCGATGGCTATTGC
AGAGGTGAGGCAGTGGTGACCCTCATTCTCAAACGGCTGCCAGACGCGCAAGCGGACAAAGATCCAATCCA
GGCTTCAATTCTGGGAATTGCCACTAATCACTCAGCCGAGGCCGCCTCTATCACTAGGCCTCATGCCGGAG
CGCAGCAAGACTTGTTCCAACAAGTTCTCACGGAGACAGGTCTTACCGCGAACGACATTAGTGTGTGCGAG
ATGCATGGTACTGGCACCCAGGCTGGAGACAGTGGTGAAACAACGTCCGTCGTGGAGACCCTAGCGCCTTT
GAACCGATCCGGCTCTGCTGTGCGAACAACACCTCTCTACATTGGCGCAGTTAAGTCCAATGTGGGTCATG
CTGAGTCCGCAGCTGGGGTCAGCAGTCTGGCCAAGATCTTGCTTATGCTCAAGCATTCCAAGATCCCTCCT
CATGTGGGCATCAAAACGAAGCTGAATCACCGGCTACCAGACCTAGCTGCACGAAATACACACATAGCGCG
GTCTGAGGTACCTTGGCCTCGGCCAAAGAATGGCAAGCGTCGTGTTCTGCTCAATAACTTCTCGGCCGCTG
GAGGTAACACGTGCCTTGTGCTTGAGGATGCGCCCGAGCCCGAGGACTCTCAAGAAGTCGACCCTAGAGAA
CATCACATTGTTGCACTCTCTGCTAAAACACCTGATTCAATGGTGAACAACCTCACGAACATGATAACCTG
GATCGACAAGCACTCTGGAGACAGCCTCGCCACCTTGCCTCAACTGTCTTACACGACAACTGCACGAAGAG
TGCACCATAGACACAGAGCCGTAGCTACCGGCACCGATCTGCTGCAAATCCGTTCGTCGCTTCAAGAACAG
CTTGACCGCCGGGTGTCCGGCGAGAGAAGTATCCCTCACCCGCCCAATGGACCTAGCTTTGTCCTTGCTTT
CACTGGCCAAGGCTCGGCGTTCGCAGGTATGGGTGTAGATCTCTACAAACGTTTCGCCTCATTTCGGTCAG
ACATTGCCCGCTATGATCAGATCTGCGAGGGTATGAGCCTGCCCTCGATCAAAGCTATGTTCGAGGACGAG
AAAGTGTTCTCCACAGCTTCACCAACTTTGCAGCAGCTCACGCATGTCTGTTTTCAGATGGCCCTGTACAG
ACTATGGAAGTCCCTCGGTGTACAGGCGAAGGCGGTAGTTGGCCACAGTTTGGGCGAGTATGCTGCACTCT
ACGCCGCTGGAGTGCTATCGCAATCCGATACGCTCTACCTGGTGGGGCGGCGTGCACAGCTGATGGAGAAA
CATCTTTCGCAAGGCACACATGCAATGCTGGCCGTCCGTGCGAAAGAAGAAGCCATTGTCGCAGCGATTGA
CGGGCCTCCAGGAGAGGCATACGAGTTCTCTTGCCGCAATGGCGAACAGCGAAACGTTCTCGGAGGCACCG
TTGCTCAAATCCAGGCGGCGAAAGCCGCGCTTGAGGCTAAGAAGATTCGATGCCAGTACTTGGACACCCCG
ATGGCATTCCACACTGGCCAAGTCGATCCGATTCTGCCCGAGCTCTTGCAGGTCGCTGCGGCATGCTCCAT
CCAGGATCCCCAAATTCCTGTCATCTCCCCAGCATATGGCAAAGTGATCAGGTCTGCCAAGGACTTCCAAC
CAGAGTACTTCACCCATCACTGCCGCAGTTCTGTCAACATGGTCGATGCTCTCCAAAGTGCAGTCGAAGAA
GGCTTACTTGACAAGAACGTTATTGGCCTCGAGATTGGCCCTGGCCCGGTCGTCACGCAGTTCGTCAAGGA
AGCTGTGGGCACAACCATGCAGACCTTTGCGTCCATCAATAAAGACAAAGACACATGGCAGCTCATGACGC
AAGCGCTAGCTAAGTTCTATCTTGCAGGTGCCAGTGTCGAGTGGTCGCGCTATCACGAAGACTTCCCTGGA
GCTCAGAAAGTCCTTGAGCTCCCGGCTTATGGCTGGGCCTTGAAAAACTACTGGCTGCAATATGTCAACGA
CTGGTCTCTGAGGAAGGGCGATCCAGCCGTGGTTGTTGCCGCGTCAAATTTGGAACTCTCTTCGTCGATAC
ACAAGGTCATAACAAACACAATCACCGCCAACAGCGACGGCGAGCTTGTCGTGGACGCAGACCTCAGTCGA
GAGGACCTGCATCCCATGGTCCAGGGGCATCAGGTCTACGGGGTTCCACTGTGCACACCTTCCGTGTACGC
GGACATCGCTCTGACACTCGGCGAGTATATTCGACAGGTCATCAAGCCAGGCGAGGTTGCACAGACATCCG
TTGAAGTAGCAGAGATGAACATTCAAAGCGCACTGGTGGCTAACAACACGGGCAGAGTGCAACTCCTTCGC
ACGTGTGCCAAGTTTGACCCCAAGGCCCAGGTAGCGTCATGCACGTTCTCTAGTATCGTCGAGCAGCATGC
TAATTGCAAGATCCGGTTCGGCAGTCTCGAGAAGGAGAAGACCGCGCTCAAGAGTGCTGCACTAGCTGCCC
AAGCCAGCATGGCCGCTCTGAAAACACAAGTTGGCCAGGATGACAACACATACCGCTTCAGCAAAGGCATG
```

```
ATTTACAAGATGATCGGCCAATTAGCTGACTTCGATGAGAAGTACCGCGGGCTCTGCGCGATCACACTCGA
CAACGACGCCATGGAAGCCTCGGGCAAAGTATCATTCAAGGGCATTCCAAACGAGGGCAAATTCCACTCTA
GCCCGGCTTATCTCGACGCGCTGTCGCAACTTGGCGGATTCGTCATGAACGCGAACGAGGGTGTGGATCTT
GAGAAGAAGTCTTTGTCAATCACGGCTGGGGTTCCATGCGTTTCTTCGCCGCTCTGGATCCAGCAATGAC
TTACTACACTCATGTGAAGATGACCCAAGGCAAAGACAAATTGTGGACTGGCGATGTCTTGATCTTCGACG
ACAAGCAAGCATTGATCGGCATTGTTGGGGGAGTGGCGTTGCAGGGCGTGCCCAAGCGACTTATGCATTAC
ATTGTTACAGCTGCCAACAAGAAAGCTTCCGGCCCGCCGACAGAGAAGAAGACCTCTAGCCCGCCAGTCGA
AAAGAAAGCCAGCGCGCCAGTCGCGCCCACGAGGCCAGCGATCCAGCGTAAGAATGCTTCGATTCCTCCAC
CTGCAACCCAAGTGACTCCGCAAAACAAGACCATCAAGACGCCAAGTGTGTCGGCACTGATAGCCCCGGCC
CTCGAGATTGTTAGCGAGGAGATCCGGATGCCAATCGACGAGCTCAAGGATGATATCGACTTCACCGATGC
TGGTCTTGACAGTCTGCTTTCCTTGGTAATTAGCAGTCGCATGCGGGACCAGCTGGGCATCGAATTCGAAT
CCGCGCAGTTCATGGAGATTGGATCTATCGGTGGACTCAAGGAGTTCTTGACCAGGCTCAGTCCCCCAGTA
GCAGTCGCCGTTGCCACTGCCGTGGAAATTGTCAAGGAGGAAGCGCTCACCTCATTGGAAGAGCTTACTGA
TCCATCCCCTAACGAGATCGGCACTGTCTGGCGCGATGCCCTCAAGATCCTGTCTGAAGAGAGTGGCCTCA
CTGATGAGGAGTTGACTGATGACACAAGTTTCGCCGACGTGGGCGTTGATAGCCTCATGAGTCTTGTGATC
ACCAGTCGCCTACGGGATGAATTGGACATCGACTTCCCGGACCGAGCATTATTCGAAGAATGCCAGACTAT
ATTTGACCTTCGCAAGAGGTTCTCTGGGTCAACGGAAAGCTTCGACTCGACGACGACCAAGCCCAGCGCTG
GTGATGCGACGCCACCTCTGACCGATTCCAGCGCGTCATCTCCGCCCTCCTCCGAGTTCGATGGCGAGACG
CCGATGACTGATCTGGACGAGGTGTTCGATTCTCCCCCAGCGCAGAAGAGGATACCATCCCCGCCCAAAGG
ACGAATCCCGCCTGCATGGTCGATGTATTTGCAAGGCTCACAGAAGCGGTCGAAGGAGATTCTTTTCTTGT
TTCCAGACGGCGCTGGCGCCGCAACTTCTTACTTGTCTTTACCTCGTTTGGGTGAAGACATTGGCGTAGTC
GCCTTCAATTCGCCTTTCATGAAGACACCGCACAAGTTTGCTGATCATACCTTACCGGACGTCATCGCGTC
CTATGTAGAAGGCATTCGAGGCCGTCAAGCGCAAGGCCCGTATCATCTGGGCGGCTGGTCTGCTGGTGGTA
TTCTGGCCTATGCCGTAGCCCAAGAACTCATCGCAGCTGGCGAGGAGGTTTCGACACTCCTCCTCATCGAC
TCGCCTTCGCCAACCAAAGGCCTAGATCGCCTTCCAACACGATTCTTCGATCACTGCACGAACGTCGGACT
CTTTGGAACGGAGCTCTCCAGAGGCAGTGGAGGTCCCAACAAGACACCCGAATGGCTGATGCCTCATTTCA
GAGCTAGTATTGAACTGCTACACGGCTACCACGCTCCTCCTATGAAGCTTGGCAACAAGACGAAAGTCATG
GTCATCTGGGCCGGCGAATGTGCATTTGATGGCGTTCGCTATGCTCACATACCGCCATCTGCAGGCGATAC
CGACGAAGATACTGAGGGTATGAAGTTTCTGACGGAGAAGAGGAAAGATTTTGGGGCCACAGAATGGGCAA
GTCTGTTCCCTGGCACTGACGTTGACGCGAGAGTTGTGGAGCGAGCATCACTTCAGCATGATGCGTGAT
TCTGGTGCACAGATGCTTGTTGAGCACATGCGAGATGGTTTGGGGATTGTTTCATCA**GCGGCCGCACTCGA**
**GCACCACCACCACCACCACTGA**
```

```
MEDGAQMRVVAFGDQTYDCSEAVSQLLRVRDDAIVVDFLERAPAVLKAELARLSSEQQEETPRFATLAELV
PRYRAGTLNPAVSQALTCIAQLGLFIRQHSSGQEAYPTAHDSCITGVCTGALTAVAVGSASSVTALVPLAL
HTVAVAVRLGARAWEIGSCLADARRGANGRYASWTSAVGGISPQDLQDRISAYTAEQALASVSVPYLSAAV
GPGQSSVSAAPVILDAFLSTLLRPLTTTRLPITAPYHAPHLFTAKDVQHVTDCLPPSEAWPTVRIPIISFS
RDEAVSRGASFPAAMSEAVRDCLIRPIALDRMAVSIANHARDLGKDSVLPSPIALSFSDKLGPQVNSHLPG
AKAPTPELTSKSIPSAIGAEQQPMAKSPIAILAASGRFPQSSSMDQFWDVLINGVDTHELVPPTRWNAATH
VSEDPKAKNVSGTGFGCWLHEAGEFDAAYFNMSPREAPQVDPAQRLALLTATEALEQAGVVPNRTSSTQKN
RVGVWYGATSNDWMETNSAQNVDTYFIPGGNRAFIPGRVNYFHKFSGPSYTIDTACSSSLAALHMACNALW
RGEVDTAIVGGTNVLTNPDMTAGLDAGHFLSRSGNCKTFDDEADGYCRGEAVVTLILKRLPDAQADKDPIQ
ASILGIATNHSAEAASITRPHAGAQQDLFQQVLTETGLTANDISVCEMHGTGTQAGDSGETTSVVETLAPL
NRSGSAVRTTPLYIGAVKSNVGHAESAAGVSSLAKILLMLKHSKIPPHVGIKTKLNHRLPDLAARNTHIAR
SEVPWPRPKNGKRRVLLNNFSAAGGNTCLVLEDAPEPEDSQEVDPREHHIVALSAKTPDSMVNNLTNMITW
IDKHSGDSLATLPQLSYTTARRVHHRHRAVATGTDLLQIRSSLQEQLDRRVSGERSIPHPPNGPSFVLAF
TGQGSAFAGMGVDLYKRFASFRSDIARYDQICEGMSLPSIKAMFEDEKVFSTASPTLQQLTHVCFQMALYR
LWKSLGVQAKAVVGHSLGEYAALYAAGVLSQSDTLYLVGRRAQLMEKHLSQGTHAMLAVRAKEEAIVAAID
GPPGEAYEFSCRNGEQRNVLGGTVAQIQAAKAALEAKKIRCQYLDTPMAFHTGQVDPILPELLQVAAACSI
QDPQIPVISPAYGKVIRSAKDFQPEYFTHHCRSSVNMVDALQSAVEEGLLDKNVIGLEIGPGPVVTQFVKE
AVGTTMQTFASINKDKDTWQLMTQALAKFYLAGASVEWSRYHEDFPGAQKVLELPAYGWALKNYWLQYVND
WSLRKGDPAVVVAASNLELSSSIHKVITNTITANSDGELVVDADLSREDLHPMVQGHQVYGVPLCTPSVYA
DIALTLGEYIRQVIKPGEVAQTSVEVAEMNIQSALVANNTGRVQLLRTCAKFDPKAQVASCTFSSIVEQHA
NCKIRFGSLEKEKTALKSAALAAQASMAALKTQVGQDDNTYRFSKGMIYKMIGQLADFDEKYRGLCAITLD
NDAMEASGKVSFKGIPNEGKFHSSPAYLDALSQLGGFVMNANEGVDLEKEVFVNHGWGSMRFFAALDPAMT
YYTHVKMTQGKDKLWTGDVLIFDDKQALIGIVGGVALQGVPKRLMHYIVTAANKKASGPPTEKKTSSPPVE
KKASAPVAPTRPAIQRKNASIPPPATQVTPQNKTIKTPSVSALIAPALEIVSEEIRMPIDELKDDIDFTDA
```

GLDSLLSLVISSRMRDQLGIEFESAQFMEIGSIGGLKEFLTRLSPPVAVAVATAVEIVKEEALTSLEELTD
PSPNEIGTVWRDALKILSEESGLTDEELTDDTSFADVGVDSLMSLVITSRLRDELDIDFPDRALFEECQTI
FDLRKRFSGSTESFDSTTTKPSAGDATPPLTDSSASSPPSSEFDGETPMTDLDEVFDSPPAQKRIPSPPKG
RIPPAWSMYLQGSQKRSKEILFLFPDGAGAATSYLSLPRLGEDIGVVAFNSPFMKTPHKFADHTLPDVIAS
YVEGIRGRQAQGPYHLGGWSAGGILAYAVAQELIAAGEEVSTLLLIDSPSPTKGLDRLPTRFFDHCTNVGL
FGTELSRGSGGPNKTPEWLMPHFRASIELLHGYHAPPMKLGNKTKVMVIWAGECAFDGVRYAHIPPSAGDT
DEDTEGMKFLTEKRKDFGATEWASLFPGTDVDARVVESEHHFSMMRDSGAQMLVEHMRDGLGIVSS**AAALE
HHHHHH**

A.1.5. wA

Expression Construct (pEwA) Gene Sequence:
ATGGAGGATCCGTACCGTGTTTACTTGTTCGGTGATCAAACTGGCGACTTTGAAGTTGGCTTACGTCGTCT
GCTTCAAGCTAAGAATCATAGCCTCTTGTCGTCCTTCCTGCAGCGTTCCTACCACGCCGTACGCCAGGAAA
TTTCACACTTACCGCCAAGCGAACGTTCTACGTTTCCACGCTTCACCTCGATCGGCGATTTACTGGCGCGC
CATTGCGAGTCTCCCGGTAACCCGGCGATTGAATCAGTCTTAACATGTATTTATCAACTGGGATGTTTTAT
CAATTATTATGGGGACCTCGGACACACCTTTCCTAGCCATAGCCAGAGCCAGCTCGTAGGTCTGTGTACGG
GTCTGCTTAGTTGCGCAGCCGTTTCTTGCGCTAGTAACATTGGCGAATTGCTTAAGCCGGCGGTCGAAGTG
GTGGTGGTGGCCCTGCGGCTGGGCCTGTGCGTCTATCGCGTGCGCAAATTGTTCGGGCAGGACCAGGCAGC
GCCTCTGAGCTGGTCAGCACTGGTTTCTGGCCTGTCCGAGTCGGAAGGTACCAGTCTGATCGATAAATTTA
CCCGGCGTAACGTTATCCCTCCTTCGTCTCGGCCGTACATTAGCGCAGTATGCGCAAACACGTTGACCATT
AGTGGTCCCCCGGTTGTTCTCAACCAGTTTCTGGATACTTTCATTTCAGGAAAGAATAAAGCTGTTATGGT
GCCAATCCACGGCCCGTTTCATGCCTCTCATCTATATGAGAAGCGGGACGTGGAGTGGATTCTTAAGTCCT
GCAACGTTGAAACAATCCGCAACCATAAGCCCCGTATTCCGGTTCTGTCAAGCAATACTGGGGAGCTCATC
GTTGTTGAAAATATGGAAGGTTTTCTCAAAATTGCTTTAGAAGAGATTCTGCTTCGCCAAATGTCTTGGGA
TAAGGTGACAGATTCATGCATATCGATTTTGAAGTCAGTCGGCGATAACAAACCCAAGAAGTTGCTTCCAA
TCTCATCTACGGCTACGCAAAGTTTATTCAACTCACTCAAGAAATCAAACCTAGTGAATATCGAAGTGGAT
GGAGGAATTAGCGACTTCGCGGCAGAAACTCAACTTGTCAACCAGACAGGCAGGGCTGAACTTTCAAAGAT
AGCCATTATTGGGATGTCGGGTCGGTTTCCTGAAGCTGATAGTCCACAGGACTTTTGGAACCTCCTATACA
AAGGCCTTGACGTGCATCGAAAAGTCCCTGAAGACAGGTGGGATGCTGATGCGCACGTTGACCTCACCGGC
ACAGCAACGAATACAAGCAAGGTCCCTTACGGATGCTGGATTCGGGAGCCTGGCTTATTCGATCCTCGCTT
CTTCAACATGTCACCCCGTGAAGCACTTCAAGCAGATCCTGCTCAACGGCTTGCATTGCTCACAGCGTATG
AGGCATTGGAAGGGGCTGGATTTGTTCCAGATAGTACGCCATCGACGCAAAGGGATAGAGTTGGTATCTTC
TACGGGATGACCAGTGATGACTACAGAGAGGTCAACAGTGGTCAGGATATCGATACCTACTTTATTCCAGG
GGGCAACCGTGCTTTTACTCCCGGCCGGATCAACTATTACTTCAAATTCAGCGGGCCCAGTGTGAGTGTTG
ATACAGCTTGCTCTTCTAGCCTCGCGGCTATCCACTTGGCCTGCAATTCCATTTGGAGAAATGACTGTGAC
ACGGCTATTACTGGTGGTGTAAACATCCTGACCAACCCAGACAACCATGCTGGTCTGGACCGTGGCCATTT
CCTTTCCAGAACCGGAAACTGCAACACGTTCGATGACGGAGCCGATGGATATTGCAGAGCAGATGGTGTTG
GCACTGTCGTTCTGAAGCGGCTTGAAGATGCTTTGGCAGATAATGATCCAATTCTTGGTGTCATTAACGGC
GCATATACGAACCACTCAGCAGAGGCAGTCTCAATCACCCGTCCCCATGTCGGCGCACAAGCATTTATCTT
TAAGAAGTTGCTAAATGAAGCAAACGTTGACCCCAAGAACATAAGCTACATCGAGATGCATGGAACTGGTA
CCCAGGCTGGGGATGCGGTGGAAATGCAATCGGTCTTGGATGTCTTTGCACCAGACCACAGACGCGGGCCA
GGACAGTCCCTACATCTTGGCTCTGCAAAATCAAATATAGGTCACGGTGAATCAGCGTCAGGTGTGACTTC
TCTAGTCAAGGTGCTGCTGATGATGAAGGAAAACATGATACCTCCGCATTGTGGCATCAAGACGAAGATCA
ATCATAATTTCCCAACTGACCTGGCGCAAAGGAATGTCCATATCGCACTGCAGCCCACAGCTTGGAATCGA
CCTTCATTTGGCAAACGCCAGATATTCTTGAACAATTTTTCCGCAGCCGGTGGAAATACCGCGCTTTTGCT
GGAAGATGGGCCCGTTTCTGATCCGGAAGGGGAGGACAAACGGCGAACGCACGTTATCACGCTTTCTGCTC
GATCTCAAACCGCTCTTCAAAACAATATCGACGCACTGTGTCAGTACATCAGTGAGCAAGAAAGACATTC
GGTGTCAAGGACAGCAATGCCCTGCCGAGTTTGGCGTATACTACGACCGCACGCCGCATTCACCACCCTTT
CCGAGTCACTGCTATTGGGTCTAGCTTTCAGGAAATGCGCGATTCTCTAATTGCGAGCTCCCGAAAGGAAT
TTGTCGCGGTGCCTGCAAAGACCCCTGGTATCGGGTTCCTCTTTACTGGCCAGGGTGCCCAATATGCAGCC
ATGGGCAAACAATTGTATGAGGACTGTTCACACTTTCGATCCGCCATAGAGCACTTGGATTGCATCAGCCA
AGGCCAGGATCTTCCATCAATTCTTCCTCTTGTTGATGGCAGCTTGCCTTTGAGCGAACTGAGTCCTGTTG
TTGTACAACTGGGGACTACATGCGTGCAGATGGCTTTGTCCAGTTTTTGGGCATCATTAGGAATAACGCCA
TCCTTCGTTCTTGGGCATAGTCTTGGGGACTTTGCTGCTATGAACGCTGCCGGAGTTCTATCGACCAGTGA
TACCATCTATGCTTGTGGTCGTCGTGCCCAGCTCCTCACAGAGCGATGCCAACCGGGAACACATGCCATGC
TGGCCATTAAAGCCCCACTAGTCGAGGTCAAACAGCTCTTGAACGAGAAGGTTCATGATATGGCATGTATT

```
AATTCGCCTTCAGAAACCGTCATTAGCGGTCCTAAGTCTAGCATTGATGAGCTCTCCCGGGCCTGCTCAGA
GAAAGGACTGAAATCAACTATTCTGACAGTGCCCTACGCTTTCCATTCTGCCCAGGTTGAACCTATTTTGG
AAGACCTTGAGAAGGCTCTGCAGGGAATCACCTTCAACAAACCGTCCGTGCCGTTTGTTTCAGCGTTGCTC
GGGGAAGTAATTACGGAAGCGGGCTCTAACATCCTAAACGCTGAATACCTGGTAAGACACTGCCGGGAAAC
TGTCAATTTTCTGAGTGCGTTTGAAGCAGTGAGAAATGCAAAGCTCGGTGGTGATCAGACTCTCTGGCTCG
AGGTCGGCCCGCATACGGTTTGTTCCGGAATGGTCAAGGCCACGCTTGGACCACAAACAACCACCATGGCG
TCTCTCCGCCGCGACGAGGACACATGGAAAGTCCTTTCCAATAGCCTGTCAAGTCTTTATTTGGCGGGCGT
TGACATCAACTGGAAGCAATATCATCAGGACTTCTCCTCCAGCCACAGAGTACTTCCGCTACCAACCTACA
AGTGGGATTTGAAGAATTACTGGATTCCCTACAGGAACAACTTCTGTCTCACCAAGGGCTCTTCTATGTCT
GCCGCGAGCGCGTCTCTACAGCCGACGTTTTTAACCACCAGCGCGCAACGCGTTGTAGAAAGCCGCGATGA
TGGTTTGACTGCTACAGTGGTGGTGCATAACGACATTGCAGATCCGGACCTGAACCGTGTCATCCAGGGCC
ATAAAGTTAATGGTGCGGCCCTGTGTCCGAGCTCTCTTTATGCAGATAGTGCTCAAACCCTTGCAGAATAC
CTCATTGAAAAGTACAAGCCAGAGCTCAAAGGCTCAGGTTTGGACGTATGCAATGTAACGGTTCCCAAACC
TCTTATTGCAAAAACTGGAAAGGAGCAATTCAGAATTTCGGCGACTGCCAATTGGGTGGACAAACACGTAT
CTGTGCAAGTCTTCTCCGTTACCGCGGAGGGCAAAAAGCTCATCGACCACGCGCACTGCGAAGTCAAACTT
TTCGACTGCATGGCCGCTGATCTTGAATGGAAGAGAGGTTCGTATCTGGTCAAGCGAAGCATTGAGCTTCT
CGAAAATAGTGCTGTGAAAGGAGATGCCCATCGCCTGCGGAGAGGAATGGTCTATAAACTCTTCAGCGCTT
TGGTCGACTACGATGAAAACTACCAATCTATCCGTGAAGTTATCCTTGATAGTGAGCATCACGAGGCGACG
GCTCTGGTCAAGTTCCAGGCACCGCAAGCCAACTTTCACCGTAACCCATACTGGATCGACAGTTTTGGGCA
TTTGTCTGGCTTCATAATGAATGCTAGCGATGGAACAGACTCCAAAAGTCAAGTCTTCGTTAATCATGGAT
GGGACTCCATGCGGTGCTTGAAGAAGTTCTCAGCAGATGTCACCTACCGTACCTATGTCAGGATGCAACCT
TGGCGCGACTCGATCTGGGCAGGAAATGTCTACATATTCGAGGGTGATGATATTATCGCAGTATTTGGCGG
TGTTAAGTTCCAAGCCCTATCGCGCAAGATCCTCGACATTGCTCTTCCGCCAGCAGGACTGTCGAAGGCGC
AGACCTCGCCAATCCAGTCCTCAGCTCCCCAGAAGCCGATTGAAACTGCAAAACCCACCTCGCGGCCAGCC
CCTCCGGTGACAATGAAATCATTCGTCAAGAAGTCCGCTGGCCCAAGTGTGGTAGTGCGAGCGCTCAATAT
CCTAGCATCAGAAGTAGGACTCTCAGAATCCGACATGTCTGACGATCTGGTGTTTGCAGACTATGGTGTCG
ATTCCCTTCTTTCCTTGACAGTAACTGGCAAGTATCGAGAGGAGCTGAACCTAGACATGGATTCTTCAGTT
TTCATTGAGCATCCAACGGTCGGCGACTTCAAGCGCTTTGTGACTCAACTCAGTCCGTCAGTAGCTAGTGA
TTCGTCGTCGACTGACCGTGAGTCTGAGTATTCATTCAATGGTGACAGCTGTTCCGGGCTATCAAGTCCAG
CATCGCCTGGCACCGTATCGCCACCAAATGAGAAGGTTATCCAAATTCATGAGAACGGCACTATGAAAGAG
ATTCGCGCCATTATAGCCGACGAAATTGGTGTCTCGGCAGACGAAATCAAGAGCGATGAAAACCTAAATGA
ATTGGGAATGGACTCGCTTCTCTCACTTACAGTCCTAGGCAAAATCCGCGAGTCACTGGACATGGACCTGC
CGGGAGAGTTCTTCATCGAGAACCAGACCCTGGACCAGATTGAGACTGCGCTGGACCTGAAGCCCAAAGCT
GTCCCGACCGCTGTCCCACAATCGCAGCCCATAACCCTACCACAATCACAATCAACGAAACAGCTTTCCAC
GCGACCTACCAGCAGCAGCGACAACCATCCCCCGGCAACCTCCATTCTCCTCCAAGGAAACCCGCGCACAG
CGTCAAAGACCCTCTTTCTCTTCCCCGACGGTTCTGGCTCCGCAACCTCCTACGCCACAATACCCGGAGTC
TCCCCAAACGTTGCCGTTTATGGCCTCAACTGCCCCTACATGAAAGCCCCCGAAAAGCTCACCTGCAGCCT
CGATTCGCTGACAACGCCGTATCTCGCCGAGATTCGCCGCCGCCAACCAACCGGCCCGTACAACCTTGGCG
GCTGGTCCGCAGGCGGGATCTGCGCATACGACGCGGCACGCAAGCTCGTACTGCAGCAGGGCGAAATAGTC
GAGACCCTTCTTCTTCTCGACACCCCCTTCCCCATTGGCCTTGAGAAACTCCCGCCCCGCCTGTACAGCTT
CTTCAACTCGATCGGCCTCTTTGGCGAAGGCAAGGCCGCGCCGCCGGCTTGGCTACTTCCCATTTTTTGG
CGTTCATTGACTCCCTCGACGCGTACAAGGCCGTGCCTCTTCCCTTCAACGAACAGGAATGGAAGGGAAAG
CTGCCAAAAACGTACCTTGTTTGGGCTAAGGATGGGGTCTGCCCTAAGCCTGGTGATCCGTGGCCTGAGCC
TGCGGAGGACGGGAGCAAGGATCCACGCGAGATGGTTTGGTTACTTTCTAACCGGACGGATCTGGGTCCGA
ATGGATGGGATACGCTTGTGGGCAAGGAAAATATTGGCGGTATTACAGTGATTCATGACGCGAATCACTTT
ACTATGACAAAGGGCGAAAAGGCAAAGGAGCTTGCTACTTTCATGAAGAATGCGTTGGGGGTTTGTGAGCG
CCGGTTGGTC**GCGGCCGCACTCGAGCACCACCACCACCACCACTGA**
```

Expression Construct (pEwA) Protein Sequence:
```
MEDPYRVYLFGDQTGDFEVGLRRLLQAKNHSLLSSFLQRSYHAVRQEISHLPPSERSTFPRFTSIGDLLAR
HCESPGNPAIESVLTCIYQLGCFINYYGDLGHTFPSHSQSQLVGLCTGLLSCAAVSCASNIGELLKPAVEV
VVVALRLGLCVYRVRKLFGQDQAAPLSWSALVSGLSESEGTSLIDKFTRRNVIPPSSRPYISAVCANTLTI
SGPPVVLNQFLDTFISGKNKAVMVPIHGPFHASHLYEKRDVEWILKSCNVETIRNHKPRIPVLSSNTGELI
VVENMEGFLKIALEEILLRQMSWDKVTDSCISILKSVGDNKPKKLLPISSTATQSLFNSLKKSNLVNIEVD
GGISDFAAETQLVNQTGRAELSKIAIIGMSGRFPEADSPQDFWNLLYKGLDVHRKVPEDRWDADAHVDLTG
TATNTSKVPYGCWIREPGLFDPRFFNMSPREALQADPAQRLALLTAYEALEGAGFVPDSTPSTQRDRVGIF
YGMTSDDYREVNSGQDIDTYFIPGGNRAFTPGRINYYFKFSGPSVSVDTACSSSLAAIHLACNSIWRNDCD
TAITGGVNILTNPDNHAGLDRGHFLSRTGNCNTFDDGADGYCRADGVGTVVLKRLEDALADNDPILGVING
```

```
AYTNHSAEAVSITRPHVGAQAFIFKKLLNEANVDPKNISYIEMHGTGTQAGDAVEMQSVLDVFAPDHRRGP
GQSLHLGSAKSNIGHGESASGVTSLVKVLLMMKENMIPPHCGIKTKINHNFPTDLAQRNVHIALQPTAWNR
PSFGKRQIFLNNFSAAGGNTALLLEDGPVSDPEGEDKRRTHVITLSARSQTALQNNIDALCQYISEQEKTF
GVKDSNALPSLAYTTTARRIHHPFRVTAIGSSFQEMRDSLIASSRKEFVAVPAKTPGIGFLFTGQGAQYAA
MGKQLYEDCSHFRSAIEHLDCISQGQDLPSILPLVDGSLPLSELSPVVVQLGTTCVQMALSSFWASLGITP
SFVLGHSLGDFAAMNAAGVLSTSDTIYACGRRAQLLTERCQPGTHAMLAIKAPLVEVKQLLNEKVHDMACI
NSPSETVISGPKSSIDELSRACSEKGLKSTILTVPYAFHSAQVEPILEDLEKALQGITFNKPSVPFVSALL
GEVITEAGSNILNAEYLVRHCRETVNFLSAFEAVRNAKLGGDQTLWLEVGPHTVCSGMVKATLGPQTTTMA
SLRRDEDTWKVLSNSLSSLYLAGVDINWKQYHQDFSSSHRVLPLPTYKWDLKNYWIPYRNNFCLTKGSSMS
AASASLQPTFLTTSAQRVVESRDDGLTATVVVHNDIADPDLNRVIQGHKVNGAALCPSSLYADSAQTLAEY
LIEKYKPELKGSGLDVCNVTVPKPLIAKTGKEQFRISATANWVDKHVSVQVFSVTAEGKKLIDHAHCEVKL
FDCMAADLEWKRGSYLVKRSIELLENSAVKGDAHRLRRGMVYKLFSALVDYDENYQSIREVILDSEHHEAT
ALVKFQAPQANFHRNPYWIDSFGHLSGFIMNASDGTDSKSQVFVNHGWDSMRCLKKFSADVTYRTYVRMQP
WRDSIWAGNVYIFEGDDIIAVFGGVKFQALSRKILDIALPPAGLSKAQTSPIQSSAPQKPIETAKPTSRPA
PPVTMKSFVKKSAGPSVVVRALNILASEVGLSESDMSDDLVFADYGVDSLLSLTVTGKYREELNLDMDSSV
FIEHPTVGDFKRFVTQLSPSVASDSSSTDRESEYSFNGDSCSGLSSPASPGTVSPPNEKVIQIHENGTMKE
IRAIIADEIGVSADEIKSDENLNELGMDSLLSLTVLGKIRESLDMDLPGEFFIENQTLDQIETALDLKPKA
VPTAVPQSQPITLPQSQSTKQLSTRPTSSSDNHPPATSILLQGNPRTASKTLFLFPDGSGSATSYATIPGV
SPNVAVYGLNCPYMKAPEKLTCSLDSLTTPYLAEIRRRQPTGPYNLGGWSAGGICAYDAARKLVLQQGEIV
ETLLLLDTPFPIGLEKLPPRLYSFFNSIGLFGEGKAAPPAWLLPHFLAFIDSLDAYKAVPLPFNEQEWKGK
LPKTYLVWAKDGVCPKPGDPWPEPAEDGSKDPREMVWLLSNRTDLGPNGWDTLVGKENIGGITVIHDANHF
TMTKGEKAKELATFMKNALGVCERRLV**AAALEHHHHHH**
```

A.1.6. Pks1

Expression Construct (pEPks1alt) Gene Sequence:

```
ATGGCCGATACAATGAGCTATTTGCTGTTTGGAGACCAGTCGCTGGATACGCATGGTTTCTTAGCTGAATT
TTGTCGTAACGGGAATCCGTCCATTTTTGGCTAAGACTTTTCTGGAGCAGGCCGGTCAGGCGCTGCGTGAAG
AGATCGACGGTCTGGGTAAACTGGAACGCTCAAAACTGCCAACGTTCCAAACACTGCGGCAGCTCAACGAG
CGCTATCACGCACAGTCAATCAAACATCCGGGCATTGATAGTGCGCTGCTCTGCACCACTCAATTAGCACA
CTACATTGATCGCACCGAAAAGGAGCCCCAGGACGCGTGTCTGCACGATCATACGTTTTTTATGGGCCTGT
GCACCGGATTGTTCGCCGCGGCGGCCATCGCAAGCACCCCATCGTGTCTACCTTAATTCCCCTCGCGGTG
CAAGTAGTTCTTATGGCATTCCGCACTGGCTCCCACGTCGGCAGCTTAGCCGAACGTCTTAGCCCGCCGGT
TGGCCAGAGCGAACCTTGGACGCATATCCTTCCTGGGTTGAAAGAAAGTGATGCGAAGGAAGCTCTGGCGA
ACTTTCATGAATCGAATTACATCTCCGTCGCCAGCCAAGCATACGTTAGCGCTGTGTCGTCATCCAGTCTG
GCAATCTCTGGGCCGCCAGCAACCCTGAAGGCTCTCGACGACCAGAACGTCTTTGGCGTCAAGTCGACGGC
GATTCCCGTCTATGGTCCCTACCATGCTGCCCATCTTCACGGCACCGCCGATGTCGAGAAGATTCTCCGCC
TCAACGACCCCAAGGTCGGCGAGATTCTCGCCAAGACCAAACCCCGTTCCGCCATCATGTCCGGCACGAAG
GGTATCTGGTTCGCCGAGACCGACACAAAGTCCCTCCTCCAGGCCGTTACCCACGAGTGTCTGGTCGACGT
CCTTCAGTTCCAGAAGGGTATCGAAGGATGCATCGAGACGGCCCGCGACTTTGAGGGATCAACCTGCCTCG
TCGTCCCCTTCGGCCCCACACACAACGCCGAGACTTTGTGCAAGCTGATCCAGGATCGCACTCAGCTGAAC
GTCGTGACACGCCACGGACCCAAGATCTCGAGGGAGAGCATCAACTCCAACATTGGAAACCACGGTTCCAG
TGGCAAGTGCAAGCTCGCCATTGTCGGCATGGCTGGTCGGTTCCCCGACGCCGCCAGTCACGAGAAGCTTT
GGGAGCTTCTTGCCAAGGGTCTCGACGTCCACCGCGTCGTTCCCGCCGACCGCTTCCCCGTCGCTACCCAC
TACGACATCACCGGCAAGGCTGTCAACACCAGTCACAGCCAGTATGGATGCTGGATCGAGAACCCTGGTTA
CTTCGATCCCAGATTCTTCAACATGTCTCCTCGCGAGGCCTTCCAGACCGACCCTATGCAGCGTATGGCTC
TGACTACGGCTTACGAGGCTCTGGAGATGTGCGGTTACGTGCCCAACCGTACTCCCTCCACGAGACTCGAC
CGTATCGGTACCTTTTACGGCCAGACCTCTGACGATTGGCGTGAAATCAACGCCGCTCAAGAGGTCGACAC
CTACTACATCACGGGTGGTGTGCGTGCCTTCGGACCTGGTCGTATCAACTACCACTTTGGTTTCAGCGGTC
CCAGTCTCAACGTCGACACTGCCTGTTCCTCCAGTGCCGCCGCTCTCAACGTGGCTTGTAACTCATTGTGG
CAAAAGGACTGTGACACCGCCATTGTTGGTGGTCTGTCTTGCATGACCAACCCCGACATTTTCGCCGGTCT
CAGTCGTGGCCAGTTCCTGTCCAAGACTGGACCTTGTGCGACATTCGACAACGGCGCTGATGGTTACTGCC
GTGCTGACGGCTGTGCCTCCGTCATCGTCAAGCGCCTGGACGATGCCCTCGCCGACAAGGACAACGTTCTT
GCCGTCATCTTGGGCACGGCCACCAACCACTCCGCCGATGCCATCTCCATCACCCACCCTCACGGTCCCAC
CCAGTCCATCCTCTCGAGGGCCATTCTCGACGATGCTGGCGTCGACCCTCTTGACGTTGACTACGTTGAGA
TGCACGGCACCGGTACCCAGGCCGGTGACGGCACAGAGATGGTGTCCGTCACCAACGTCTTTGCCCCCGCC
GACCGCAAGAGACCTGCTGACAGGCCTTTGTACCTCGGTGCTGTCAAGTCCAACATCGGTCACGGAGAGGC
```

```
TGCTTCCGGTGTCACCGCCTTGACCAAGGTGCTCATGATGATGCGGAAGAACGCCATCCCTCCCCACGTCG
GTATCAAGAAGGAGATCAACAAGACCTTCCCCAAGGACTTGTCGGAGCGCAACGTCAACATTGCCTTCCAC
TTGACGCCCTTCAAGCGCCGAGACGGCAAGCCCAGACGCATCTTCGTCAACAACTTCAGTGCCGCCGGTGG
TAACACCGGACTGTTGCTCGAGGACGCGCCCCTGATCCCTGCCGCCGAGGTTGACCCCCGCAATGTACAGG
TCATCACTGTCACTGGCAAGTCCAAGGCCGCCATGATCCGCAACGCCGAGAGACTCGTCGGCTGGATGGAG
CAGAACCCCCAGACTCCTCTGTCCCACGTTGCTTACACGACCACCGCACGTCGCATCCAGCACTACTGGCG
CATGAACGTCGCGGCTTCTGACCTGCCCGAGGCTCAGCGTCTCATCAAGGACAGGCTCAAGGAGAACTTCA
GCCCCATCTCCACCCAGCAGCCCAAGGTCGCTTTCATGTTCACTGGCCAGGGTTCGCACTACGCCGGCCTC
GGCAAGGACCTTTACGCTCACTACCGCGTATTCCGTGAGAGCATCGACGAGTTCAACCAGCTCGCCCAGAT
CCACGGCTTCCCCAGCTTCCTGGCTTTGATCGATGGTAGCGAGCCCGATGTTGCCAAGCTTTCGCCCGTCA
TCGTTCAGCTCGGACTCTGCTGCTTCGAGATGGCCCTTGCCCGTCTCTGGGCTTCCTGGGGCATTCGTCCC
AGCGCCGTCATGGGACACTCTCTCGGAGAGTACGCCGCGCTCAACGCTGCCGGTGTCTTGTCTGCCAGCGA
CACCATCTACCTCGTCGGAGCTCGTGCTCAGCTTCTGGTCCAGAAGTGCACCGCCGGAACCCATGCCATGT
TGGCCGTCACTGGCCCTGTCGATGCCGTCATGGAGGCTCTCGGGTCCCAGGCTGAGGCCATCAACGTTGCT
TGCATCAACGGACCCCGCGAGACTGTACTCAGCGGCACTGCTGCGAAGGTTAGCGAGATTTCCGCGCAGCT
GGGCACTTCCGGCTTCAAGTGCACCCAGCTCAAGGTGCCTTTCGCTTTCCACTCTGCTCAGGTCGATCCCA
TCCTTGACGACTTTGAGACTCTCGCCCGCGCCGTGAGCTTCGAGAGGCCCCAGGTGCCCATCATCTCTCCT
CTCCTCGGAAAGATGGTCGAGAGCGAGCCCGTCAACGCTGCTTACCTCCGCAATCACGCTCGCGATGCCGT
CAACTTCCTTGGCGGCCTCGTCCATGCTCAGCAATCTGGCTCCATCGACGAGAAGACCGTCTGGCTCGAGG
TCGGACCCCATCCCGTCTGCGCCAACTTTGTCAAGTCTTCCTTCGGCATCAGCTCTGTTGCTGTTCCTTCC
CTGCGCCGCAATGAGCCGACCTACAAGATCCTCAGCAGCACTCTTTGCACTCTGCATACTGCGGGTGTCAA
CCTGGACTGGGATGAGTTCCACCGCGACTTCACCGACTGCACCCGCCTGTTGGACCTGCCGACCTACTCCT
TCGACGAGAAGAACTACTGGCTCCAGTACACTGGTGACTGGTGTCTCACCAAGAACCGTGGTCCTTCTGCT
GTCAAGGCTCCTCTCCAGATCGAGCCTGCCAGACCCAAACTGGCGACCACCTCCATCCACGCCATCACCAA
CGAGGACATCAAGGACGATATCGCCATCATCGATACCGAGACCAACCTGTCTCGCCCCGACACACGTCCTT
TGGTCGAGGGCCACTTGTGCAATGGCACCCCTCTTTGCCCTTCCACGCTCTACGCTGACATGGCGATGACC
GTCGCCGACTACGCCTACAAGACGTTGCGTCCCGGCACCGAGAACATTGGTCTGAACGTGGCCAATGTCGA
GGTGCCCAAGACCCTCATCTTCGATGACAAGGTTGAGGCGCACATCTTGCGCACTACCGTCACTGCCAACG
TCGGCCTCGGCTACGCCGACGTTAGCTTCCACACCGGCGAGGGTAGCAAGAAGACCGAGCACGCGACCTGC
AAGGTCGTCTACGGCGATACCGAGCAGTGGGCTGACGAGTTCGAGCGCGTCGCCTACCTCATCAAGGGTCG
TATCACCGCTCTTGAGGAGGCTGAGCGCCAGGGCAACGCCTCCAAGATTGGCCGTGGCTTGACTTACAAGC
TCTTCAGCGCCCTTGTTGACTACGACCGCAAGTACCAGGGTATGGAGGAGGTCGTCCTCGACAGCAAGGCT
TGCGAGGCTACCGCCAAGGTGTCTTTCCAGACGACCGAGAAGGATGGAAACTTCTTCTTCAACCCTTACTG
GATTGACAGCTGTTGCCACATTTCTGGCTTCATCATCAACGGTACCGACGCCATTGACTCTCGCGAGCAGG
TGTTCATCTCGCACGGCTGGGGCTCCATGAGGTTCACCGAGAAGCTTGATGCCAACAAGACCTACCGCTCC
TACGTTCGCATGCAGCCTGTCAAGGGTACCAAGACCTATGCTGGTGATGCTTACGTCTTCGACGGCGACCG
CATCGTCGGTGTTGCTGGCGACGTCAGGTTCGCCTCCATCCCGCGCAAGGTTCTTAACCTGGTTCTGCCGC
CTCGTGGAAGTGCTCTTGCTGGCGCTGCTCCTGCTGCCTCGGCTTCAAAGGCTGCCGCCCCCAAGGCTGCC
TCCCCTGCCAAGGCTGCTCCCAAAGCGAAGGCTGCCAAGGGCTCCAAGCAGGTCACCGCCAGCAACCTCAA
GGCTGTCAATGCCAAGCTTGCTACGCGCTCTGTCGTGCAGGATGTCTTTGACATCTTGGCCAAGGAAGTTG
GTGTCACCCACGACGAGCTTGCCGACAACATCGCCTTCACTGATTTGGGCTGTGACTCTCTGATGGCCCTG
ACTGTCTCCGGTCGCATGCGCGAGGAGCTCGATATTGACATCGACTCGCACGCCTTTGTCGAGCACCCCAC
CATTGGCGCCTTCAAGACTTTCCTCGCCCAGTTCGAGACTCCCGGCCGCAAGGAGAGCTACGTCCAAGACT
CTGGCGAGTCTAGCGGTTCCGTCTCTGAGACCCCCGAGCTCGAGTCCGACTCCAATGTCACTACTCCCTAC
GAGGAGAGTGACCGGTCGGTCAAGGGCGAGGGAGATGAGGCCGATTCCGACGACTTGCAAAGCATTCTCCG
CGAGACCATTGCCACCGAGATGGGCGTCGAGGTTGACGAGATCATTGCTGCTCCTGACTTGGCTGCCCTGG
GTATGGATTCTCTGATGAGCTTGTCCATTCTCGGAACTCTCCGTGAGAAGTCTGGCCAGGACATTCCCAAC
GATCTCTTTGTCACCAACCCGTCTCTGCTGGAGGTTGAGAAGGCTCTGGGCATCGGCCCGAAGCCCAAGCC
CGCGGCTGCCCCCAAGCCGGCCAAGTCTGCTCCTGCAGCTTCCCGCCGCGAGAAGGTTGAGCCCACCAAGG
AGATCAACACTCACCCGGGAAACACGACCGCCTCCATCACTAAGCCTCCTCCGCCGACCGAGATCATCGAC
AACTATCCTCATCGCAAGGCGACTTCCATCCTTCTGCAAGGCAGCACCCGCACCGCCACAAAGAACCTGTG
GATGGTCCCTGACGGTTCCGGCTGCGCTACCTCCTACACCGAGATCAGCCAGGTCTCGTCCAACTGGGCTG
TCTGGGGTCTCTTCTCCCCCTTCATGAAGACTCCCGAGGAATACAAGTGCGGTGTCTACGGCATGGCTGCG
AAGTTCATCGAAGCCATGAAGGCGAGACAGTCCAAAGGCCCCTACTCCTTGGCGGGTTGGTCTGCGGGCGG
TGTCATTGCCTACGAAATTGTCAACCAGCTCACCAAGGCCGGCGAGACTGTCGAGAACCTCATCATCATCG
ACGCTCCCTGCCCTGTCACCATTGAGCCTCTCCCCCGCAGCCTCCACGCCTGGTTCGCCTCTATCGGTCTC
CTCGGCGAAGGTGACGACGAGGCTGCCAAGAAGATTCCTTCCTGGCTCCTCCCCCACTTCGCTGCCAGCGT
CACTGCTCTGAGCAACTACACTGCCGAGCCTATTCCAAAGGAGAAGTGCCCGAACGTGATGGCCATTTGGT
```

181

```
GTGAGGATGGTGTGTGTCATCTTCCCACCGATCCCCGCCCGGATCCCTACCCGACGGGCCACGCTCTCTTC
CTGCTGGACAACAGGACCGACTTTGGTCCTAACCGTTGGGATGAGTACCTGGATGTCAACAAGTTCAGGAC
CAGGCACATGCCCGGTAACCACTTCTCCATGATGCATGGCGACTACGCCAAGCAACTTGGAACTTTCCTTC
GTGAAGCGGTTTGCGAGGCGGCCGCACTCGAGCACCACCACCACCACCACTGA
```

Expression Construct (pEPks1alt) Protein Sequence:
```
MADTMSYLLFGDQSLDTHGFLAEFCRNGNPSILAKTFLEQAGQALREEIDGLGKLERSKLPTFQTLRQLNE
RYHAQSIKHPGIDSALLCTTQLAHYIDRTEKEPQDACLHDHTFFMGLCTGLFAAAAIASTPSVSTLIPLAV
QVVLMAFRTGSHVGSLAERLSPPVGQSEPWTHILPGLKESDAKEALANFHESNYISVASQAYVSAVSSSSL
AISGPPATLKALDDQNVFGVKSTAIPVYGPYHAAHLHGTADVEKILRLNDPKVGEILAKTKPRSAIMSGTK
GIWFAETDTKSLLQAVTHECLVDVLQFQKGIEGCIETARDFEGSTCLVVPFGPTHNAETLCKLIQDRTQLN
VVTRHGPKISRESINSNIGNHGSSGKCKLAIVGMAGRFPDAASHEKLWELLAKGLDVHRVVPADRFPVATH
YDITGKAVNTSHSQYGCWIENPGYFDPRFFNMSPREAFQTDPMQRMALTTAYEALEMCGYVPNRTPSTRLD
RIGTFYGQTSDDWREINAAQEVDTYYITGGVRAFGPGRINYHFGFSGPSLNVDTACSSSAAALNVACNSLW
QKDCDTAIVGGLSCMTNPDIFAGLSRGQFLSKTGPCATFDNGADGYCRADGCASVIVKRLDDALADKDNVL
AVILGTATNHSADAISITHPHGPTQSILSRAILDDAGVDPLDVDYVEMHGTGTQAGDGTEMVSVTNVFAPA
DRKRPADRPLYLGAVKSNIGHGEAASGVTALTKVLMMMRKNAIPPHVGIKKEINKTFPKDLSERNVNIAFH
LTPFKRRDGKPRRIFVNNFSAAGGNTGLLLEDAPLIPAAEVDPRNVQVITVTGKSKAAMIRNAERLVGWME
QNPQTPLSHVAYTTTARRIQHYWRMNVAASDLPEAQRLIKDRLKENFSPISTQQPKVAFMFTGQGSHYAGL
GKDLYAHYRVFRESIDEFNQLAQIHGFPSFLALIDGSEPDVAKLSPVIVQLGLCCFEMALARLWASWGIRP
SAVMGHSLGEYAALNAAGVLSASDTIYLVGARAQLLVQKCTAGTHAMLAVTGPVDAVMEALGSQAEAINVA
CINGPRETVLSGTAAKVSEISAQLGTSGFKCTQLKVPFAFHSAQVDPILDDFETLARAVSFERPQVPIISP
LLGKMVESEPVNAAYLRNHARDAVNFLGGLVHAQQSGSIDEKTVWLEVGPHPVCANFVKSSFGISSVAVPS
LRRNEPTYKILSSTLCTLHTAGVNLDWDEFHRDFTDCTRLLDLPTYSFDEKNYWLQYTGDWCLTKNRGPSA
VKAPLQIEPARPKLATTSIHAITNEDIKDDIAIIDTETNLSRPDTRPLVEGHLCNGTPLCPSTLYADMAMT
VADYAYKTLRPGTENIGLNVANVEVPKTLIFDDKVEAHILRTTVTANVGLGYADVSFHTGEGSKKTEHATC
KVVYGDTEQWADEFERVAYLIKGRITALEEAERQGNASKIGRGLTYKLFSALVDYDRKYQGMEEVVLDSKA
CEATAKVSFQTTEKDGNFFFNPYWIDSCCHISGFIINGTDAIDSREQVFISHGWGSMRFTEKLDANKTYRS
YVRMQPVKGTKTYAGDAYVFDGDRIVGVAGDVRFASIPRKVLNLVLPPRGSALAGAAPAASASKAAAPKAA
SPAKAAPKAKAAKGSKQVTASNLKAVNAKLATRSVVQDVFDILAKEVGVTHDELADNIAFTDLGCDSLMAL
TVSGRMREELDIDIDSHAFVEHPTIGAFKTFLAQFETPGRKESYVQDSGESSGSVSETPELESDSNVTTPY
EESDRSVKGEGDEADSDDLQSILRETIATEMGVEVDEIIAAPDLAALGMDSLMSLSILGTLREKSGQDIPN
DLFVTNPSLLEVEKALGIGPKPKPAAAPKPAKSAPAASRREKVEPTKEINTHPGNTTASITKPPPPTEIID
NYPHRKATSILLQGSTRTATKNLWMVPDGSGCATSYTEISQVSSNWAVWGLFSPFMKTPEEYKCGVYGMAA
KFIEAMKARQSKGPYSLAGWSAGGVIAYEIVNQLTKAGETVENLIIIDAPCPVTIEPLPRSLHAWFASIGL
LGEGDDEAAKKIPSWLLPHFAASVTALSNYTAEPIPKEKCPNVMAIWCEDGVCHLPTDPRPDPYPTGHALF
LLDNRTDFGPNRWDEYLDVNKFRTRHMPGNHFSMMHGDYAKQLGTFLREAVCEAAALEHHHHHH
```

A.2. *NR-PKS Protein Sequence Alignments*

Protein alignments for individual NR-PKS domains were conducted in ClustalX using the BLOSUM60 scoring matrix.[1] Individual domains were aligned with sequences with known crystal or solution structures. Display files were created using the ENDscript server.[2] Alignments for SAT, KS-MAT, PT, ACP, and TE are presented below.

```
Sc_MAT                      β1            TT      α1            α2
                          ────►        TT    ℓℓℓℓ        ℓℓℓℓ......ℓℓℓℓℓℓℓℓℓ
                          1           10          20          30

Sc_MAT        ..................MLVLVAPGQGAQTPGFLTDWLALPGAADR........VAAWSDAI
Ap_PksA_SAT   ..............MAQSRQLFLFGDQTADFVPKLRSLLSVQDSPILA.AFLDQSHYVV
Gf_Pks4_SAT   ...............MSSADVYVFGDQSTPVLDKLQALVRVKDNALLT.SFLGEAFLAV
At_ACAS_SAT   MDFTPSTGTPGEFRRMKLTYFTNEFPSDDLPGLARQLHLHSKDRRHHILARFLQDATLAI
Cn_CTB1_SAT   ............MEDGAQMRVVAFGDQTYDCSEAVSQLLRVRDDAIVV.DFLERAPAVL
An_wA_SAT     ...............MEDPYRVYLFGDQTGDFEVGLRRLLQAKNHSLLS.SFLQRSYHAV
Cl_Pks1_SAT   ..............MADTMSYLLFGDQSLDTHGFLAEFCRNGNPSILAKTFLEQAGQAL


Sc_MAT            α3          α4              α5
              ℓℓℓℓℓℓ       ℓℓℓℓ     ℓℓℓℓℓℓℓℓℓ............ℓℓℓℓℓℓℓℓℓ
              40          50          60            70          80

Sc_MAT        GLDLAHFGTKADADEIRDTSVAQPLLV............AAGILSAAALGTQTSVADATG
Ap_PksA_SAT   RAQMLQSMNTVDHKLARTADLRQMVQK.YVDGKLTP.AFRTALVCLCQLGCFIREYEESG
Gf_Pks4_SAT   RREIVSLSSLERKSIPEAESLSLLLEG.VRRSEPHA.ALDSAFVCIYEIGYYIDYLARSD
At_ACAS_SAT   REEVAQLPPALKDLLPPFESVLTFVEYPEMRKGPLCGSIDGVLLSTVELATFIGYFEEFP
Cn_CTB1_SAT   KAELARLSSEQQEETPRFATLAELVPR..YRAGTLNPAVSQALTCIAQLGLFIRQHSSGQ
An_wA_SAT     RQEISHLPPSERSTFPRFTSIGDLLAR.HCESPGNP.AIESVLTCIYQLGCFINYYGDLG
Cl_Pks1_SAT   REEIDGLGKLERSKLPTFQTLRQLNERYHAQSIKHP.GIDSALLCTTQLAHYIDRTEKEP


Sc_MAT           β2           α6                α7
              ────►         ℓℓℓℓℓℓℓℓ    ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ     ───....
              90          100         110         120         130

Sc_MAT        PGFTPG....AVAGHSVGEITAAVFAGVLDDTAALSLVRRRGLAMAEAAAVTETG....
Ap_PksA_SAT   NMYPQP.SDSYVLGFCMGSLAAVAVSCSRSLSELLPIAVQTVLIAFRLGLCALEMRDRVD
Gf_Pks4_SAT   KQHPPA.APSLLLGICTGSIAAAAVSCAKDVFEISRLGVEAATVAFRLGMHVRRRAENLG
At_ACAS_SAT   ETYDFESAHTYLAGLGIGLLSAAAVSLSRTILADIAYVGSEVVRMAFRLGVLVDNVSEQLO
Cn_CTB1_SAT   EAYPTA.HDSCITGVCTGALTAVAVGSASSVTALVPLALHTVAVAVRLGARAWEIGSCLA
An_wA_SAT     HTFPSH.SQSQLVGLCTGLLSCAAVSCASNIGELLKPAVEVVVVALRLGLCVYRVRKLFG
Cl_Pks1_SAT   QDACLH.DHTFFMGLCTGLFAAAAIASTPSVSTLIPLAVQVVLMAFRTGSHVGSLAERLS


Sc_MAT           β3               α8          β4    TT   β5       α9
              ────────►         ℓℓℓℓℓℓℓℓℓℓ   ────►  TT  ────►   ℓℓℓℓℓℓℓℓℓ
              140         150         160         170         180

Sc_MAT        ............MSALLGGDPEVSVAHLERLGLTPANVNGAGQIVAAGTMEQLAALNEDK
Ap_PksA_SAT   GCSDDRGD...PWSTIVWGLDPQQARDQIEVFCRTTNVPQTRRPWISCIS.KNAITLSGS
Gf_Pks4_SAT   ..YSTPS...SWSMILSSNQEELVSEALKEFSKEKNLTYSSRPYISATG.PGFTTISGP
At_ACAS_SAT   PRELASHGTPDSWAYVLPNVTREAVQQELDVIHSGEGIPETGKIFVSAFS.QSSVTVSGP
Cn_CTB1_SAT   DARRGANGRYASWTSAVGGISPQDLQDRISAYTAEQALASVSVPYLSAAVGPGQSSVSAA
An_wA_SAT     QDQAAPL....SWSALVSGLSESEGTSLIDKFTRRNVIPPSSRPYISAVC.ANTLTISGP
Cl_Pks1_SAT   PPVGQSE....PWTHILPGLKESDAKEALANFHESNYISVASQAYVSAVS.SSSLAISGP


Sc_MAT        T...........T    β6        TT  η1    α10
              T...........T  ────►      TT  ℓℓℓ ℓℓℓℓℓℓℓℓℓℓℓℓℓ
              190         200         210         220

Sc_MAT        PE.............GVRKVVPLKVAG.AFHTRHMAPAVDKLAEAAK.......ALTPADP
Ap_PksA_SAT   PSTLRAFC.AMPQMAQHRTAPIPICL.PAHNGALFTQADITTILDT...TPTTPWEQLPG
Gf_Pks4_SAT   PSILESVK.SCDTFSGKRLYPAPIYG.PYHNSSSYSESSLEHGLAS..ILEDVGFLENEM
At_ACAS_SAT   PSRLKDLFRTSDFFRDRRFFSLPVFGGLCHAKHIYTETHVQQVVRT..KPMDMLSARVLP
Cn_CTB1_SAT   PVILDAFL..STLLRPLTTTRLPITA.PYHAPHLFTAKDVQHVTDC....LPPSEAWPTV
An_wA_SAT     PVVLNQFL.DTFISGKNKAVMVPIHG.PPHASHLYEKRDVEWILKS..CNVET.IRNHKP
Cl_Pks1_SAT   PATLKALD.DQNVFG.VKSTAIPVYG.PYHAAHLHGTADVEKILRLNDPKVGEILAKTKP


Sc_MAT          β7           α11          β8      α12
              ────►        ℓℓℓℓℓℓℓℓℓℓℓ  ────► ℓℓℓℓℓℓℓℓℓ     ─
              230         240         250         260         270

Sc_MAT        KVTYVSNKDG....RAVASGTEVLDRLVGQVANPVRWDLCMETFKELGVT.........A
Ap_PksA_SAT   QIPYISHVTG.NVVQTSNYRDLIEVALSETLLEQVRLDLVETGLPRLLQSRQVKSVTIVP
Gf_Pks4_SAT   LIPIISCASG.SRLDQLSFGNLLKNVLSSALSQQIRMDLVTDALVETVS......GTEAT
At_ACAS_SAT   RIPIFSPSSG.SPFPAATATELFEGIISEILTQVIQWDNVIQGALDQINILSPSEFQVLV
Cn_CTB1_SAT   RIPIISFSRDEAVSRGASFPAAMSEAVRDCLIRPIALDRMAVSIANHARDLGKDSVLPSP
An_wA_SAT     RIPVLSSNTG.ELIVVENMEGFLKIALEEILLRQMSWDKVTDSCISILKSVG..DNKPKK
Cl_Pks1_SAT   RSAIMSGTKG.IWFAETDTKSLLQAVTHECLVDVLQFQKGIEGCIETARDFEGSTCLVVP
```

183

```
                      β9              α13             β10        η2         α14
Sc_MAT            ────────▶      000000000      ────────▶  0000000000000000000
                            280            290        300        310
Sc_MAT         IIEVCPGGTLTGLAKRALPGVKTLALKTPDDLDAARELVAEHTQA....
Ap_PksA_SAT    FLTRMNETMSNILPDSFIS..TETRTDTGRAIPASGRPGAG........
Gf_Pks4_SAT    LIPVNAQTTVCSLADWLAK..RGATTRIGPTLESLTKDRAEPNLAPGDE
At_ACAS_SAT    FRISLPIHDLMAAVNTELKGFQATTKEIMPWVSHTAKDRIPREPS....
Cn_CTB1_SAT    IALSFSDKLGPQVNSHLPGAKAPTPELTSKSIPSAIGAEQQPMA.....
An_wA_SAT      LLPISSTATQSLFNSLKKS..NLVNIEVDGGISDFAAETQLVNQTGRAE
Cl_Pks1_SAT    FGPTHNAETLCKLIQDRTQ..LNVVTRHGPKISRESINSNIGNHGSSG.
```

Appendix Figure A.1 │ Alignment of NR-PKS SAT domains. NR-PKS sequences are aligned with the fabD MAT domain of both fatty acid and polyketide biosynthesis in *Streptomyces coelicolor*. The secondary structure is from a crystal structure of fabD MAT domain (PDB: 1MN2).[3]

Sequence alignment: *Se_DEBS_M5_KS-AT*

Block 1 — β1 / TTT / α2 / β2 / η1 (positions 1–40)

```
Se_DEBS_M5_KS-AT  ...IAIVGMACRFPGDVDSPESFWEFVSGGGDAIAEAPADRGWEPDPD..........A
Ap_PksA_KS-MAT    KCKLAIVSMSGRFP.ESPTTESFWDLLYKCLDVCKEVPRRRWDINTHVDPSGKARNKGAT
Gf_Pks4_KS-MAT    .NKIAIIGFSGRFP.EADNLDEFWDLLIRGLDVHKPVPEE.RFARDHYDPTGQRKNTSQV
At_ACAS_KS-MAT    QSKIAIVGMSCRLPGGATNTEKFWDVLEQGLDVYRTIPPDRFDVNTHYDPAGKRVNASHT
Cn_CTB1_KS-MAT    KSPIAILAASGRFP.QSSSMDQFWDVLINGVDTHELVPPTRWNAATHVSEDPKAKNVSGT
An_wA_KS-MAT      LSKIAIIGMSGRFP.EADSPQDFWNLLYKCLDVHRKVPEDRWDADAHVDLTGTATNTSKV
Cl_Pks1_KS-MAT    KCKLAIVGMAGRFP.DAASHEKLWELLAKGLDVHRVVPADRFPVATHYDITGKAVNTSHS
```

Block 2 — β3 / β4 / η2 / α3 / α4 / η3 / β5 (positions 50–100)

```
Se_DEBS_M5_KS-AT  RLCGMLAAAGDFDAGFFGISPREALAMDPQQRIMLEISWEALERAGHDPVSLRGSATG.V
Ap_PksA_KS-MAT    KWGCWLDFSGDFDPRFFGISPKEAPQMDPAQRMALMSTYEAMERAGLVPDTTPSTQRDRI
Gf_Pks4_KS-MAT    QYGCWLKSAGYFDTQFFHMSPKEAMQTDPAQRLALLTAYEALEMAGVVPDRTPSTQRNRV
At_ACAS_KS-MAT    PYGCFIEEPGLFDAPFFNMSPREAQQTDPMQRLALVTAYEALERAGYVPNRTPATNKHRI
Cn_CTB1_KS-MAT    GFGCWLHEAGEFDAAYFNMSPREAPQVDPAQRLALLTATEALEQAGVVPNRTSSTQKNRV
An_wA_KS-MAT      PYGCWIREPGLFDPRFFNMSPREALQADPAQRLALLTAYEALEGAGFVPDSTPSTQRDRV
Cl_Pks1_KS-MAT    QYGCWIENPGYFDPRFFNMSPREAFQTDPMQRMALTTAYEALEMCGYVPNRTPSTRLDRI
```

Block 3 — TT / η4 / α5 / α6 / β6 / η5 (positions 110–160)

```
Se_DEBS_M5_KS-AT  FTGVCTVDYGPRPDEAPDEVLGYVGTCTASSVASGRVAYCLGLEGPAMTVDTACSSGLTA
Ap_PksA_KS-MAT    GVFHCVTSNDWMETNTAQNIDTYFITGGNRGFIPGRINFCFEFAGPSYTNDTACSSSLAA
Gf_Pks4_KS-MAT    GVYYGTTSNDWGEVNSSQVDTYYIPGANRAFIPGRVNYFFKFTGPSIAVDTACSSSLAA
At_ACAS_KS-MAT    GTFYGQASDDYREVNTAQDVDTYFITGGCRAFGPGRINYFFKFWGPSYSIDTACSSSLAT
Cn_CTB1_KS-MAT    GVWYGATSNDWMETNSAQNVDTYFIPGGNRAFIPGRVNYFHKFSGPSYTIDTACSSSLAA
An_wA_KS-MAT      GIFYGMTSDDYREVNSGQDIDTYFIPGGNRAFTPGRINYYFKFSGPSVSVDTACSSSLAA
Cl_Pks1_KS-MAT    GTFYGQTSDDWREINAAQEVDTYYITGGVRAFGPGRINYHFGFSGPSLNVDTACSSSAAA
```

Block 4 — α7 / β7 / α8 / TT / TT TT (positions 170–220)

```
Se_DEBS_M5_KS-AT  LHLAMESLRRDECGLALAGGVTVMSSPGAFTEFRSQGGLAADG.RCKPFSKAADGFGLAE
Ap_PksA_KS-MAT    IHLACNSLWRGDCDTAVAGGTNMIYTPDGHTGLDKGFFLSRTG.NCKPYDDKADGYCRAE
Gf_Pks4_KS-MAT    INLAITSLKNRDCNTAIAGGTNVMTNPDNFAGLDRGHFLSRTG.NCKAFNDGADGYCRAD
At_ACAS_KS-MAT    VEAACTSLWNGSTDTAVVGGVNVLTNSDAFAGLSRGHFLSKIPGACKTWDCNADGYCRAD
Cn_CTB1_KS-MAT    LHMACNALWRGEVDTAIVGGTNVLTNPDMTAGLDAGHFLSRSG.NCKTFDDEADGYCRGE
An_wA_KS-MAT      IHLACNSIWRNDCDTAITGGVNILTNPDNHAGLDRGHFLSRTG.NCNTFDDGADGYCRAD
Cl_Pks1_KS-MAT    LNVACNSLWQKDCDTAIVGGLSCMTNPDIFAGLSRGQFLSKTG.PCATFDNGADGYCRAD
```

Block 5 — β8 / α9 / β9 / α10 (positions 230–280)

```
Se_DEBS_M5_KS-AT  GAGVLVLQRLSAARREGRPVLAVLRGSAVNQDGASNGLTAPSGPAQQRVIRRALENAGVR
Ap_PksA_KS-MAT    GVGTVFIKRLEDALADNDPILGVILDAKTNHSAMSESMTRPHVGAQIDNMTAALNTTGLH
Gf_Pks4_KS-MAT    GIGTLILKRLPDAIADSDPIFGVILGAHTNHSAESVSITRPLADAQEYLFKKLLNETGIH
At_ACAS_KS-MAT    GVISLVMKRLEDAQADNDNILGVILGAATNHSADAVSITHPHAGAQAHLFRDVLRNAGVD
Cn_CTB1_KS-MAT    AVVTLILKRLPDAQADKDPIQASILGIATNHSAEAASITRPHAGAQQDLFQQVLTETGLT
An_wA_KS-MAT      GVGTVVLKRLEDALADNDPILGVINGAYTNHSAEAVSITRPHVGAQAFIFKKLLNEANVD
Cl_Pks1_KS-MAT    GCASVIVKRLDDALADKDNVLAVILGTATNHSADAISITHPHGPTQSILSRAILDDAGVD
```

Block 6 — η6 / β10 / α11 / TT / β11 / α12 / η7 (positions 290–330)

```
Se_DEBS_M5_KS-AT  AGDVDYVEAHGTGTRLGDPIEVHALLSTYGAERDPD.....DPLWIGSVKSNIGHTQAAA
Ap_PksA_KS-MAT    PNDFSYIEMHGTGTQVGDAVEMESVLSVFAPSETAR..KADQPLFVGSAKANVGHGEGVS
Gf_Pks4_KS-MAT    PHDVSYVEMHGTGTQAGDAVEMRSVLNSFAFDHSR...PRDKSLYLGSVKANVGHAESAS
At_ACAS_KS-MAT    SHDVSYVELHGTGTQAGDFEEMKSVTDVFAPLTKRR..SPNQPLYVGAVKANVGHGEAVA
Cn_CTB1_KS-MAT    ANDISVCEMHGTGTQAGDSGETTSVVETLAPLNRSGSAVRTTPLYIGAVKSNVGHAESAA
An_wA_KS-MAT      PKNISYIEMHGTGTQAGDAVEMQSVLDVFAPDHRR...GPGQSLHLGSAKSNIGHGESAS
Cl_Pks1_KS-MAT    PLDVDYVEMHGTGTQAGDGTEMVSVTNVFAPADRKR..PADRPLYLGAVKSNIGHGEAAS
```

Block 7 — α13 / β12 / TT / TT / β13 / β14 (positions 340–390)

```
Se_DEBS_M5_KS-AT  GVAGVMKAVLALRHGEMPRTLHFDEP.......SPQIEWDLGAVSVVSQARSWPAGERPRR
Ap_PksA_KS-MAT    GVTSLIKVLMMMQHDTIPPHCGIKPGSKINRNFPDLGARNVHIAFEP..KPWPRTHTPRR
Gf_Pks4_KS-MAT    GVLAIIKVLLMMQKNTIPPHCGIKTKINQG.FPKDLDHRGVRIALKDSVDWSRPEGGKRR
At_ACAS_KS-MAT    GVTALLKVLLMLQKSVIPPHVGIKNSINPQ.IPKDLDKRNLHIPYEK.QSWKSTPGKSRI
Cn_CTB1_KS-MAT    GVSSLAKILLMLKHSKIPPHVGIKTKLNHR..LPDLAARNTHIARSE.VPWPRPKNGKRR
An_wA_KS-MAT      GVTSLVKVLLMMKENMIPPHCGIKTKINHN.FPTDLAQRNVHIALQP.TAWNRPSFGKRQ
Cl_Pks1_KS-MAT    GVTALTKVLMMMRKNAIPPHVGIKKEINKT.FPKDLSERNVNIAFHL.TPFKRRDGKPRR
```

Se_DEBS_M5_KS-AT

β15   TT   β16   β17   α14
400   410   420   430   440   450

```
Se_DEBS_M5_KS-AT   AGVSSFGISGTNAHVIVEEAPEADEP...EPAPDSGPVPLVLSGRDEQAMRAQAGRLADH
Ap_PksA_KS-MAT     VLINNFSAAGGNTALIVEDAPERHWP...TEKDPRSSHIVALSAHVGASMKTNLERLHQY
Gf_Pks4_KS-MAT     VLVNNFSAAGGNTSLLLEDGPAVHPARQHQDGDARTEHVVAVSARSTKALEENLKALEAY
At_ACAS_KS-MAT     AVVNNFSAAGGNTSVVLEEGPVTELT....GVDPRPSHVVAISAKSKVSLKGNLERFAAY
Cn_CTB1_KS-MAT     VLLNNFSAAGGNTCLVLEDAPEPEDS...QEVDPREHHIVALSAVTPDSMVNNLTNMITW
An_wA_KS-MAT       IFLNNFSAAGGNTALLLEDGPVSDPE....GEDKRRTHVITLSARSQTALQNNIDALCQY
Cl_Pks1_KS-MAT     IFVNNFSAAGGNTGLLLEDAPLIPAA....EVDPRNVQVITVTGKSKAAMIRNAERLVGW
```

α15   β18   α16   TT
460   470   480   490   500

```
Se_DEBS_M5_KS-AT   LAREP.......RNSLRDTGFTLATRRSAWEHRAVVVG.DRDEALAGLRAVADGRIADRT
Ap_PksA_KS-MAT     LLKN.......PHTDLAQLSYTTARRWHYLHRVSVTGASVEEVTRKLEMAIQNG....D
Gf_Pks4_KS-MAT     IAN....SWAPEGELLSQLSYTTARRVHHSRRVAFVTNGLDDLRKSLLKAATDAG...Q
At_ACAS_KS-MAT     IDTNP.......GVSLSHLSYTTMARRHHHNHRLAAAVSDAEQLKKQLTSWMQSVESHRP
Cn_CTB1_KS-MAT     IDKHS....GDSLATLPQLSYTTTARRVHHRHRAVATGTDLLQIRSSLQEQLDRRVSGER
An_wA_KS-MAT       ISEQEKTFGVKDSNALPSLAYTTTARRIHHPFRVTAIGSSFQEMRDSLIASSRKE.....
Cl_Pks1_KS-MAT     MEQNP.......QTPLSHVAYTTTARRIQHYWRMNVAASDLPEAQRLIKDRLKENFSP..
```

β19   β20   TT   α17   α18   η8   α19
510   520   530   540   550   560

```
Se_DEBS_M5_KS-AT   ATGQARTRRGVAMVFPGQGAQWQGMARDLLRESQVFADSIRDCERALAPHVDWSLTDLLS
Ap_PksA_KS-MAT     GVSRPKSKPKILFAFTGQGSQYATMGKQVYDAYPSFREDLEKFDRLAQSHGFPSFLHVCT
Gf_Pks4_KS-MAT     VKGIPAVSPKVGFLFTGQGAQETAMAIGYYKSFSSFRSDIHQLDSIATLQGLPSVLPLIH
At_ACAS_KS-MAT     ISATGP..PPVAFAFTGQGASYKSMNVELFHTLPSFREQMMHLDALAQQQGFPSFIPAID
Cn_CTB1_KS-MAT     SIPHPPNGPSFVLAFTGQGSAFAGMGVDLYKRFASFRSDIARYDQICEGMSLPSIKAMFE
An_wA_KS-MAT       FVAVPAKTPGIGFLFTGQGAQYAAMGKQLYEDCSHFRSAIEHLDCISQGQDLPSILPLVD
Cl_Pks1_KS-MAT     .ISTQQ..PKVAFMFTGQGSHYAGLGKDLYAHYRVFRESIDEFNQLAQIHGFPSFLALID
```

α20   β21   α21
570   580   590   600   610

```
Se_DEBS_M5_KS-AT   G...ARPLDRVDVVQPALFAVMVSLAALWRSHGVEPAAVVGHSQGEIAAAHVAGALTLED
Ap_PksA_KS-MAT     SPKGDVEEMAPVVVQLAITCLQMALTNLMTSFGIRPDVTVGHSLGEFAALYAAGVLSASD
Gf_Pks4_KS-MAT     G.TTPVEDLSAVVVQLGTCIIQISLARFWISLGITPQYVIGHSLGEYAALQIAGVLHVND
At_ACAS_KS-MAT     GSHPQDHAHSQVVTQLALTGTQIALAKYWMSLGVRPEVVVGHSLGEFAALHIAGVLSAGD
Cn_CTB1_KS-MAT     D.EKVFSTASPTLQQLTHVCFQMALYRLWKSLGVQAKAVVGHSLGEYAALYAAGVLSQSD
An_wA_KS-MAT       G.SLPLSELSPVVVQLGTTCVQMALSSFWASLGITPSFVLGHSLGDFAAMNAAGVLSTSD
Cl_Pks1_KS-MAT     GSEPDVAKLSPVIVQLGLCCFEMALARLWASWGIRPSAVMGHSLGEYAALNAAGVLSASD
```

α22   η9   β22   α23   η10   β23   TT   β24
620   630   640   650   660   670

```
Se_DEBS_M5_KS-AT   AAKLVAVRSRVLRRLGGQGGMASFGLGTEQAAERIGRFAG...ALSIASVNGPRSVVVAG
Ap_PksA_KS-MAT     VVYLVGQRAELLQERCQRGTHAMLAVKATPEALSQWIQ...DHDCEVACINGPEDTVLSG
Gf_Pks4_KS-MAT     AIFLCGHRAALLDKKCTAYTHGMVAVKAAADDLRQHISS..DLKVEIACVNGAEDTVLSG
At_ACAS_KS-MAT     TLFLVGRRAQLLEQHCVQGSYQMLAVRASVSQIEEIADG...RLYEVACINGPKETVLSG
Cn_CTB1_KS-MAT     TLYLVGRRAQLMEKHLSQGTHAMLAVRAKEEAIVAAIDGPPGEAYEFSCRNGEQRNVLGG
An_wA_KS-MAT       TIYACGRRAQLLTERCQPGTHAMLAIKAPLVEVKQLLN...EKVHDMACINSPSETVISG
Cl_Pks1_KS-MAT     TIYLVGARAQLLVQKCTAGTHAMLAVTGPVDAVMEALGSQ.AEAINVACINGPRETVLSG
```

α24   β25   η11   α25   β26   T
680   690   700   710   720   730

```
Se_DEBS_M5_KS-AT   ESGPLDELIAECEAEAHKARRIPVDYASHSPQVESLREELLTELA.GISPVSADVALYST
Ap_PksA_KS-MAT     TTKNVAEVQRAMTDNGIKCTLLKLPFAFHSAQVQPILDDFEALAQ.GATFAKPQLLILSP
Gf_Pks4_KS-MAT     PNADIESLCGKLTQAGYKLHKLEIPFAFHSSQVPILDDLEELAS.QVGFHEPKLPIVSP
At_ACAS_KS-MAT     TRQEIRNIAEHLMTKGYKCTALEVAFAGHSAQLDPILDTYEQIATKGAIFHPPNLPIISP
Cn_CTB1_KS-MAT     TVAQIQAAKAALEAKKIRCQYLDTPMAFHTGQVDPILPELLQVAA.ACSIQDPQIPVISP
An_wA_KS-MAT       PKSSIDELSRACSEKGLKSTILTVPYAFHSAQVEPILEDLEKALQ.GITFNKPSVPFVSA
Cl_Pks1_KS-MAT     TAAKVSEISAQLGTGSGFKCTQLKVPFAFHSAQVDPILDDFETLAR.AVSFERPQVPIISP
```

TT   η12   α26   α27   β27
740   750   760   770   780

```
Se_DEBS_M5_KS-AT   TTGQPIDTAT...MDTAYWYANLREQVRFQDATRQLAEAG....FDAFVEVSPHPVLTVG
Ap_PksA_KS-MAT     LLRTEIHEQG..VVTPSYVAQHCRHTVDMAQALRSAREKGLIDDKTLVIELGPKPLISGM
Gf_Pks4_KS-MAT     LLRTLLTGD...TLGPQYIRRHCRETVDPFLGAIKMAESQGIMDRSGMCIEIGAHPILTRM
At_ACAS_KS-MAT     LLGKVIFDDKT..VNATYMRRASRETVNFHAALETAQRISTVDDTTAWVEIGPHPVCMGF
Cn_CTB1_KS-MAT     AYGKVIRSAK..DFQPEYFTHHCRSSVNMVDALQSAVEEGLLDKNVIGLEIGPGPVVTQF
An_wA_KS-MAT       LLGEVITEAGSNILNAEYLVRHCRETVNFLSAFEAVRNAKLGGDQTLWLEVGPHTVCSGM
Cl_Pks1_KS-MAT     LLGKMVESEP...VNAAYLRNHARDAVNFLGGLVHAQQSGSIDEKTVWLEVGPHPVCANF
```

186

Se_DEBS_M5_KS-AT

α28   β28   α29   η13

```
                    790        800        810        820        830        840
Se_DEBS_M5_KS-AT  IEATLDSALPADAGACVVGTLRRDRGGLADFHTALGEAYAQGVEVDWSPAFADARP...
Ap_PksA_KS-MAT    VKMTLG......DKISTLPTLAPNKAIWPSLQKILTSVYTGGWDINWKKYHAPFASSQKV
Gf_Pks4_KS-MAT    VKSIIG......QDFRCLASLRRKEDHFKTLADSLCALHLAGFSVNWDEYHRDFASSRNV
At_ACAS_KS-MAT    IRSTLQ......STALTVPSLRRGEESWVTITRSLSSLHCAGVEVHWNEFHRPFEQALRL
Cn_CTB1_KS-MAT    VKEAVG......TTMQTFASINKDKDTWQLMTQALAKFYLAGASVEWSRYHEDFPGAQKV
An_wA_KS-MAT      VKATLG......PQTTTMASLRRDEDTWKVLSNSLSSLYLAGVDINWKQYHQDFSSSHRV
Cl_Pks1_KS-MAT    VKSSFG......ISSVAVPSLRRNEPTYKILSSTLCTLHTAGVNLDWDEFHRDFTDCTRL
```

β29

```
Se_DEBS_M5_KS-AT

                    850        860        870
Se_DEBS_M5_KS-AT  VELPVYPFQRQRYWLPIPTGGRARDEDDDWRYQ.....................
Ap_PksA_KS-MAT    VDLPSYGWDLKDYYIPYQGDWCLHRHQQDCKCAAPGHEIKTADYQVPPESTPHRPS
Gf_Pks4_KS-MAT    LQLPKYSWQLANYWMQYKYSWCLTKGDAPVENGPVGAVVQAR..............
At_ACAS_KS-MAT    LDLPTYSWNDKNYWIQYNGDWALTKGNTFYSSQQQNSAAVDELP............
Cn_CTB1_KS-MAT    LELPAYGWALKNYWLQYVNDWSLRKGDPAVVAAS....................
An_wA_KS-MAT      LPLPTYKWDLKNYWIPYRNNFCLTKGSSMSAASASLQP.................
Cl_Pks1_KS-MAT    LDLPTYSFDEKNYWLQYTGDWCLTKNRGPSAVKAPLQIEPARPK..........
```

Appendix Figure A.2│Alignment of NR-PKS KS-MAT didomains. NR-PKS sequences are aligned with the module 5 KS-AT didomain of 6-deoxyerythronolide B synthase (DEBS) from *Saccharopolyspora erythraea*. The secondary structure is from a crystal structure of DEBS module 5 KS-AT (PDB: 2HG4).[4]

```
                              β1        β2            α1    β3   β4
Ap_PksA_PT    1        10         20          30        40        50

Ap_PksA_PT   .EIKTTTTLHRVVEETTKPLG.ATLVVETDISRKDVNGLARGHLVDGIPLCTPSFYADIA
Gf_Pks4_PT   RALRLSDSVHNVIEQVHGDKR.SSITVESDMHDPSLLAIAQNHRVNGLTMAPSTLFADIA
At_ACAS_PT   .SGPRTSTVQKIVEESFDGRA.ARVVMQSDLMQSDLLEAAYGHKMNCCGVVTSSIHADVG
Cn_CTB1_PT   SNLELSSSIHKVITNTITANSDGELVVDADLSREDLHPMVQGHQVYGVPLCTPSVYADIA
An_wA_PT     ..TFLTTSAQRVVESRDDGLT.ATVVVHNDIADPDLNRVIQGHKVNGAALCPSSLYADSA
Cl_Pks1_PT   ..KLATTSIHAITNEDIKDDI.AIIDTETNLSRPDTRPLVEGHLCNCTPLCPSTLYADMA


                 α2                       β5                    β6       η1
Ap_PksA_PT      60        70        80        90       100       110

Ap_PksA_PT   MQVGQYSMQRLRAGHPGAGAIDGLVDVSDMVVDKALVPHGKGPQLLRTTLTMEWPPKAAA
Gf_Pks4_PT   FTLAKHLIQNHGLDTHTN.....LPSINNMAVEKALIVG.ETGPQL.....FRASLDMDW
At_ACAS_PT   FTLGQYVYKKLNPNTKVP.....AMNMASLEVLKGLVANKNTDKPQRF...QVTVTTTDI
Cn_CTB1_PT   LTLGEYIRQVIKPGEVAQT....SVEVAEMNIQSALVANNTGRVQLLR.........T
An_wA_PT     QTLAEYLIEKYKPELKGS.....GLDVCNVTVPKPLIA..KTGKEQ.....FRISATANW
Cl_Pks1_PT   MTVADYAYKTLRPGTENI.....GLNVANVEVPKTLIFDDKVEAHI.....LRTTVTANV


                    β7              β8          α3              α4
Ap_PksA_PT     120       130       140       150       160       170

Ap_PksA_PT   TTRSAKVKFATYFADGKLDTEHASCTVRFTS.DAQLKSLRRSVSEYKTHIRQLHDGH.AK
Gf_Pks4_PT   TTMRGSVRIFSVGANGKQTTLHAVCDVAVENPSSHRESWQSNAYLIQRGIKQLVQGA.SD
At_ACAS_PT   NSRILQLEWRNVHAHGPAEEPFASAKIYYCDASEWLLSWRPTLHLVQGRIQALERLA.EA
Cn_CTB1_PT   CAKFDPKAQVASCTFSSIVEQHANCKIRFGSLEKEKTALKSAALAAQASMAALKTQVGQD
An_wA_PT     VDKHVSVQVFSVTAEGKKLIDHAHCEVKLFDCMAADLEWKRGSYLVKRSIELLENSA.VK
Cl_Pks1_PT   GLGYADVSFHTG..EGSKKTEHATCKVVYGDTEQWADEFERVAYLIKGRITALEEAE.RQ


                β9    α5        β10  η2   β11   η3   β12  η4
Ap_PksA_PT     180       190       200       210       220       230

Ap_PksA_PT   GQFMRYNRKTGYKLMS.SMARFNPDYMLLDYIVLNEAENEAASGVDFSLGSSEGTFAAHP
Gf_Pks4_PT   GSAHMMRRGLLYKIFS.NSVQYGSAFQGIEQVWFDSEGLEGTGKVFMPSGKD..TFALNP
At_ACAS_PT   GIANRFSGRMAYNLFANSLVDYAGKYRGMQSVVMH..ELEAFADVTLTVEKAG.TWTVPP
Cn_CTB1_PT   DNTYRFSKGMIYKMIG.QLADFDEKYRGLCAITLDNDAMEASGKVSFKGIPNEGKFHSSP
An_wA_PT     GDAHRLRRGMVYKLFS.ALVDYDENYQSIREVILDSEHHEATALVKFQAPQA..NFHRNP
Cl_Pks1_PT   GNASKIGRGLTYKLFS.ALVDYDRKYQGMEEVVLDSKACEATAKVSFQTTEKDGNFFFNP


                     α6            TT          β13        TT    β14
Ap_PksA_PT     240       250       260       270       280       290

Ap_PksA_PT   AHVDAITQVAGFAMNANDNVDIEKQVYVNHGWDSFQIYQPLDNSKSYQVYTKMGQAKE.N
Gf_Pks4_PT   YCCDSLGHITGFIMNCSDSLDLDDHVYINHGWRTLRLVEPYQCDVQYQTYVKMQAVGSDD
At_ACAS_PT   YFIDSVAHLAGFIMNVSDAIDNQKNFCVTPGWNSMRFAAPLVAGGRYRSYVKMIPTVEDD
Cn_CTB1_PT   AYLDALSQLGGFVMNANEGVDLEKEVVINHGWDSMRFFAALDPAMTYYTHVKMTQGKD..
An_wA_PT     YWIDSFGHLSGFIMNASDGTDSKSQVFVNHGWDSMRCLKKFSADVTYRTYVRMQPWR..D
Cl_Pks1_PT   YWIDSCCHISGFIINGTDAIDSREQVFISHGWGSMRFTEKLDANKTYRSYVRMQPVKG.T


                   β15       T.T       β16          α7
Ap_PksA_PT     300       310       320       330       340       350

Ap_PksA_PT   DLVHGDVVVLDG.EQIVAFFRGLTLRSVPRGALRVVLQTTVKKADRQLGFKTMPSPPPP
Gf_Pks4_PT   STYSGDVHVLRD.GKIIGICGGVTFKKVARKVLEMLLPKPS....GAKAKHGVVKHVAP
At_ACAS_PT   SVYLGDVYILQN.DMIIGMVGGIKFRRYPRLLLNRFFSPSDDSTAKTAAGETPPAPTTT
Cn_CTB1_PT   KLWTGDVLIFDDKQALIGIVGGVALQGVPKRLMHYIVTAANKKASGPPTEKKTSSPPVE
An_wA_PT     SIWAGNVYIFEG.DDIIAVFGGVKFQALSRKILDIALPPAG....LSKAQTSPIQSSAP
Cl_Pks1_PT   KTYAGDAYVFDG.DRIVGVAGDVRFASIPRKVLNLVLPPRGSALAGAAPAASASKAAAP
```
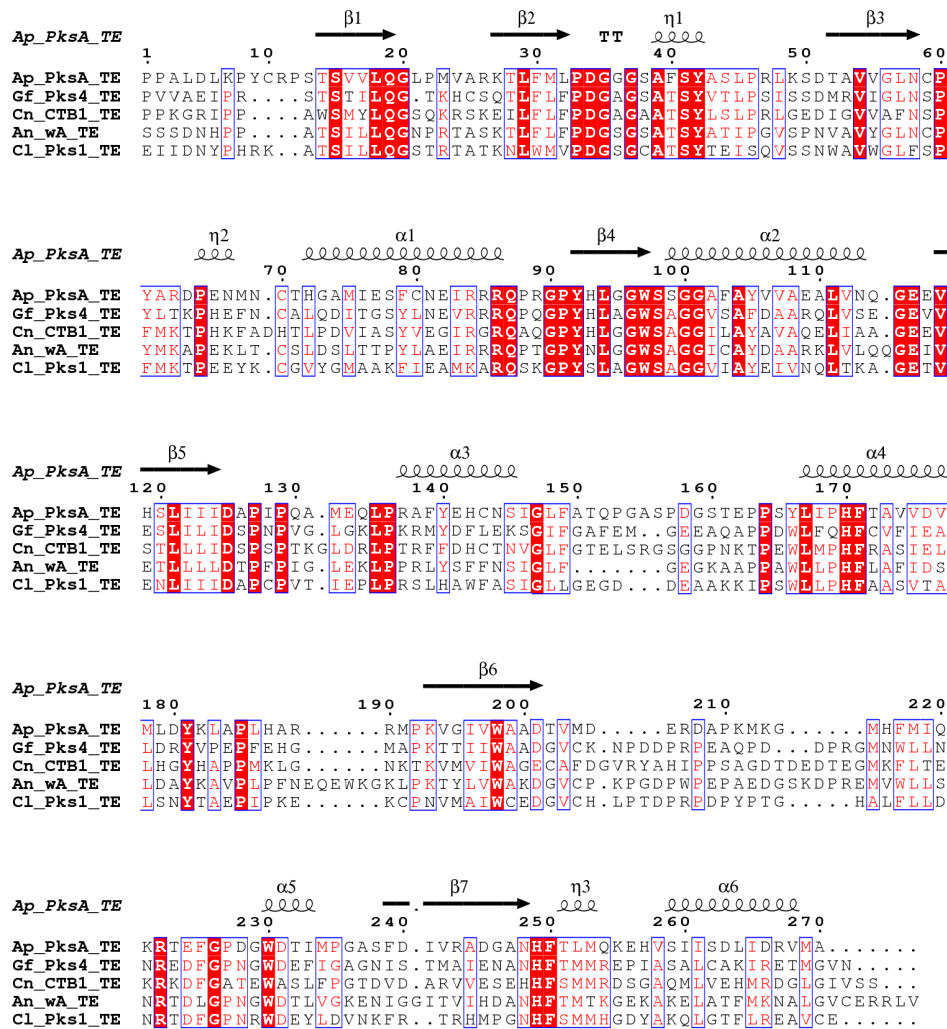
Appendix Figure A.3 | Alignment of NR-PKS PT domains. The secondary structure is from a crystal structure of PksA PT domain (PDB: 3HRQ).[5]

```
                                α1                    η1              α2                  α3
Ap_PksA_ACP              ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ        ℓℓℓ    TT    ℓℓℓℓ      ℓℓℓℓℓℓℓℓℓℓℓℓℓ
                        1         10          20          30          40          50              60
                        .         .           .           .           .           .               .
Ap_PksA_ACP    KGVGVSNEKLDAVMRVVSEESGIALEELTDDSNFADMGIDSLSSMVIGSRFREDLGLDLG
Gf_Pks4_ACP    SPVGSASGLIQKALEIIADEIGVDISQLTDTTLLADLGVDSLMSLTILGNFREELDLDI.
At_ACAS_ACP    TSPVDSNSTAARALALVAKEAGMEVTDLQDDAIFANLGVDSLMSLVIAEKFREELGVVV.
Cn_CTB1_ACP1   IKTPSVSALIAPALEIVSEEIRMPIDELKDDIDFTDAGLDSLLSLVISSRMRDQLGIEF.
Cn_CTB1_ACP2   PSPNEIGTVWRDALKILSEESGLTDEELTDDTSFADVGVDSLMSLVITSRLRDELDIDFP
An_wA_ACP1     VKKSAGPSVVVRALNILASEVGLSESDMSDDLVFADYGVDSLLSLTVTGKYREELNLDM.
An_wA_ACP2     VIQIHENGTMKEIRAIIADEIGVSADEIKSDENLNELGMDSLLSLTVLGKIRESLDMDLP
Cl_Pks1_ACP1   NAKLATRSVVQDVFDILAKEVGVTHDELADNIAFTDLGCDSLMALTVSGRMREELDIDI.
Cl_Pks1_ACP2   GDEADSDDLQSILRETIATEMGVEVDEIIAAPDLAALGMDSLMSLSILGTLREKSGQDIP


                                α4
Ap_PksA_ACP              ℓℓℓℓℓℓℓℓℓ
                        70          80
                        .           .
Ap_PksA_ACP    PEFSLFIDCTTVRALKDFMLGSGD
Gf_Pks4_ACP    .PAAQFYEFSTVQDLKSFLGANDQ
At_ACAS_ACP    .AGSLFLEYPTVGDLKSWLLEYYS
Cn_CTB1_ACP1   .ESAQFMEIGSIGGLKEFLTRLSP
Cn_CTB1_ACP2   .DRALFEECQTIFDLRKRFSGSTE
An_wA_ACP1     .DSSVFIEHPTVGDFKRFVTQLSP
An_wA_ACP2     ..GEFFIENQTLDQIETALDLKPK
Cl_Pks1_ACP1   .DSHAFVEHPTIGAFKTFLAQFET
Cl_Pks1_ACP2   ..NDLFVTNPSLLEVEKALGIGPK
```

Appendix Figure A.4 │ Alignment of NR-PKS ACP domains. Individual tandem ACPs are aligned, separately. The secondary structure is from a NMR solution structure of PksA ACP domain (PDB: 2KR5).[6]

```
                                  β1               β2          TT   η1              β3
Ap_PksA_TE   1        10          20          30            40          50          60
Ap_PksA_TE   PPALDLKPYCRPSTSVVLQGLPMVARKTLFMLPDGGGSAFSYASLPRLKSDTAVVGLNCP
Gf_Pks4_TE   PVVAEIPR....STSTILQG.TKHCSQTLFLFPDGAGSATSYVTLPSISSDMRVIGLNSP
Cn_CTB1_TE   PPKGRIPP....AWSMYLQGSQKRSKEILFLFPDGAGAATSYLSLPRLGEDIGVVAFNSP
An_wA_TE     SSSDNHPP....ATSILLQGNPRTASKTLFLFPDGSGSATSYATIPGVSPNVAVYGLNCP
Cl_Pks1_TE   EIIDNYPHRK..ATSILLQGSTRTATKNLWMVPDGSGCATSYTEISQVSSNWAVWGLFSP

                 η2              α1                β4          α2
Ap_PksA_TE            70          80          90          100         110
Ap_PksA_TE   YARDPENMN.CTHGAMIESFCNEIRRRQPRGPYHLGGWSSGGAFAYVVAEALVNQ.GEEV
Gf_Pks4_TE   YLTKPHEFN.CALQDITGSYLNEVRRRQPQGPYHLAGWSAGGVSAFDAARQLVSE.GEVV
Cn_CTB1_TE   FMKTPHKFADHTLPDVIASYVEGIRGRQAQGPYHLGGWSAGGILAYAVAQELIAA.GEEV
An_wA_TE     YMKAPEKLT.CSLDSLTTPYLAEIRRRQPTGPYNLGGWSAGGICAYDAARKLVLQQGEIV
Cl_Pks1_TE   FMKTPEEYK.CGVYGMAAKFIEAMKARQSKGPYSLAGWSAGGVIAYEIVNQLTKA.GETV

              β5                α3                          α4
Ap_PksA_TE   120         130         140         150         160         170
Ap_PksA_TE   HSLIIIDAPIPQA.MEQLPRAFYEHCNSIGLFATQPGASPDGSTEPPSYLIPHFTAVVDV
Gf_Pks4_TE   ESLILIDSPNPVG.LGKLPKRMYDFLEKSGIFGAFEM..GEEAQAPPDWLFQHFCVFIEA
Cn_CTB1_TE   STLLLIDSPSPTKGLDRLPTRFFDHCTNVGLFGTELSRGSGGPNKTPEWLMPHFRASIEL
An_wA_TE     ETLLLLDTPFPIG.LEKLPRLYSFFNSIGLF.......GEGKAAPPAWLLPHFLAFIDS
Cl_Pks1_TE   ENLIIIDAPCPVT.IEPLPRSLHAWFASIGLLGEGD...DEAAKKIPSWLLPHFAASVTA

                               β6
Ap_PksA_TE   180         190         200         210         220
Ap_PksA_TE   MLDYKLAPLHAR......RMPKVGIVWAADTVMD.....ERDAPKMKG......MHFMIQ
Gf_Pks4_TE   LDRYVPEPFEHG......MAPKTTIIWAADGVCK.NPDDPRPEAQPD...DPRGMNWLLN
Cn_CTB1_TE   LHGYHAPPMKLG......NKTKVMVIWAGECAFDGVRYAHIPPSAGDTDEDTEGMKFLTE
An_wA_TE     LDAYKAVPLPFNEQEWKGKLPKTYLVWAKDGVCP.KPGDPWPEPAEDGSKDPREMVWLLS
Cl_Pks1_TE   LSNYTAEPIPKE......KCPNVMAIWCEDGVCH.LPTDPRPDPYPTG.....HALFLLD

                    α5          β7      η3        α6
Ap_PksA_TE   230         240         250         260         270
Ap_PksA_TE   KRTEFGPDGWDTIMPGASFD.IVRADGANHFTLMQKEHVSIISDLIDRVMA.......
Gf_Pks4_TE   NREDFGPNGWDEFIGAGNIS.TMAIENANHFTMMREPIASALCAKIRETMGVN.....
Cn_CTB1_TE   KRKDFGATEWASLFPGTDVD.ARVVESEHHFSMMRDSGAQMLVEHMRDGLGIVSS...
An_wA_TE     NRTDLGPNGWDTLVGKENIGGITVIHDANHFTMTKGEKAKELATFMKNALGVCERRLV
Cl_Pks1_TE   NRTDFGPNRWDEYLDVNKFR..TRHMPGNHFSMMHGDYAKQLGTFLREAVCE......
```

Appendix Figure A.5│Alignment of NR-PKS TE domains. The secondary structure is from a crystal structure of PksA TE domain (PDB: 3ILS).[7]

A.3. References

[1]   M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, D. G. Higgins, *Bioinformatics* **2007**, *23*, 2947-2948.

[2]   X. Robert, P. Gouet, *Nucleic Acids Res* **2014**, *42*, W320-324.

[3]   A. T. Keatinge-Clay, A. A. Shelat, D. F. Savage, S. C. Tsai, L. J. Miercke, J. D. O'Connell, 3rd, C. Khosla, R. M. Stroud, *Structure* **2003**, *11*, 147-154.

[4]   Y. Tang, C.-Y. Kim, I. I. Mathews, D. E. Cane, C. Khosla, *Proc Natl Acad Sci USA* **2006**, *103*, 11124-11129.

[5]   J. M. Crawford, T. P. Korman, J. W. Labonte, A. L. Vagstad, E. A. Hill, O. Kamari-Bidkorpeh, S. C. Tsai, C. A. Townsend, *Nature* **2009**, *461*, 1139-U1243.

[6]   P. Wattana-Amorn, C. Williams, E. Ploskon, R. J. Cox, T. J. Simpson, J. Crosby, M. P. Crump, *Biochemistry* **2010**, *49*, 2186-2193.

[7]   T. P. Korman, J. M. Crawford, J. W. Labonte, A. G. Newman, J. Wong, C. A. Townsend, S. C. Tsai, *Proc Natl Acad Sci USA* **2010**, *107*, 6246-6251.

Appendix B: General Experimental Methods

B.1. *Transformation of E. coli by Electroporation*

Electroporation was used to transform *E. coli* BL21(DE3) with expression constructs. *E. coli* BL21(DE3) cells were made electrocompetent by standard protocols.[1] *E. coli* BL21(DE3) cells were grown in 1 L LB broth in a baffled flask at 37 °C with constant shaking at 250 rpm until OD_{600} 0.5 was achieved. The cells were immediately chilled on ice and harvested by centrifugation at $4000 \times g$, for 20 min at 4 °C. The pelleted cells were washed with 1 L of sterile, ice-cold ddH_2O, centrifuged as above, suspended in 0.5 L of sterile, ice-cold ddH_2O, centrifuged as above, suspended in 20 mL of sterile, ice-cold 10% glycerol, centrifuged as above, and suspended in 2–3 mL of sterile, ice-cold 10% glycerol. The resulting suspension was divided into 80 μL aliquots, flash frozen in an ethanol and dry-ice bath, and stored at −80 °C, until further use.

Electrocompetent cells were thawed on ice and 18 μL cell suspension was mixed with 2 μL of plasmid DNA for transformation. The resulting mixture was placed between the electrodes of an ice-cold Potter cuvette and pulsed with a 9–10 kV/cm electric field for 10 ms, at 4 °C. The mixture was inoculated into 180 μL LB broth and incubated at 37 °C with constant shaking at 250 rpm for 1 h. The resulting inoculum was streaked on an LB agar plate containing the appropriate antibiotic and incubated, inverted, at 37 °C, overnight.

B.2. *Heterologous Expression of Proteins in E. coli*

All proteins were expressed in *E. coli* BL21(DE3) unless otherwise stated. Most protein expression constructs in this study were prepared with a T7 promoter and the lac operon for IPTG inducible expression. Overnight starter cultures were inoculated from

−80 °C 20% glycerol cell stocks into LB broth supplemented with 50 μg/mL kanamycin and incubated at 37 °C with shaking at 250 rpm. Starter cultures were used to inoculate LB broth supplemented with 50 μg/mL kanamycin in baffled flasks by a 1:100 dilution. Cultures were incubated at 37 °C with shaking at 250 rpm until $OD_{600 \, nm}$ 0.7 was achieved (roughly 2.5 h). The cultures were immediately submerged in ice water for 30 min. Protein expression was induced with the addition of 1 mM IPTG from a 1 M stock. Induction was carried out at 19 °C with shaking at 250 rpm for 16 h. Cells were harvested by centrifugation at 4100 × g at 4 °C for 20 min. Cell pellets were flash frozen in liquid nitrogen and stored at −80 °C until further use.

B.3. *Nickel Affinity Protein Purification*

In the main, proteins in these studies were purified from frozen cell pellets and harbored 6 × His tags for nickel affinity purification. Cell pellets were suspended in lysis buffer to a final concentration of 0.25 g/mL at 4 °C. Lysis buffer was composed of 50 mM potassium phosphate pH 8.0, 300 mM NaCl, 10 mM imidazole, and 10% glycerol. The cell suspension was lysed by sonication (50% duty cycle, 10 s pulses over 3.33 min) on ice. The lysate was immediately separated by centrifugation at 25000 × g at 4 °C for 20 min. High density nickel agarose beads were added to the decanted lysate and allowed to incubate at 4 °C with gentle mixing for 1 h. The nickel beads were packed into a column under gravity and washed three times with 10 column volumes of lysis buffer and two times with 10 column volumes of wash buffer at 4 °C. Wash buffer was composed of 50 mM potassium phosphate pH 8.0, 200 mM NaCl, 50 mM imidazole, and 10% glycerol. Recombinant proteins were eluted in 4 column volumes of elution buffer at 4

°C. Elution buffer was composed of 50 mM potassium phosphate pH 8.0, 300 mM NaCl, 250 mM imidazole, and 10% glycerol.

B.4.   *Protein Concentration Determination by the Bradford Assay*

The Bradford assay was used to determine protein concentration for *in vitro* reactions of purified proteins.[2] A calibration curve was prepared using bovine serum albumin as a standard (Figure B.1). A second order polynomial regression was used to generate a standard curve for calculating protein concentration from the absorbance at 595 nm (Equation B.1). A 1 × Bio-Rad protein assay dye was prepared by diluting a 5 × stock solution in water. Purified protein was diluted to a final volume of 100 µL in water. To this dilution, 5 mL of 1 × Bio-Rad protein assay dye was added and the solution was mixed by inversion thrice. The absorbance at 595 nm was recorded and Equation B.1 was used to calculate the purified protein concentration accounting for the initial dilution.
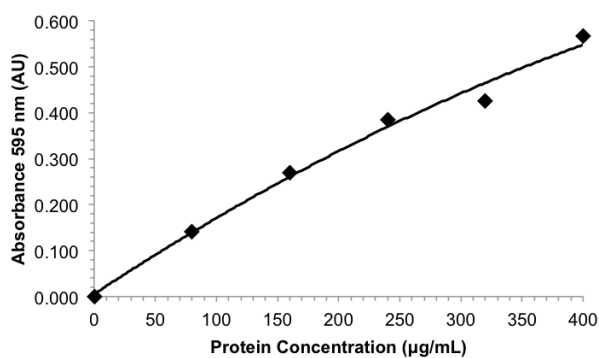
Appendix Figure B.1│Bradford assay calibration curve for determining protein concentration. The calibration curve calculated in Equation B.1 is displayed.

$$-9.93 \times 10^{-7}\, x^2 + 1.75 \times 10^{-3}\, x + 5.43 \times 10^{-3} = y \qquad (\text{Equation B. 1})$$

where *x* is the protein concentration in µg/mL and *y* is the absorbance at 595 nm in AU.

B.5. *Touchdown Polymerase Chain Reaction for DNA Amplification*

Standard PCR reactions were conducted using the touchdown PCR protocol.[3] Touchdown PCR reactions were prepared as 50 μL solutions of 1 × HF Buffer, 3% dimethyl sulfoxide (DMSO), 0.2 mM each dNTP, 0.2 μM each primer, 30 ng DNA template, and 1 U Phusion polymerase (New England Biolabs, Ipswich, MA, USA). The template DNA contained the target sequence for amplification. The primers flanked the target template sequence. Touchdown PCR reactions were cycled as follows: 98 °C for 30 s; followed by 10 cycles of 98 °C for 10 s, 68 °C decreasing by 1 °C each cycle for 30 s, and 72 °C for 15 s/kbp target sequence; followed by twenty cycles of 98 °C for 10 s, 58 °C for 30 s, and 72 °C for 15 s/kbp target sequence; followed by 72 °C for 300 s, then held at 4 °C until further use. Touchdown PCR products were run on a 1.1% agarose gel in their totality.[4] Correctly sized products were purified using a GeneJET gel extraction kit (Thermo Scientific, Waltham, MA, USA) according to the manufacturer's instructions. Purified products were stored at −20 °C until further use.

B.6. *Overlap Extension PCR and Polymerase Chain Assembly*

Overlap extension PCR was used to synthetically splice individual exons during expression construct assembly. Overlap extension PCR reactions were prepared as 50 μL solutions of 1 × HF Buffer, 3% dimethyl sulfoxide (DMSO), 0.2 mM each dNTP, 0.2 μM each outside primer, 30 ng each internal DNA fragment, and 1 U Phusion polymerase (New England Biolabs, Ipswich, MA, USA). The internal DNA fragments contained a 20–30 bp overlap sequence at the desired splicing site. The outside primers flanked the desired total DNA target for amplification. Overlap extension PCR reactions were cycled as follows: 98 °C for 30 s; followed by 30 cycles of 98 °C for 10 s, 58 °C for 30 s, and 72

°C for 15 s/kbp target sequence; followed by 72 °C for 300 s, then held at 4 °C until further use. Overlap extension PCR products were run on a 1.1% agarose gel in their totality.[4] Correctly sized products were purified using a GeneJET gel extraction kit (Thermo Scientific, Waltham, MA, USA) according to the manufacturer's instructions. Purified products were stored at −20 °C until further use. Polymerase chain assembly (PCA) was used to generate entirely synthetic exons (synthons) to be used in overlap extension PCR. PCA was conducted in the same manner as overlap extension PCR with 40 bp oligonucleotides containing nesting 20 bp overlaps and no outside primers.

B.7.   *Gibson Assembly for Plasmid Construction*

Gibson assembly was used to produce chimeric NR-PKS expression constructs according to established protocols.[5] Gibson assembly reaction were prepared as 20 μL solutions of equimolar DNA fragments (between 10 and 100 ng), 1 × ISO Buffer, 0.08 U T4 exonuclease, 0.5 U Phusion polymerase, and 80 U *Taq* DNA ligase. The DNA fragments contained 20–40 bp overlaps at their termini and were the necessary pieces generate the desired plasmid. A 5 × ISO Buffer solution was prepared according to established protocols[5] and consisted of 25% (*w/w*) poly(ethylene glycol) average molecular weight 8000 Da (PEG-8000), 500 mM Tris-HCl pH 7.5, 50 mM MgCl$_2$, 50 mM DTT, 1 mM each dNTP, and 5 mM NAD$^+$. Gibson assembly reactions were incubated at 50 °C for 1 h followed by dilution with 80 μL of water. The resulting dilution was transformed into *E. coli* BL21(DE3) cells by electroporation, plated on LB agar supplemented with 50 μg/mL kanamycin, and incubated at 37 °C for 16 h. Resulting colonies were inoculated into LB broth supplemented with 50 μg/mL kanamycin and

incubated at 37 °C with shaking at 250 rpm for 16 h. Plasmids from the resulting colonies were purified using the GeneJET Miniprep plasmid kit.

B.8. Standard Culture Recipes

Luria-Bertani (LB) broth was prepared by dissolving 10 g tryptone, 5 g yeast extract, and 10 g NaCl in 1 L of ddH$_2$O. The solution was adjusted to pH 7.0 and autoclaved for 25 min.

LB agar was prepared by dissolving 10 g tryptone, 5 g yeast extract, 10 g NaCl, and 7.5 g agar in 1 L of ddH$_2$O. The solution was adjusted to pH 7.0 and autoclaved for 25 min. The mixture was allowed to cool with constant stirring. Once the solution had reached a manageable temperature (about 45-50 °C), the appropriate antibiotic(s) were added to achieve the desired concentration and the LB agar was poured into Petri dishes. The LB agar was allowed to set overnight and the resulting plates were stored at 4 °C, inverted, until further use. Plates were dried, inverted, in as sterile hood immediately prior to use.

Potato Dextrose Agar (PDA) was prepared by dissolving 38 g of Difco™ potato dextrose agar in 1 L of ddH$_2$O with gentle heating. The solution was autoclaved for 25 min and allowed to cool with constant stirring. Once the solution had read a manageable temperature (about 45-50 °C), the PDA was poured in a sterile hood into Petri dishes to yield a bed thickness of about 3 mm. The PDA was to set for 1 h and the resulting plates were stored at 4 °C, inverted, until further use. Plates were dried, inverted, in a sterile hood immediately prior to use.

B.9. References

[1]   W. J. Dower, J. F. Miller, C. W. Ragsdale, *Nucleic Acids Res* **1988**, *16*, 6127-6145.
[2]   M. M. Bradford, *Anal Biochem* **1976**, *72*, 248-254.
[3]   D. J. Korbie, J. S. Mattick, *Nature Protocols* **2008**, *3*, 1452-1456.
[4]   D. Voytas, in *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc., **2001**.
[5]   D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison, III, H. O. Smith, *Nature Methods* **2009**, *6*, 343-U341.

Appendix C: Supplementary Material to Chapter 2

C.1. Deconstructed CTB1 Expression Construct Cloning Details

CTB1 was deconstructed into mono- to multidomain fragments for ease of

expression as detailed in Chapter 2. Specific sequence information relevant to cloning

deconstructed CTB1 expression constructs are detailed in this section.

C.1.1. Primer Sequences

Appendix Table C.1│Primers used to clone deconstructed CTB1 expression constructs. Bold sequences indicate restriction enzyme sites used for cloning. Underlined sequences indicate a position of site-directed mutagenesis.

Primer	Sequence
CTB1-ex3-5	GAACAGGCGCTTGCATCCGTGTCCGTTCCAT
CTB1-ex3-3	CTTCACGCGGACTCATGTTGAAGTAAGC
CTB1-ex6-5	TATCCCGGGTGGCAATCGCGCATTCATC
CTB1-MAT6-3	TTAT**GCGGCCGC**TGACGCGGCAACAAC
CTB1-PT-5	GGAATTC**CATATG**TCAAATTTGGAACTCTCTTCGTC
CTB1-ex6-3	ATTAGCATGCTGCTCGACGATACTAGAGAACGTGCATGA
CTB1-ex7-5	CACGTTCTCTAGTATCGTCGAGCAGCATGCTAAT
CTB1-ex7-3	CTTTCTCAAGATCCACACCCTCGTTCGCGTTCAT
CTB1-ex8-5	ATGAACGCGAACGAGGGTGTGGATCTTGAGAAAGAA
CTB1-PT-3	GTAA**GCGGCCGC**AATCGAAGCATTCTTACGCTG
CTB1-ACP-5	GTAA**CATATG**AGCGCGCCAGTCGCGCCCACG
CTB1-ACP3s	GTAA**GCGGCCGC**CTACTTCTGCGCTGGGGGAGAATC
CTB1-ex8-3	CTTGTGCGGTGTCTTCATGAAAGGCGAATTG
CTB1-ex9-5	CCTTTCATGAAGACACCGCACAAGTTTGCTG
CTB1-3	GTAA**GCGGCCGC**TGATGAAACAATCCCCAAACC
CTB1-TE-5	GTAA**CATATG**AGGATACCATCCCCGCCCAAA
CTB1-3-Stop	GTAA**GCGGCCGC**CTATGATGAAACAATCCCCAAACC
CTB1-S2008A-5	GGCGGCTGGG<u>C</u>TGCTGGTGGT
CTB1-S2008A-3	ACCACCAGCA<u>G</u>CCCAGCCGCC
CTB1-H2171Q-5	AGCGAGCATCA<u>G</u>TTCAGCATG
CTB1-H2171Q-3	CATGCTGAACT<u>G</u>ATGCTCGCT

C.1.2. Synthon of CTB1 Exons 4 and 5

Exons 4 and 5 of CTB1 comprise 67 codons of the final protein. Due to their

small sizes, a codon-optimized synthon of exons 4 and 5 was prepared by polymerase

chain assembly (PCA), as described in Chapter 2. The sequences of the synthon and the

primers used to prepare it are as follows:

Synthon DNA Sequence:
CCGCGTGAAGCGCCGCAGGTTGATCCAGCACAGCGTCTGGCCCTGCTGACCGCAACTGAAGCTCTGGAACA
GGCAGGTGTTGTTCCGAACCGTACCTCTAGCACCCAGAAGAACCGTGTAGGCGTGTGGTACGGCGCTACTA
GCAACGACTGGATGGAGACTAACTCTGCGCAGAACGTGGATACCTACTTTATCCCGGGT

Synthon Protein Sequence:
PREAPQVDPAQRLALLTATEALEQAGVVPNRTSSTQKNRVGVWYGATSNDWMETNSAQNVDTYFIPG

Appendix Table C.2│Primers used to generate synthon of CTB1 exons 4 and 5.

Primer	Sequence
CTB1-S21	TTCAACATGAGTCCGCGTGAAGCGCCGCAGGTTGATCCAG
CTB1-S22	GCGTCTGGCCCTGCTGACCGCAACTGAAGCTCTGGAACAG
CTB1-S23	GTGTTGTTCCGAACCGTACCTCTAGCACCCAGAAGAACCG
CTB1-S24	GGCGTGTGGTACGGCGCTACTAGCAACGACTGGATGGAGA
CTB1-S25	CTCTGCGCAGAACGTGGATACCTACTTTATCCCGGGTGGC
CTB1-S2-1	TGCGCGATTGCCACCCGGGATAAAGTA
CTB1-S2-2	TCCACGTTCTGCGCAGAGTTAGTCTCCATCCAGTCGTTGC
CTB1-S2-3	AGCGCCGTACCACACGCCTACACGGTTCTTCTGGGTGCTA
CTB1-S2-4	TACGGTTCGGAACAACACCTGCCTGTTCCAGAGCTTCAGT
CTB1-S2-5	GTCAGCAGGGCCAGACGCTGTGCTGGATCAACCTGCGGCG

C.2. *Deconstructed CTB1 Proteins*

Deconstructed CTB1 mono- to multidomain fragments were expressed and

purified as described in Chapter 2. Specific characteristics of these proteins are presented

in this section.

Appendix Table C.3│Details of deconstructed CTB1 expression constructs.

Protein	Encodes	Plasmid	Vector	Tag	MW (kDa)
CTB1 SAT-KS-MAT	M1-S1293	pECTB1-NKA6	pET-24a(+)	C-His	140.1
CTB1 PT	S1293-I1654	pECTB1-PT	pET-24a(+)	C-His	41.2
CTB1 ACP$_2$	S1637-K1909	pECTB1-ACP	pET-28a(+)	N-His	31.7
CTB1 TE	R1910-S2196	pECTB1-TE	pET-28a(+)	N-His	33.6
CTB1 TE-S2008A	R1910-S2196	pECTB1-TE-S2008A	pET-28a(+)	N-His	33.6
CTB1 TE-H2171Q	R1910-S2196	pECTB1-TE-H2171Q	pET-28a(+)	N-His	33.6
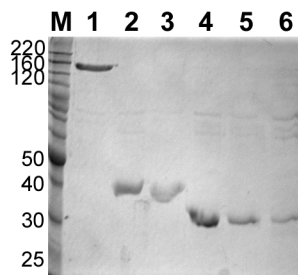
Appendix Figure C.1 | 12% SDS-PAGE of deconstructed CTB1 protein fragments. Benchmark Protein ladder (Invitrogen, **M**), CTB1 SAT-KS-MAT (**1**), CTB1 PT (**2**), CTB1 ACP$_2$ (**3**), CTB1 TE (**4**), CTB1 TE-S2008A (**5**), and CTB1 TE-H2171Q (**6**).

C.3. *Nor-toralactone Synthesis*

A synthetic standard of *nor*-toralactone was prepared by Dr. Jonathan R. Scheerer

(Johns Hopkins University, Baltimore, MD, USA) as follows:
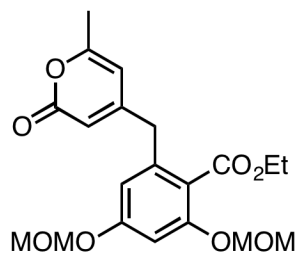
Ethyl 2,4-bis(methoxymethoxy)-6-((6-methyl-2-oxo-2*H*-pyran-4-yl)methyl)benzoate.

A colorless solution of ethyl 2,4-bis(methoxymethoxy)-6-methylbenzoate[1] (100

mg, 0.35 mmol) in THF (0.7 mL) was added dropwise to a freshly prepared −78 °C

solution of LDA in THF (0.5 M, 1.54 mL, 0.77 mmol). After allowing the resulting deep

red solution to equilibrate at −78 °C for 20 min, a solution of 4-methoxy-6-methyl-2*H*-

pyran-2-one[1] (50 mg, 0.35 mmol) in THF (0.7 mL) was added slowly. The resulting

orange solution was allowed to react at −78 °C for 1 h, quenched (sat. NH$_4$Cl, 10 mL),

diluted (Et$_2$O, 30 mL), washed (3 × sat. NH$_4$Cl), concentrated and purified by flash

chromatography (4:1 hexane:EtOAc → 3:1 hexane:EtOAc); yielding a colorless oil (99

mg, 72%). ^{1}H NMR (400 MHz, CDCl$_3$, δ): 6.77 (d, *J* = 2.0 Hz, 1H), 6.48 (d, *J* = 2.0 Hz,

1H), 5.89 (s, 1H), 5.87 (s, 1H), 5.17 (s, 2H), 5.15 (s, 2H), 4.30 (q, *J* = 7.2 Hz, 2H), 3.66

(s, 2H), 3.47 (s, 3H), 3.44 (s, 3H), 2.17 (s, 3H), 1.29 (t, *J* = 7.2 Hz, 3H).
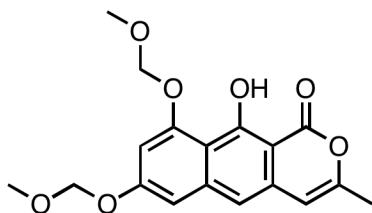
10-hydroxy-7,9-bis(methoxymethoxy)-3-methyl-1*H*-benzo[*g*]isochromen-1-one.

To a 0 °C colorless solution of the ester above (117 mg, 0.30 mmol) in THF (5

mL), LiHMDS (1 M solution in hexane, 0.66 mL, 0.66 mmol) was added. The resulting

dark brown solution was allowed to warm to room temperature over 3 h, quenched (sat.

NH₄Cl), diluted (Et₂O), washed (3 × sat. NH₄Cl), concentrated and purified by flash

chromatography on silica gel (4:1 hexane:EtOAc, R$_f$ 0.4) to afford the tricyclic pyrone as

a colorless solid (69 mg, 67 %). ¹H NMR (400 MHz, CDCl₃, δ): 12.99 (s, 1H), 6.94 (s,

1H), 6.92 (d, *J* = 2.4 Hz, 1H), 6.77 (d, *J* = 2.4 Hz, 1H), 6.20 (s, 1H), 5.36 (s, 2H), 5.28 (s,

2H), 3.61 (s, 3H), 3.52 (s, 3H), 2.22 (s, 3H).
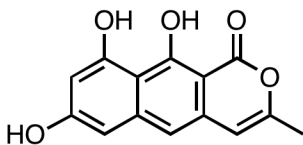
7,9,10-trihydroxy-3-methyl-1*H*-benzo[*g*]isochromen-1-one.

The di-MOM protected pyrone above (69 mg, 0.20 mmol) was dissolved in a

solution of MeOH (10 mL) and conc. HCl (0.4 mL), and allowed to react at room

temperature, under Ar for 4 h. The reaction mixture was diluted (EtOAc), washed (3 ×

sat. NH₄Cl), dried (MgSO₄) and concentrated to a light yellow solid (45 mg, 87%). ¹H

NMR (400 MHz, d₆-acetone, δ) 7.07 (s, 1H), 6.73 (d, *J* = 2.2 Hz, 1H), 6.49 (d, *J* = 2.2

Hz, 1H), 2.26 (s, 3H); HRMS–ESI-IT-TOF (*m/z*): [MH$^+$] calcd for C$_{14}$H$_{11}$O$_5$ 259.061; found 259.060.

C.4. *References*

[1]   J. M. Crawford, P. M. Thomas, J. R. Scheerer, A. L. Vagstad, N. L. Kelleher, C. A. Townsend, *Science* **2008**, *320*, 243-246.

Appendix D: Supplementary Material to Chapter 3

D.1. Deconstructed NR-PKS Expression Construct Cloning Details

The NR-PKSs used in this study were deconstructed into SAT-KS-MAT tridomain fragments and PT, ACP_n, and TE monodomain fragments for ease of selective domain-swapping as detailed in Chapter 3. Specific sequence information relevant to cloning deconstructed NR-PKSs and the CTB1/Pks1 chimeric NR-PKS expression constructs are detailed in this section.

D.1.1. Primer Sequences

Appendix Table D.1│Primers used to clone new constructs in this study.

Primer	Sequence
Pks4-SAT-s1	GTAACATATGGCGAGCAGCGCCGACGTGTATGTGTTC
Pks4-NKA-3	GTAAGCGGCCGCCGCTTGTACAACTGCACCGACC
CTB1-S3+1	AAGCGCTAGCTAAGTTCTATCTTGCAGGTGCCAGTGTCGA
CTB1-ex6-3	ATTAGCATGCTGCTCGACGATACTAGAGAACGTGCATGA
CTB1-ex7-5	CACGTTCTCTAGTATCGTCGAGCAGCATGCTAAT
CTB1-3	GTAAGCGGCCGCTGATGAAACAATCCCCAAACC

D.1.2. Primers for Gibson Assembly of Chimeric NR-PKS

Appendix Table D.2│Primers used for Gibson assembly of the CTB1/Pks1 chimeric NR-PKS.

Primer	Sequence
M4P6-5	CCGTGGTTGTTGCCGCGTCACCCAAACTGGCGACCACCTC
M4P6-3	GAGGTGGTCGCCAGTTTGGGTGACGCGGCAACAACCACGG
T7	TAATACGACTCACTATAGGG
T7 Term	GCTAGTTATTGCTCAGCGG

D.2. Deconstructed NR-PKSs Used in This Study

Deconstructed NR-PKS fragments were expressed and purified as described in Chapter 3. Specific characteristics of these proteins are presented in this section.

Appendix Table D.3│Protein constructs used in this study. Experimental cloning details are provided in the given references. Asterisks indicate revised sequence numbering described in the given reference.

Protein	Encodes	Plasmid	Vector	Tag	MW (kDa)	Ref.
PksA SAT-KS-MAT	M1-S1294	pENKA4	pET24a	C-His	143.7	[1]
PksA PT	E1305-I1677	pEPT2	pET24a	C-His	42.4	[1]
PksA ACP	I1677-S1839	pEACP41	pET28a	N-His	20.6	[2]
PksA TE	D1812-A2109	pETE2	pET28a	N-His	36.4	[1]
ACAS PT	S1318-A1669	pEACAS-PT2	pET24a	C-His	40.0	[3]
ACAS ACP	A1668-S1771	pEACAS-ACP2	pET28a	N-His	13.3	[3]
ACTE	Full protein	pEACTE	pET28a	N-His	37.5	[3]
Pks4 SAT-KS-MAT	M1-A1283*[4]	pEPks4-NKAn	pET24a	C-His	140.8	This study
Pks4 PT	R1284-P1631*[4]	pEPks4-PT	pET24a	C-His	39.6	[3]
Pks4 ACP	P1631-K1764*[4]	pEPks4-ACP	pET28a	N-His	16.2	[3]
Pks4 TE	P1765-N2036*[4]	pEPks4-TE	pET28a	N-His	32.2	[3]
CTB1 SAT-KS-MAT	M1-S1293	pECTB1-NKA6	pET24a	C-His	140.1	[5]
CTB1 PT	S1293-I1654	pECTB1-PT	pET24a	C-His	41.2	[5]
CTB1 ACP$_2$	S1637-K1909	pECTB1-ACP	pET28a	N-His	31.7	[5]
CTB1 TE	R1910-S2196	pECTB1-TE	pET28a	N-His	33.6	[5]
CTB1 full-length	Full protein	pECTB1	pET24a	C-His	238.0	This study
wA PT	T1287-P1619	pEwA-PT	pET24a	C-His	38.5	[3]
wA ACP$_2$	P1619-H1879	pEwA-ACP	pET28a	N-His	30.4	[3]
wA TE	L1859-V2157	pEwA-TE2	pET28a	N-His	34.9	[3]
Pks1 SAT-KS-MAT	M1-K1291*[6]	pEPks1-NKA	pET24a	C-His	141.3	[4]
Pks1 PT	K1291-P1640*[6]	pEPks1-PT	pET24a	C-His	39.4	[3]
Pks1 ACP$_2$	P1640-K1908*[6]	pEPks1-ACP	pET28a	N-His	30.9	[3]
Pks1 TE	P1881-E2183*[6]	pEPks1-TEalt	pET28a	N-His	35.7	[6]
Pks1 full-length	Full protein	pEPks1alt	pET24a	C-His	238.1	[6]
CTB1 SAT-KS-MAT/Pks1 PT-ACP$_2$-TE Chimera	CTB1: M1-S1293 Pks1: P1290-E2183*	pEM4P6	pET24a	C-His	237.1	This study
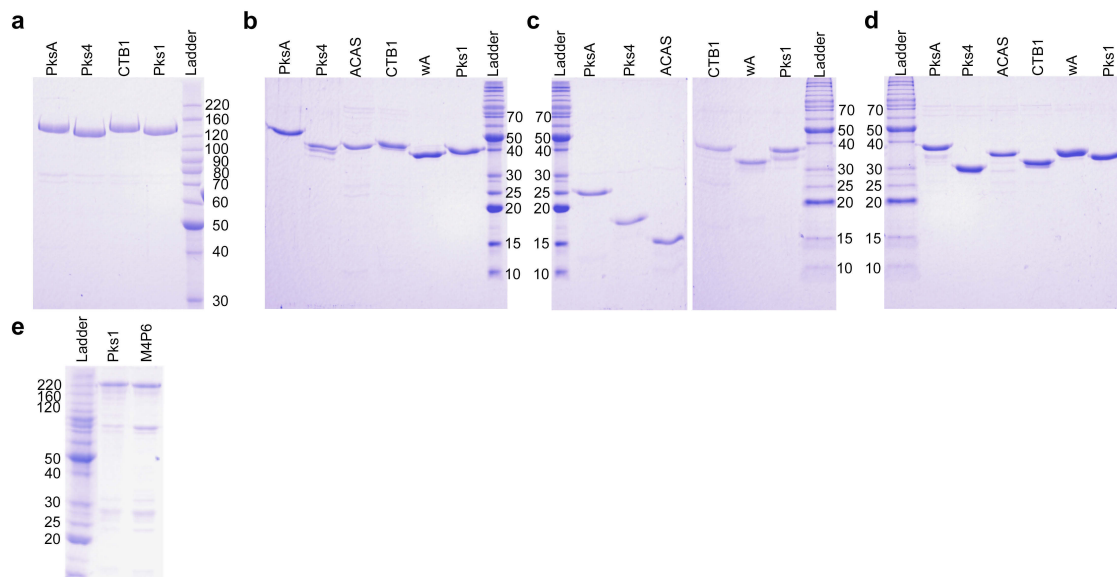
Appendix Figure D.1│Purified proteins used in this study separated by SDS-PAGE. BenchMark Protein Ladder was used as a molecular weight standard (indicated in kDa). A) 10% SDS-PAGE of SAT-KS-MAT proteins, B) 15% SDS-PAGE of PT proteins, C) 15% SDS-PAGE of ACP$_n$ proteins, D) 15% SDS-PAGE of TE proteins, and E) 12% SDS-PAGE of full-length PKS.

D.3. *Combinatorial Reactions of CTB1 SAT-KS-MAT*

Previously, we conducted and reported combinatorial reactions including CTB1 SAT-KS-MAT.[3] The results of these reactions are presented in Figure D.2.
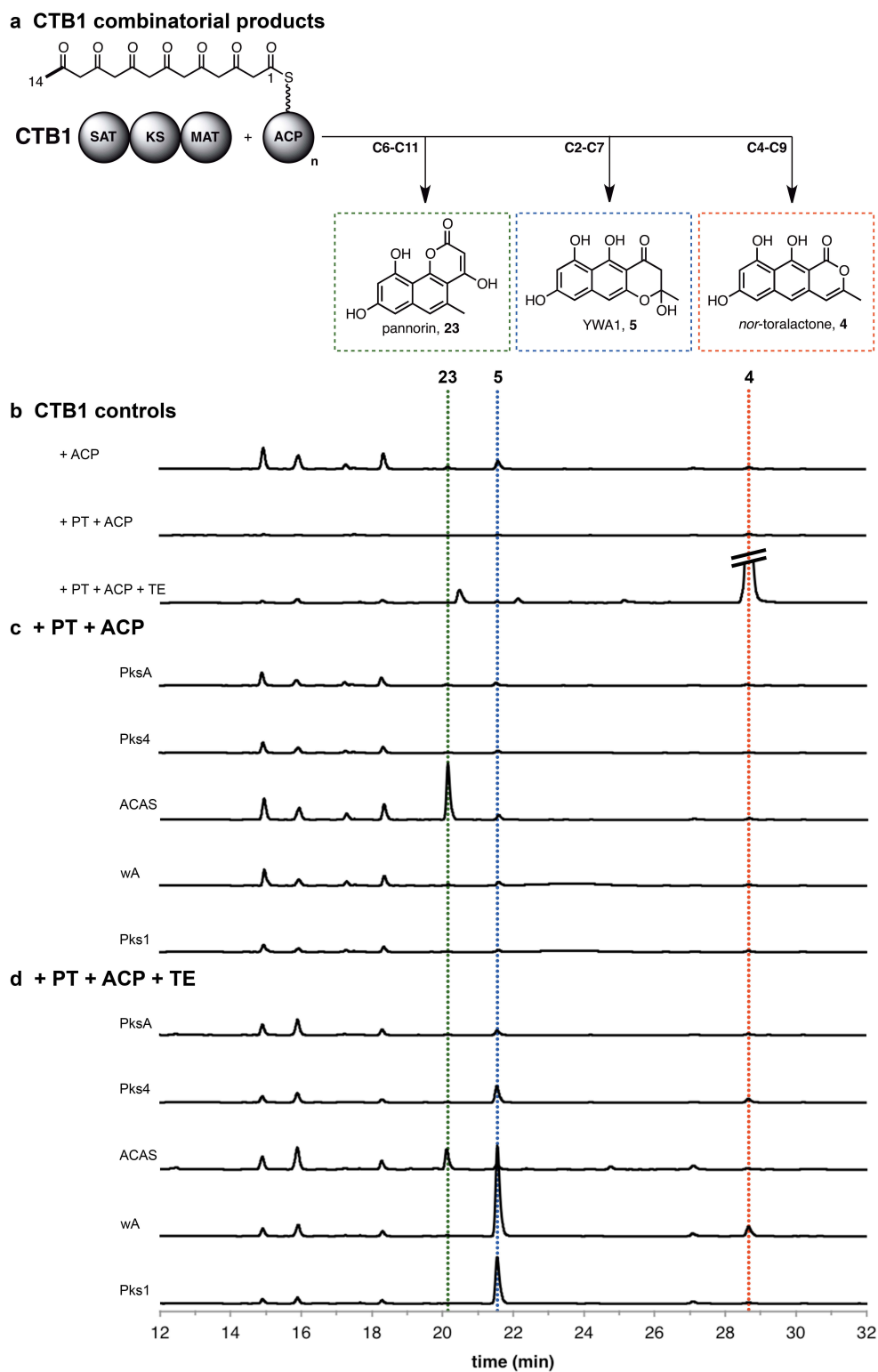
Appendix Figure D.2│Product analysis of combinatorial reactions with CTB1 SAT-KS-MAT.
Proposed products of combinatorial reactions are presented (**a**). The chromatograms at 280 nm for products of cognate control reactions (**b**), non-cognate combinatorial reactions with the indicated PT and ACP$_n$ (**c**), and non-cognate combinatorial reactions with the indicated PT, ACP$_n$, and TE (**d**) are presented.

D.4. Polyketide Product Chemical Characterization

All mass data unless otherwise stated was collected on a Shimadzu LC-IT-TOF (Shimadzu Corporation, Kyoto, Japan) in positive ion mode fitted with a Luna C18(2) column (2.0 × 150 mm, 3 μm [Phenomenex, Torrence, CA]) using a linear gradient of 5% to 85% solvent B over 30 min at 0.2 mL/min. HPLC retention times are reported for a Prodigy 5u ODS3 column (4.6 × 250 mm, 5 μm [Phenomenex]) using a linear gradient of 5% to 85% solvent B over 30 min at 1 mL/min. Solvent A was water + 0.1% formic acid and solvent B was acetonitrile + 0.1% formic acid in both cases. The λ_{max} for each compound was recorded for the HPLC peak maximum at 280 nm.
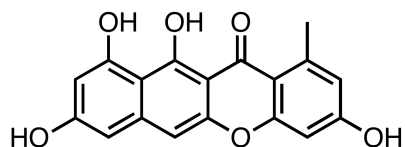
3,8,10,11-tetrahydroxy-1-methyl-12*H*-benzo[*b*]xanthen-12-one (pre-bikaverin, 2).

UV (acetonitrile [*aq*.]) λ_{max}, nm: 226, 254, 276, 302, 332, 348, 428. HPLC retention time: 33.0 min. HRMS–ESI-IT-TOF (*m/z*): [MH$^+$] calcd for $C_{18}H_{13}O_6$, 325.0712; found, 325.0703. These data are in complete agreement with literature values for **2**.[7]
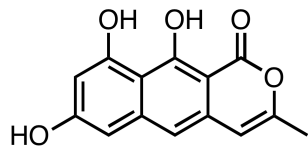

7,9,10-trihydroxy-3-methyl-1*H*-benzo[*g*]isochromen-1-one (*nor*-toralactone, 4).

UV (acetonitrile [*aq*.]) λ_{max}, nm: 268, 278, 394. HPLC retention time: 28.7 min. HRMS–ESI-IT-TOF (*m/z*): [MH$^+$] calcd for $C_{14}H_{11}O_5$, 259.0607; found, 259.0606. These data are in complete agreement with literature values for **4**.[5]
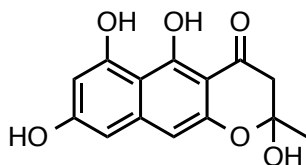
2,5,6,8-tetrahydroxy-2-methyl-2,3-dihydro-4*H*-benzo[*g*]chromen-4-one (YWA1, 5).

UV (acetonitrile [*aq*.]) λ_{max}, nm: 228, 280, 324, 334, 408. HPLC retention time: 21.6 min.

HRMS–ESI-IT-TOF (*m/z*): [MH$^+$] calcd for $C_{14}H_{13}O_6$, 277.0712; found, 277.0670. These

data are in complete agreement with literature values for **5**.[8]
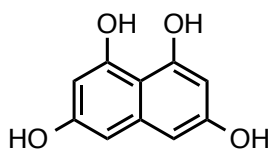
naphthalene-1,3,6,8-tetraol (THN, 6).

UV (acetonitrile [*aq*.]) λ_{max}, nm: 266, 300. HPLC retention time: 14.2 min. HRMS–

ESI/APCI-TOF (*m/z*): [MH$^+$] calcd for $C_{10}H_9O_4$, 193.0501; found, 193.0494. MS data

was collected on a Waters Acquity/Xeno-G2 UPLC-MS in positive ion mode. These data

are in complete agreement with literature values for **6**.[6]
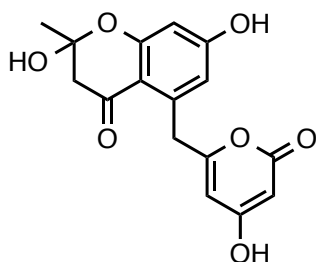
6-((3,6-dihydroxy-6-methyl-8-oxo-5,6,7,8-tetrahydronaphthalen-1-yl)methyl)-4-

hydroxy-2*H*-pyran-2-one (SEK4, 7).

UV (acetonitrile [*aq*.]) λ_{max}, nm: 280. HPLC retention time: 15.9 min. HRMS–ESI-IT-

TOF (*m/z*): [MH$^+$] calcd for $C_{16}H_{15}O_7$ 319.0818; found, 319.0806. SEK4 coelutes with its

isomer SEK4b, **8**. These data are in complete agreement with literature values for **7**.[9]
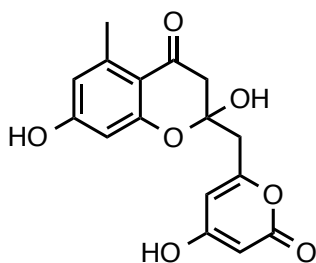
2,5,7-trihydroxy-2-((4-hydroxy-2-oxo-2H-pyran-6-yl)methyl)chroman-4-one (SEK4b, 8).

UV (acetonitrile [*aq.*]) λ_{max}, nm: 280. HPLC retention time: 15.9 min. HRMS–ESI-IT-TOF (*m/z*): [MH$^+$] calcd for $C_{16}H_{15}O_7$, 319.0818; found, 319.0806. SEK4b coelutes with its isomer SEK, **7**. These data are in complete agreement with literature values for **8**.[9c, 10]
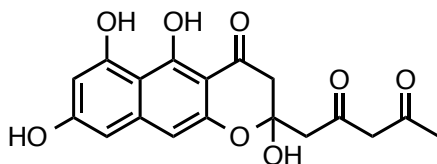
1-(2,5,6,8-tetrahydroxy-4-oxo-3,4-dihydro-2H-benzo[g]chromen-2-yl)pentane-2,4-dione (11).

UV (acetonitrile [*aq.*]) λ_{max}, nm: 230, 284, 326, 338, 416. HPLC retention time: 17.6 min. HRMS–ESI/APCI-TOF (*m/z*): [MH$^+$] calcd for $C_{18}H_{17}O_8$, 361.0923; found, 361.0913; [MH$^+$-H$_2$O] calcd for $C_{18}H_{15}O_7$, 343.0818; found, 343.0808. MS data was collected on a Waters Acquity/Xeno-G2 UPLC-MS in positive ion mode. The λ_{max} and shape of the UV-vis spectrum for this molecule are similar to those for YWA1 (**5**) with which it likely shares the same core architecture (see Figures 3.5, D.3).[8] This peak converts to pre-bikaverin (**2**) with time. Taken together, these data support the proposed structural assignment. This structure is consistent with polyketide cyclization logic.
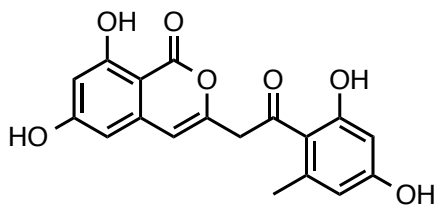
3-(2-(2,4-dihydroxy-6-methylphenyl)-2-oxoethyl)-6,8-dihydroxy-1*H*-isochromen-1-one (SMA93, 12).

UV (acetonitrile [*aq*.]) λ_{max}, nm: 244, 278, 328. HPLC retention time: 20.0 min. HRMS–ESI-IT-TOF (*m/z*): [MH$^+$] calcd for $C_{12}H_{11}O_5$, 235.0607; found, 235.0603. These data are in complete agreement with literature values for **12**.[11]

R = CH$_2$COCH$_2$COCH$_3$, **13**
R = CH$_2$COCH$_3$, **14**
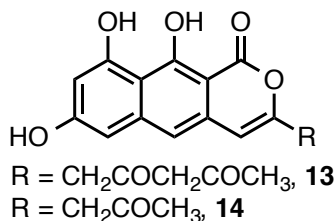
1-(7,9,10-trihydroxy-1-oxo-1*H*-benzo[*g*]isochromen-3-yl)pentane-2,4-dione (13) or

7,9,10-trihydroxy-3-(2-oxopropyl)-1*H*-benzo[*g*]isochromen-1-one (14).

UV (acetonitrile [*aq*.]) λ_{max}, nm: 270, 280, 392. HPLC retention time: 24.8 min. HRMS–ESI-IT-TOF (*m/z*): [MH$^+$] calcd for $C_{18}H_{15}O_7$ (**13**), 345.0818; mass not found; [MH$^+$] calcd for $C_{16}H_{13}O_6$ (**14**), 301.0712; mass not found. The λ_{max} and shape of the UV-vis spectrum for this molecule are similar to those for *nor*-toralactone (**4**) and norpyrone (**19**), suggesting a shared core architecture (see Figures 3.6, D.3).[1, 5] Because a mass for this peak could not be found, we propose structures **13** and **14** as the most reasonable assignments for this product. Both are consistent with polyketide cyclization logic.
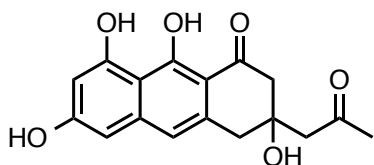
3,6,8,9-tetrahydroxy-3-(2-oxopropyl)-3,4-dihydroanthracen-1(2*H*)-one (15).

UV (acetonitrile [*aq*.]) λ_{max}, nm: 226, 272, 398. HPLC retention time: 20.9 min. HRMS–

ESI-IT-TOF (*m/z*): [MH⁺] calcd for $C_{17}H_{17}O_6$, 317.1025; found, 317.1015. The λ_{max} and

shape of the UV-vis spectrum for this molecule are similar to those for atrochrysone,

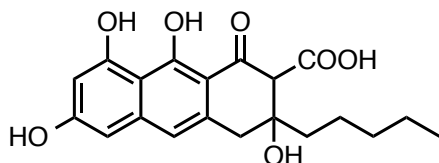suggesting a shared core architecture (see Figures 3.7, D.3).[12] Taken together, these data

support the proposed structural assignment. This assignment is in agreement with

polyketide cyclization logic.
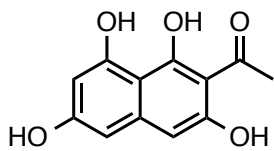
1-(1,3,6,8-tetrahydroxynaphthalen-2-yl)ethan-1-one (ATHN, 16).

UV (acetonitrile [*aq*.]) λ_{max}, nm: 230, 280, 320, 330, 404. HPLC retention time: 24.9 min.

HRMS–ESI-IT-TOF (*m/z*): [MH⁺] calcd for $C_{12}H_{11}O_5$, 235.0607; found, 235.1326. These

data are in complete agreement with literature values for **16**.[6]
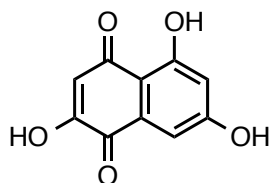
2,5,7-trihydroxynaphthalene-1,4-dione (flaviolin, 17).

UV (acetonitrile [*aq*.]) λ_{max}, nm: 216, 262, 306, 446. HPLC retention time: 18.5 min.

HRMS–ESI/APCI-TOF (*m/z*): [MH⁺] calcd for $C_{10}H_7O_5$, 207.0294; found, 207.0287. MS

data was collected on a Waters Acquity/Xeno-G2 UPLC-MS in positive ion mode. These data are in complete agreement with literature values for **17**.[6]
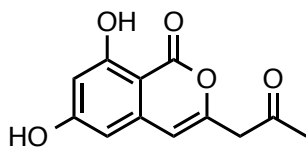
6,8-dihydroxy-3-(2-oxopropyl)-1*H*-isochromen-1-one (18).

UV (acetonitrile [*aq*.]) λ_{max}, nm: 244, 278, 328. HPLC retention time: 20.0 min. HRMS– ESI-IT-TOF (*m/z*): [MH⁺] calcd for $C_{12}H_{11}O_5$, 235.0607; found, 235.0603. These data are in complete agreement with literature values for **18**.[6]
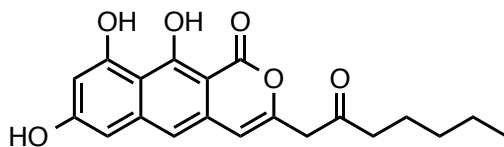
7,9,10-trihydroxy-3-(2-oxoheptyl)-1*H*-benzo[*g*]isochromen-1-one (norpyrone, 19).

UV (acetonitrile [*aq*.]) λ_{max}, nm: 272, 280, 390. HPLC retention time: 35.1 min. HRMS– ESI-IT-TOF (*m/z*): [MH⁺] calcd for $C_{20}H_{21}O_6$, 357.1338; found, 357.1340. These data are in complete agreement with literature values for **19**.[1]
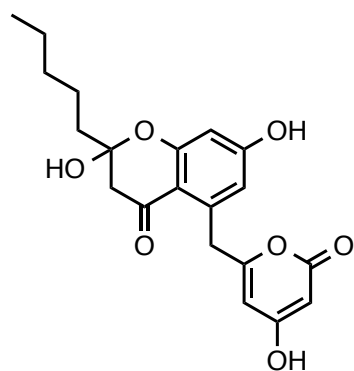
2,7-dihydroxy-5-((4-hydroxy-2-oxo-2*H*-pyran-6-yl)methyl)-2-pentylchroman-4-one (Hex-SEK4, 20).

UV (acetonitrile [*aq*.]) λ_{max}, nm: 232, 278. HPLC retention time: 24.7 min. HRMS–ESI-IT-TOF (*m/z*): [MH$^+$] calcd for $C_{20}H_{23}O_7$, 375.1444; found, 375.1441. Hex-SEK4 coelutes with its isomer Hex-SEK4b, **21**. These data are in complete agreement with literature values for **20**.[13]
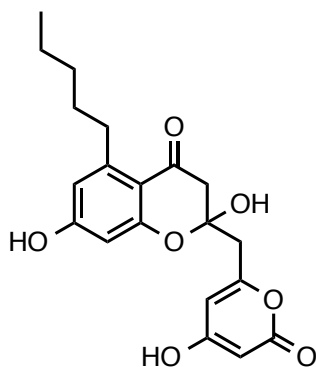
2,7-dihydroxy-2-((4-hydroxy-2-oxo-2*H*-pyran-6-yl)methyl)-5-pentylchroman-4-one (Hex-SEK4b, 21).

UV (acetonitrile [*aq*.]) λ_{max}, nm: 232, 278. HPLC retention time: 24.7 min. HRMS–ESI-IT-TOF (*m/z*): [MH$^+$] calcd for $C_{20}H_{23}O_7$, 375.1444; found, 375.1441. Hex-SEK4b coelutes with its isomer Hex-SEK4, **20**. These data are in complete agreement with literature values for **21**.[3]
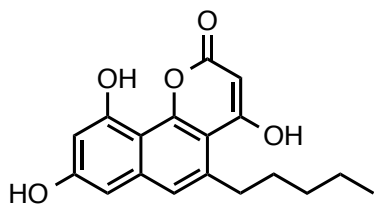
4,8,10-trihydroxy-5-pentyl-2*H*-benzo[*h*]chromen-2-one (Hex-pannorin, 22).

UV (acetonitrile [*aq*.]) λ_{max}, nm: 232, 278, 288, 322, 368. HPLC retention time: 28.5 min. HRMS–ESI-IT-TOF (*m/z*): [MH$^+$] calcd for $C_{18}H_{19}O_5$, 315.1233; found, 315.1220. The λ_{max} and shape of the UV-vis spectrum for this molecule are similar to those for pannorin (**23**), suggesting a shared core architecture (see Figures 3.11, D.5).[14] Taken together,

these data support the proposed structural assignment. This assignment is in agreement

with polyketide cyclization logic.
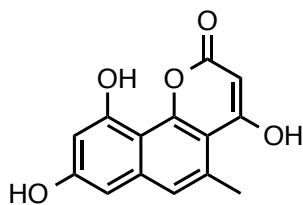
4,8,10-trihydroxy-5-methyl-2*H*-benzo[*h*]chromen-2-one (pannorin, 23).

UV (acetonitrile [*aq*.]) λ_{max}, nm: 230, 276, 288, 320, 364. HPLC retention time: 20.1 min.

HRMS–ESI-IT-TOF (*m/z*): calcd for $C_{14}H_{11}O_5$ [MH$^+$], 259.0607; found, 259.0606. These

data are in complete agreement with literature values for **23**.[5, 14]

3,6,8,9-tetrahydroxy-1-oxo-3-pentyl-1,2,3,4-tetrahydroanthracene-2-carboxylic acid

(Hex-atrochrysone carboxylic acid, 24).

UV (acetonitrile [*aq*.]) λ_{max}, nm: 224, 268, 318, 378. HPLC retention time: 29.0 min.

HRMS–ESI-IT-TOF (*m/z*): [MH$^+$], calcd for $C_{20}H_{23}O_7$, 375.1444; found, 375.1440. The

λ_{max} and shape of the UV-vis spectrum for this molecule are similar to those for

atrochrysone, suggesting a shared core architecture (see Figures 3.7, D.5).[12] Taken

together, these data support the proposed structural assignment. This assignment is in

agreement with polyketide cyclization logic.
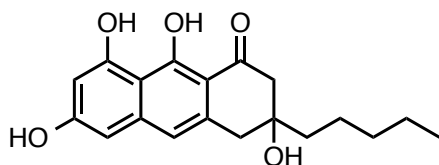
3,6,8,9-tetrahydroxy-3-pentyl-3,4-dihydroanthracen-1(2*H*)-one (Hex-atrochrysone, 25).

UV (acetonitrile [*aq.*]) λ_{max}, nm: 228, 272, 320, 294. HPLC retention time: 32.6 min.

HRMS–ESI-IT-TOF (*m/z*): [MH$^+$] calcd for $C_{19}H_{23}O_5$, 331.1546; found, 331.1535. The

λ_{max} and shape of the UV-vis spectrum for this molecule are similar to those for

atrochrysone, suggesting a shared core architecture (see Figures 3.7, D.5).[12] Taken

together, these data support the proposed structural assignment. This assignment is in

agreement with polyketide cyclization logic.
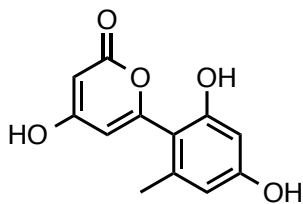
6-(2,4-dihydroxy-6-methylphenyl)-4-hydroxy-2*H*-pyran-2-one (31).

UV (acetonitrile [*aq.*]) λ_{max}, nm: 304. HPLC retention time: 14.3 min. HRMS–ESI/APCI-

TOF (*m/z*): [MH$^+$] calcd for $C_{12}H_{11}O_5$, 235.0607; found, 235.0604. MS data was collected

on a Waters Acquity/Xeno-G2 UPLC-MS in positive ion mode. A synthetic standard of

this material was provided by Dr. Ikuro Abe. HPLC chromatogram for this product is

presented in Figure D.7. These data are in complete agreement with the standard.[9c, 15]
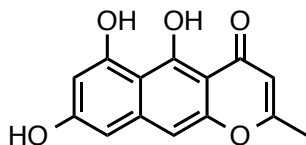
5,6,8-trihydroxy-2-methyl-4*H*-benzo[*g*]chromen-4-one (*nor*-rubrofusarin, 32).
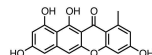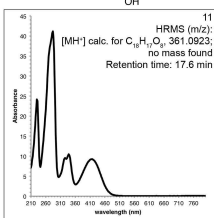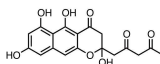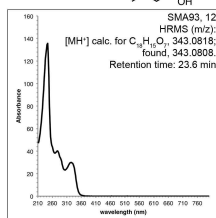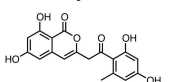
UV (acetonitrile [*aq*.]) λ_{max}, nm: 224, 278, 328, 412. HPLC retention time: 27.0 min.

HRMS–ESI/APCI-TOF (*m/z*): [MH$^+$] calcd for $C_{14}H_{11}O_5$, 259.0607; found, 259.0599. MS

data was collected on a Waters Acquity/Xeno-G2 UPLC-MS in positive ion mode. These

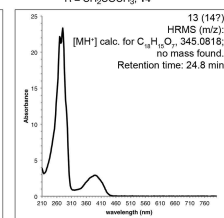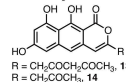data are in complete agreement with literature values for **32**.[16]
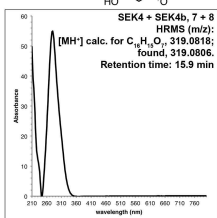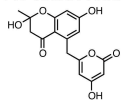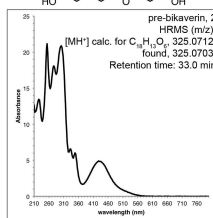
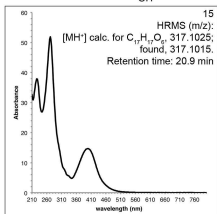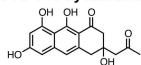**Appendix Figure D.3│Proposed structures, HRMS data, UV-vis spectra, and retention times for products of combinatorial reactions with Pks4 SAT-KS-MAT.** Previously characterized species include SEK4 (**7**),[9a] SEK4b (**8**),[10] SMA93 (**12**),[7] pre-bikaverin (**2**).[7]
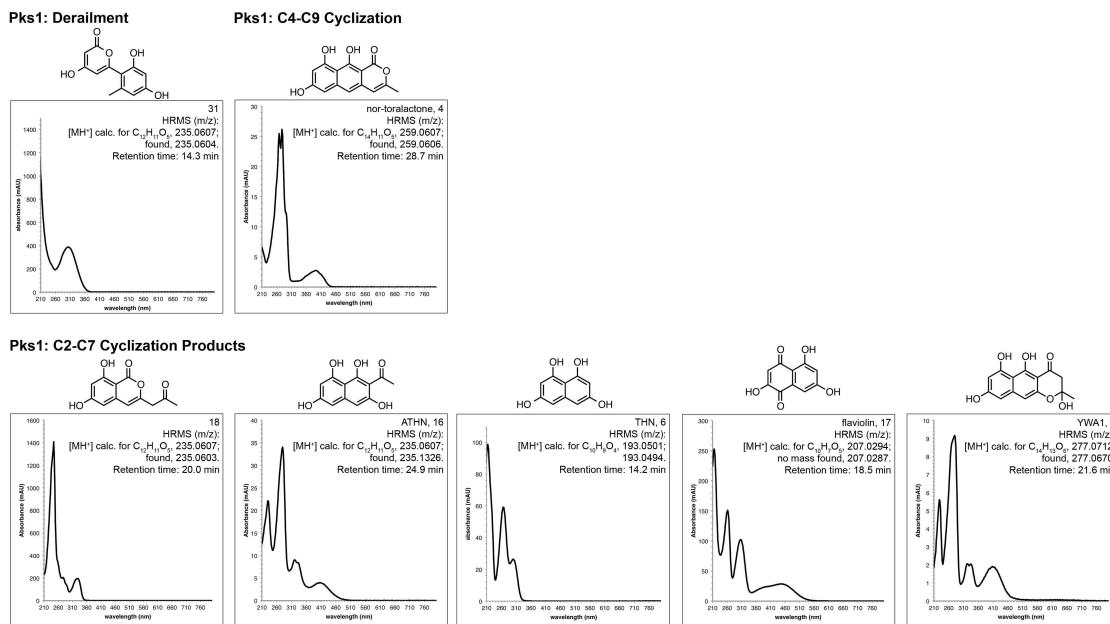
Appendix Figure D.4│ Proposed structures, HRMS data, UV-vis spectra, and retention times for products of combinatorial reactions with Pks1 SAT-KS-MAT. Previously characterized species include 6-(2,4-dihydroxy-6-methylphenyl)-4-hydroxy-2-pyrone (**31**), isocoumarin **18**,[6] ATHN (**16**),[6] THN (**6**), flaviolin (**17**),[6] YWA1 (**5**),[17] and *nor*-toralactone (**4**).[5]
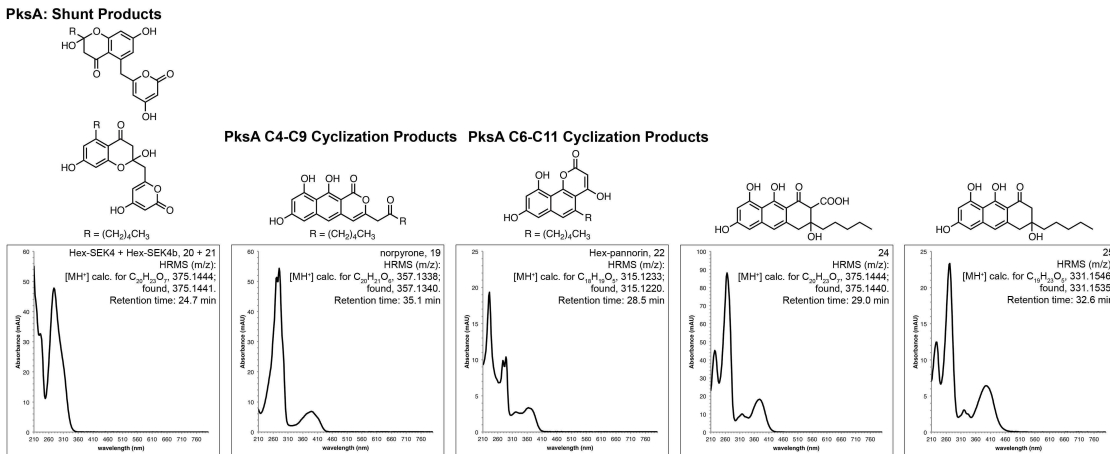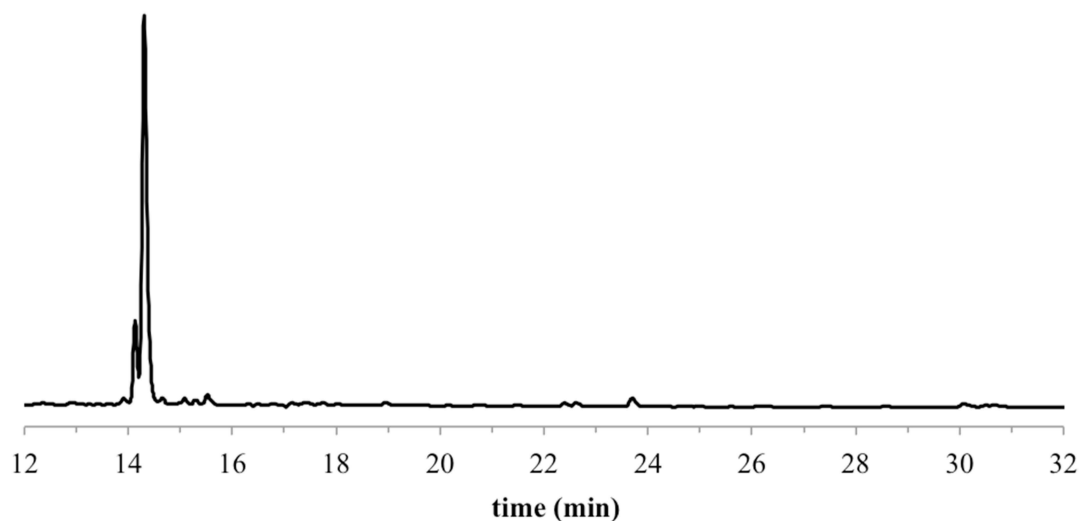

Appendix Figure D.5│ Proposed structures, HRMS data, UV-vis spectra, and retention times for products of combinatorial reactions with PksA SAT-KS-MAT. Previously characterized species include hex-SEK4 (**20**),[13] hex-SEK4b (**21**),[13] and norpyrone (**19**).[1]

218

Appendix Figure D.6 │ HPLC chromatogram (280 nm) of a synthetic standard of 6-(2,4-dihydroxy-6-methylphenyl)-4-hydroxy-2-pyrone (31).
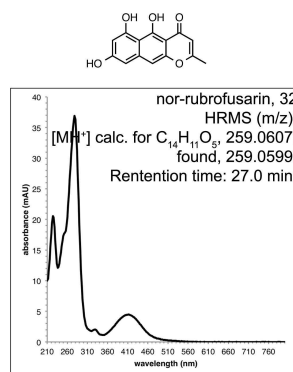
Appendix Figure D.7 │ HRMS, UV-vis, and retention time data for *nor*-rubrofusarin (32). Data is in accord with literature values.[16]

D.5. References

[1]   J. M. Crawford, P. M. Thomas, J. R. Scheerer, A. L. Vagstad, N. L. Kelleher, C. A. Townsend, *Science* **2008**, *320*, 243-246.

[2]   J. M. Crawford, B. C. R. Dancy, E. A. Hill, D. W. Udwary, C. A. Townsend, *Proc Natl Acad Sci USA* **2006**, *103*, 16728-16733.

[3]   A. L. Vagstad, A. G. Newman, P. A. Storm, K. Belecki, J. M. Crawford, C. A. Townsend, *Angew Chem Int Ed Engl* **2013**, *52*, 1718-1721.

[4]   J. M. Crawford, A. L. Vagstad, K. R. Whitworth, K. C. Ehrlich, C. A. Townsend, *Chembiochem* **2008**, *9*, 1019-1023.

[5]   A. G. Newman, A. L. Vagstad, K. Belecki, J. R. Scheerer, C. A. Townsend, *Chem Commun* **2012**, *48*, 11772-11774.

[6]   A. L. Vagstad, E. A. Hill, J. W. Labonte, C. A. Townsend, *Chem Biol* **2012**, *19*, 1525-1534.

[7]   S. M. Ma, J. Zhan, K. Watanabe, X. Xie, W. Zhang, C. C. Wang, Y. Tang, *J Am Chem Soc* **2007**, *129*, 10642-10643.

[8]   A. Watanabe, I. Fujii, U. Sankawa, M. E. Mayorga, W. E. Timberlake, Y. Ebizuka, *Tetrahedron Lett* **1999**, *40*, 91-94.

[9]   (a) H. Fu, S. Ebertkhosla, D. A. Hopwood, C. Khosla, *J Am Chem Soc* **1994**, *116*, 4166-4170; (b) W. J. Zhang, Y. R. Li, Y. Tang, *Proc Natl Acad Sci USA* **2008**, *105*, 20683-20688; (c) I. Abe, S. Oguro, Y. Utsumi, Y. Sano, H. Noguchi, *J Am Chem Soc* **2005**, *127*, 12709-12716.

[10]   H. Fu, D. A. Hopwood, C. Khosla, *Chem Biol* **1994**, *1*, 205-210.

[11]   S. M. Ma, J. X. Zhan, X. K. Xie, K. J. Watanabe, Y. Tang, W. J. Zhang, *J Am Chem Soc* **2008**, *130*, 38-39.

[12]   (a) T. Awakawa, K. Yokota, N. Funa, F. Doi, N. Mori, H. Watanabe, S. Horinouchi, *Chem Biol* **2009**, *16*, 613-623; (b) M. Gill, P. M. Morgan, *ARKIVOC* **2001**, *2001*, 145-156.

[13]   A. L. Vagstad, S. B. Bumpus, K. Belecki, N. L. Kelleher, C. A. Townsend, *J Am Chem Soc* **2012**, *134*, 6865-6877.

[14]   H. Ogawa, K. Hasumi, K. Sakai, S. Murakawa, A. Endo, *J Antibiot* **1991**, *44*, 762-767.

[15]   K. Springob, S. Samappito, A. Jindaprasert, J. Schmidt, J. E. Page, W. De-Eknamkul, T. M. Kutchan, *FEBS J* **2007**, *274*, 406-417.

[16]   R. J. N. Frandsen, C. Schutt, B. W. Lund, D. Staerk, J. Nielsen, S. Olsson, H. Giese, *J Biol Chem* **2011**, *286*, 10419-10428.

[17]   I. Fujii, A. Watanabe, U. Sankawa, Y. Ebizuka, *Chem Biol* **2001**, *8*, 189-197.

Appendix E: Supplementary Material to Chapter 4

E.1. Cloning of Chimeric NR-PKS
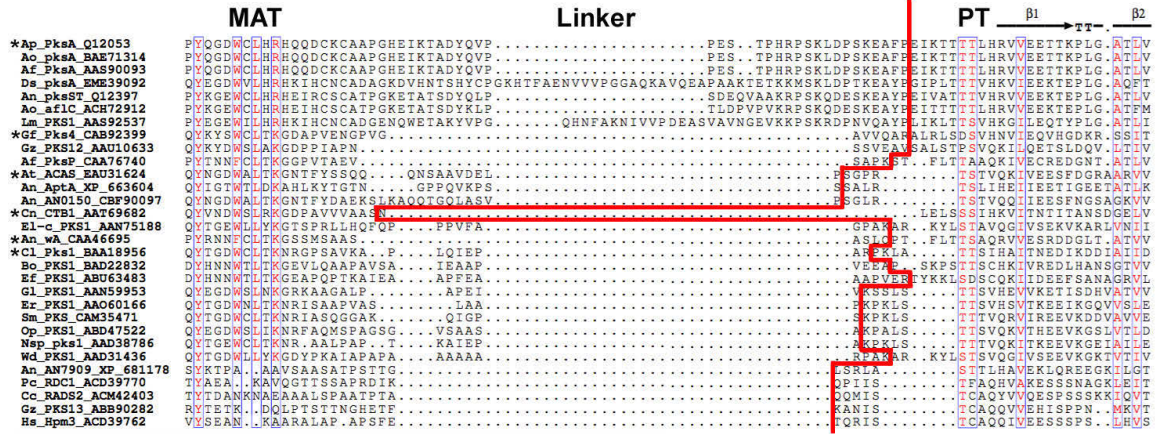
Appendix Figure E.1│Multiple sequence alignment of NR-PKSs for the MAT-PT linker region. NR-PKS sequences used in this alignment correspond to known polyketide products. Accession numbers for each sequence are indicated. Stars indicate sequences used in this study (PksA from *A. parasiticus*, Pks4 from *G. fujikuroi*, ACAS from *A. terreus*, CTB1 from *C. nicotianae*, wA from *A. nidulans*, and Pks1 from *C. lagenaria*). The secondary structure at the N-terminus of the PksA PT domain is shown. The crossover site for two-part heterocombinations is shown as a red line. It was selected to be 4–5 amino acids upstream of the conserved Thr at the N-terminal boundary of the PT. This figure was prepared using the ENDscript server.[1]

E.1.1. Gibson Assembly of Chimeric NR-PKS Constructs

Appendix Table E.1│PCR reagents used to generate DNA parts for Gibson assembly of two-part NR-PKS chimeras.

PCR Product	Template	Forward Primer	Reverse Primer
M1P2-N	pEPksA	T7	M1P2-3
M1P3-N	pEPksA	T7	M1P3-3
M1P4-N	pEPksA	T7	M1P4-3
M1*P5-N	pEPksA	T7	M1*P5-3
M1*P6-N	pEPksA	T7	M1*P6-3
M2P1-N	pEPks4	T7	M2P1-3
M2P3-N	pEPks4	T7	M2P3-3
M2P4-N	pEPks4	T7	M2P4-3
M2P5-N	pEPks4	T7	M2P5-3
M2P6-N	pEPks4	T7	M2P6-3
M3P1-N	pEACAS	T7	M3P1-3
M3P2-N	pEACAS	T7	M3P2-3
M3P4-N	pEACAS	T7	M3P4-3
M3P5-N	pEACAS	T7	M3P5-3
M3P6-N	pEACAS	T7	M3P6-3

M4P1-N	pECTB1	T7	M4P1-3
M4P2-N	pECTB1	T7	M4P2-3
M4P3-N	pECTB1	T7	M4P3-3
M4P5-N	pECTB1	T7	M4P5-3
M4P6-N	pECTB1	T7	M4P6-3
M5P1-N	pEwANKA2	T7	M5P1-3
M5P2-N	pEwANKA2	T7	M5P2-3
M5P3-N	pEwANKA2	T7	M5P3-3
M5P4-N	pEwANKA2	T7	M5P4-3
M5P6-N	pEwANKA2	T7	M5P6-3
M6P1-N	pEPks1alt	T7	M6P1-3
M6P2-N	pEPks1alt	T7	M6P2-3
M6P3-N	pEPks1alt	T7	M6P3-3
M6P4-N	pEPks1alt	T7	M6P4-3
M6P5-N	pEPks1alt	T7	M6P5-3
M1P2-C	pEPks4	M1P2-5	T7 term
M1P3-C	pEACAS	M1P3-5	T7 term
M1P4-C	pECTB1	M1P4-5	T7 term
M1*P5-C	pEwa-PTA	M1*P5-5	T7 term
M1*P6-C	pEPks1alt	M1*P6-5	T7 term
M2P1-C	pEPksA	M2P1-5	T7 term
M2P3-C	pEACAS	M2P3-5	T7 term
M2P4-C	pECTB1	M2P4-5	T7 term
M2P5-C	pEwa-PTA	M2P5-5	T7 term
M2P6-C	pEPks1alt	M2P6-5	T7 term
M3P1-C	pEPksA	M3P1-5	T7 term
M3P2-C	pEPks4	M3P2-5	T7 term
M3P4-C	pECTB1	M3P4-5	T7 term
M3P5-C	pEwa-PTA	M3P5-5	T7 term
M3P6-C	pEPks1alt	M3P6-5	T7 term
M4P1-C	pEPksA	M4P1-5	T7 term
M4P2-C	pEPks4	M4P2-5	T7 term
M4P3-C	pEACAS	M4P3-5	T7 term
M4P5-C	pEwa-PTA	M4P5-5	T7 term
M4P6-C	pEPks1alt	M4P6-5	T7 term
M5P1-C	pEPksA	M5P1-5	T7 term
M5P2-C	pEPks4	M5P2-5	T7 term
M5P3-C	pEACAS	M5P3-5	T7 term
M5P4-C	pECTB1	M5P4-5	T7 term
M5P6-C	pEPks1alt	M5P6-5	T7 term
M6P1-C	pEPksA	M6P1-5	T7 term
M6P2-C	pEPks4	M6P2-5	T7 term
M6P3-C	pEACAS	M6P3-5	T7 term
M6P4-C	pECTB1	M6P4-5	T7 term
M6P5-C	pEwA-PTA	M6P5-5	T7 term

Appendix Table E.2│DNA components used in Gibson assembly reactions of two-part NR-PKS chimeras.

Plasmid	Part 1	Part 2	Part 3
pEM1P2-2	M1P2-N	M1P2-C	pET-24a(+)-NdeI-NotI
pEM1P3-2	M1P3-N	M1P3-C	pET-24a(+)-NdeI-NotI
pEM1P4-1	M1P4-N	M1P4-C	pET-24a(+)-NdeI-NotI
pEM1*P5-1[a]	M1*P5-N[a]	M1*P5-C[a]	pET-24a(+)-NdeI-NotI
pEM1*P6-2[a]	M1*P6-N[a]	M1*P6-C[a]	pET-24a(+)-NdeI-NotI
pEM2P1-3	M2P1-N	M2P1-C	pET-24a(+)-NdeI-NotI
pEM2P3-2	M2P3-N	M2P3-C	pET-24a(+)-NdeI-NotI
pEM2P4-4	M2P4-N	M2P4-C	pET-24a(+)-NdeI-NotI
pEM2P5-2	M2P5-N	M2P5-C	pET-24a(+)-NdeI-NotI
pEM2P6-3	M2P6-N	M2P6-C	pET-24a(+)-NdeI-NotI
pEM3P1-4	M3P1-N	M3P1-C	pET-24a(+)-NdeI-NotI
pEM3P2-1	M3P2-N	M3P2-C	pET-24a(+)-NdeI-NotI
pEM3P4-2	M3P4-N	M3P4-C	pET-24a(+)-NdeI-NotI
pEM3P5-1	M3P5-N	M3P5-C	pET-24a(+)-NdeI-NotI
pEM3P6-3	M3P6-N	M3P6-C	pET-24a(+)-NdeI-NotI
pEM4P1-7	M4P1-N	M4P1-C	pET-24a(+)-NdeI-NotI
pEM4P2-3	M4P2-N	M4P2-C	pET-24a(+)-NdeI-NotI
pEM4P3-8	M4P3-N	M4P3-C	pET-24a(+)-NdeI-NotI
pEM4P5-1	M4P5-N	M4P5-C	pET-24a(+)-NdeI-NotI
pEM4P6-1	M4P6-N	M4P6-C	pET-24a(+)-NdeI-NotI
pEM5P1-1	M5P1-N	M5P1-C	pET-24a(+)-NdeI-NotI
pEM5P2-1	M5P2-N	M5P2-C	pET-24a(+)-NdeI-NotI
pEM5P3-1	M5P3-N	M5P3-C	pET-24a(+)-NdeI-NotI
pEM5P4-1	M5P4-N	M5P4-C	pET-24a(+)-NdeI-NotI
pEM5P6-1	M5P6-N	M5P6-C	pET-24a(+)-NdeI-NotI
pEM6P1-1	M6P1-N	M6P1-C	pET-24a(+)-NdeI-NotI
pEM6P2-2	M6P2-N	M6P2-C	pET-24a(+)-NdeI-NotI
pEM6P3-1	M6P3-N	M6P3-C	pET-24a(+)-NdeI-NotI
pEM6P4-2	M6P4-N	M6P4-C	pET-24a(+)-NdeI-NotI
pEM6P5-1	M6P5-N	M6P5-C	pET-24a(+)-NdeI-NotI

[a] A silent mutation is introduced at the crossover site to make synthesis of these constructs easier.

E.1.2. Gibson Assembly of Engineered Topopyrone Synthases

Appendix Table E.3│PCR reagents used to generate DNA parts for Gibson assembly of topopyrone synthases.

PCR Product	Template	Forward Primer	Reverse Primer
M2P4A1-N	pEM2P4-4	T7	P4A1-3
M2P1-G1491L-N	pEM2P1-3	T7	P1-G1491L-40-3
M2P1-G1491I-N	pEM2P1-3	T7	P1-G1491I-40-3
M2P1-G1491V-N	pEM2P1-3	T7	P1-G1491V-3
M2P1-G1491A-N	pEM2P1-3	T7	P1-G1491A-3
M2P4A1-I1468V-N	pEM2P4A1	T7	P4-I1468V-40-3
M2P4A1-I1468A-N	pEM2P4A1	T7	P4-I1468V-40-3
M2P4A1-I1468G-N	pEM2P4A1	T7	P4-I1468V-40-3
M2P1^{P-}-G1491L-N	pEM2P1-GL	T7	P1-H1345A-3
Pks1^{P-}-N	pEPks1alt	T7	P6-H1330A-3
M2P4A1-C	pEPksA	P4A1-5	T7 term
M2P1-G1491L-C	pEM2P1-3	P1-G1491L-40-5	T7 term
M2P1-G1491I-C	pEM2P1-3	P1-G1491I-40-5	T7 term
M2P1-G1491V-C	pEM2P1-3	P1-G1491V-5	T7 term
M2P1-G1491A-C	pEM2P1-3	P1-G1491A-5	T7 term
M2P4A1-I1468V-C	pEM2P4A1	P4-I1468V-40-5	T7 term
M2P4A1-I1468A-C	pEM2P4A1	P4-I1468A-40-5	T7 term
M2P4A1-I1468G-C	pEM2P4A1	P4-I1468G-40-5	T7 term
M2P1^{P-}-G1491L-C	pEM2P1-GL	P1-H1345A-5	T7 term
M2P1^{P-T-}-G1491L-C1	pEM2P1-GL	P1-H1345A-5	T1-S1937A-3
M2P1^{P-T-}-G1491L-C2	pEM2P1-GL	T1-S1937A-5	T7 term
Pks1^{P-}-C	pEPks1alt	P6-H1330S-5	T7 term
Pks1^{P-T-}-C1	pEPks1alt	P6-H1330S-5	T6-S2009A-3
Pks1^{P-T-}-C2	pEPks1alt	T6-S2009A-5	T7 term

Appendix Table E.4│ DNA components used in Gibson assembly reactions of topopyrone synthases.

Plasmid	Part 1	Part 2	Part 3	Part 4
pEM2P4A1	M2P4A1-N	M2P4A1-C	pET-24a(+)-NdeI-NotI	
pEM2P1-GL	M2P1-G1491L-N	M2P1-G1491L-C	pET-24a(+)-NdeI-NotI	
pEM2P1-GI	M2P1-G1491I-N	M2P1-G1491I-C	pET-24a(+)-NdeI-NotI	
pEM2P1-GV	M2P1-G1491V-N	M2P1-G1491V-C	pET-24a(+)-NdeI-NotI	
pEM2P1-GA	M2P1-G1491A-N	M2P1-G1491A-C	pET-24a(+)-NdeI-NotI	
pEM2P4A1-IV	M2P4A1-I1468V-N	M2P4A1-I1468V-C	pET-24a(+)-NdeI-NotI	
pEM2P4A1-IA	M2P4A1-I1468A-N	M2P4A1-I1468A-C	pET-24a(+)-NdeI-NotI	
pEM2P4A1-IG	M2P4A1-I1468G-N	M2P4A1-I1468G-C	pET-24a(+)-NdeI-NotI	
pEM2P1^{P-}-GL	M2P1^{P-}-G1491L-N	M2P1^{P-}-G1491L-C	pET-24a(+)-NdeI-NotI	
pEM2P1^{P-T-}-GL	M2P1^{P-}-G1491L-N	M2P1^{P-T-}-G1491L-C1	M2P1^{P-T-}-G1491L-C2	pET-24a(+)-NdeI-NotI
pEPks1^{P-}	Pks1^{P-}-N	Pks1^{P-}-C	pET-24a(+)-NdeI-NotI	
pEPks1^{P-T-}	Pks1^{P-}-N	Pks1^{P-T-}-C1	Pks1^{P-T-}-C2	pET-24a(+)-NdeI-NotI

E.1.3. Primer Sequences

Appendix Table E.5│Sequences for primers used in this study.

Primer	Sequence
T7	TAATACGACTCACTATAGGG
M1P2-5	CTAGCAAGGAGGCCTTCCCCGCTTTACGTCTTTCTGATAG
M1P3-5	CTAGCAAGGAGGCCTTCCCCTCCGGCCCGAGGACTTCAAC
M1P4-5	CTAGCAAGGAGGCCTTCCCCAATTTGGAACTCTCTTCGTC
M1*P5-5	CTAGCAAGGAGGCCTTCCCACCGACGTTTTTAACCACCAG
M1*P6-5	CTAGCAAGGAGGCCTTCCCACCCAAACTGGCGACCACCTC
M2P1-5	GTGCAGTTGTACAAGCGCGCGAAATCAAGACCACCACGAC
M2P3-5	GTGCAGTTGTACAAGCGCGCTCCGGCCCGAGGACTTCAAC
M2P4-5	GTGCAGTTGTACAAGCGCGCAATTTGGAACTCTCTTCGTC
M2P5-5	GTGCAGTTGTACAAGCGCGCCCGACGTTTTTAACCACCAG
M2P6-5	GTGCAGTTGTACAAGCGCGCCCCAAACTGGCGACCACCTC
M3P1-5	CCGCCGTGGATGAATTGCCAGAAATCAAGACCACCACGAC
M3P2-5	CCGCCGTGGATGAATTGCCAGCTTTACGTCTTTCTGATAG
M3P4-5	CCGCCGTGGATGAATTGCCAAATTTGGAACTCTCTTCGTC
M3P5-5	CCGCCGTGGATGAATTGCCACCGACGTTTTTAACCACCAG
M3P6-5	CCGCCGTGGATGAATTGCCACCCAAACTGGCGACCACCTC
M4P1-5	CCGTGGTTGTTGCCGCGTCAGAAATCAAGACCACCACGAC
M4P2-5	CCGTGGTTGTTGCCGCGTCAGCTTTACGTCTTTCTGATAG
M4P3-5	CCGTGGTTGTTGCCGCGTCATCCGGCCCGAGGACTTCAAC
M4P5-5	CCGTGGTTGTTGCCGCGTCACCGACGTTTTTAACCACCAG
M4P6-5	CCGTGGTTGTTGCCGCGTCACCCAAACTGGCGACCACCTC
M5P1-5	CCGCGAGCGCGTCTCTACAGGAAATCAAGACCACCACGAC
M5P2-5	CCGCGAGCGCGTCTCTACAGGCTTTACGTCTTTCTGATAG
M5P3-5	CCGCGAGCGCGTCTCTACAGTCCGGCCCGAGGACTTCAAC
M5P4-5	CCGCGAGCGCGTCTCTACAGAATTTGGAACTCTCTTCGTC
M5P6-5	CCGCGAGCGCGTCTCTACAGCCCAAACTGGCGACCACCTC
M6P1-5	TCCAGATCGAGCCTGCCAGAGAAATCAAGACCACCACGAC
M6P2-5	TCCAGATCGAGCCTGCCAGAGCTTTACGTCTTTCTGATAG
M6P3-5	TCCAGATCGAGCCTGCCAGATCCGGCCCGAGGACTTCAAC
M6P4-5	TCCAGATCGAGCCTGCCAGAAATTTGGAACTCTCTTCGTC
M6P5-5	TCCAGATCGAGCCTGCCAGACCGACGTTTTTAACCACCAG
P4A1-5	AAGTGACTCCGCAAAACAAGGCAAAAGGAGTAGGTGTCAG
P1-G1491L-40-5	CGATACAATAGGAAGACCTTGTACAAGCTCATGAGCAGC
P1-G1491I-40-5	CGATACAATAGGAAGACCATTTACAAGCTCATGAGCAGC
P1-G1491V-5	TGCGATACAATAGGAAGACCGTGTAC
P1-G1491A-5	TGCGATACAATAGGAAGACCGCGTAC
P4-I1468V-40-5	CGCTTCAGCAAAGGCATGGTGTACAAGATGATCGGCCAA
P4-I1468A-40-5	CGCTTCAGCAAAGGCATGGCGTACAAGATGATCGGCCAA
P4-I1468G-40-5	CGCTTCAGCAAAGGCATGGGCTACAAGATGATCGGCCAA
P1-H1345A-5	AACGGCCTCGCTCGAGGGGCGCTTGTCGATGGGATCCCT
T1-S1937A-5	TATCACCTGGGCGGCTGGGCGTCCGGTGGTGCATTCGCT
P6-H1330S-5	CGTCCTTTGGTCGAGGGCGCGTTGTGCAATGGCACCCCT
T6-S2009A-5	TACTCCTTGGCGGGTTGGGCGGCGGGCGGTGTCATTGCC

T7 term	GCTAGTTATTGCTCAGCGG
M1P2-3	CTATCAGAAAGACGTAAAGCGGGGAAGGCCTCCTTGCTAG
M1P3-3	GTTGAAGTCCTCGGGCCGGAGGGGAAGGCCTCCTTGCTAG
M1P4-3	GACGAAGAGAGTTCCAAATTGGGGAAGGCCTCCTTGCTAG
M1*P5-3	CTGGTGGTTAAAAACGTCGGTGGGAAGGCCTCCTTGCTAG
M1*P6-3	GAGGTGGTCGCCAGTTTGGGTGGGAAGGCCTCCTTGCTAG
M2P1-3	GTCGTGGTGGTCTTGATTTCGCGCGCTTGTACAACTGCAC
M2P3-3	GTTGAAGTCCTCGGGCCGGAGCGCGCTTGTACAACTGCAC
M2P4-3	GACGAAGAGAGTTCCAAATTGCGCGCTTGTACAACTGCAC
M2P5-3	CTGGTGGTTAAAAACGTCGGGCGCGCTTGTACAACTGCAC
M2P6-3	GAGGTGGTCGCCAGTTTGGGGCGCGCTTGTACAACTGCAC
M3P1-3	GTCGTGGTGGTCTTGATTTCTGGCAATTCATCCACGGCGG
M3P2-3	CTATCAGAAAGACGTAAAGCTGGCAATTCATCCACGGCGG
M3P4-3	GACGAAGAGAGTTCCAAATTTGGCAATTCATCCACGGCGG
M3P5-3	CTGGTGGTTAAAAACGTCGGTGGCAATTCATCCACGGCGG
M3P6-3	GAGGTGGTCGCCAGTTTGGGTGGCAATTCATCCACGGCGG
M4P1-3	GTCGTGGTGGTCTTGATTTCTGACGCGGCAACAACCACGG
M4P2-3	CTATCAGAAAGACGTAAAGCTGACGCGGCAACAACCACGG
M4P3-3	GTTGAAGTCCTCGGGCCGGATGACGCGGCAACAACCACGG
M4P5-3	CTGGTGGTTAAAAACGTCGGTGACGCGGCAACAACCACGG
M4P6-3	GAGGTGGTCGCCAGTTTGGGTGACGCGGCAACAACCACGG
M5P1-3	GTCGTGGTGGTCTTGATTTCCTGTAGAGACGCGCTCGCGG
M5P2-3	CTATCAGAAAGACGTAAAGCCTGTAGAGACGCGCTCGCGG
M5P3-3	GTTGAAGTCCTCGGGCCGGACTGTAGAGACGCGCTCGCGG
M5P4-3	GACGAAGAGAGTTCCAAATTCTGTAGAGACGCGCTCGCGG
M5P6-3	GAGGTGGTCGCCAGTTTGGGCTGTAGAGACGCGCTCGCGG
M6P1-3	GTCGTGGTGGTCTTGATTTCTCTGGCAGGCTCGATCTGGA
M6P2-3	CTATCAGAAAGACGTAAAGCTCTGGCAGGCTCGATCTGGA
M6P3-3	GTTGAAGTCCTCGGGCCGGATCTGGCAGGCTCGATCTGGA
M6P4-3	GACGAAGAGAGTTCCAAATTTCTGGCAGGCTCGATCTGGA
M6P5-3	CTGGTGGTTAAAAACGTCGGTCTGGCAGGCTCGATCTGGA
P4A1-3	CTGACACCTACTCCTTTTGCCTTGTTTTGCGGAGTCACTT
P1-G1491L-40-3	GCTGCTCATGAGCTTGTACAAGGTCTTCCTATTGTATCG
P1-G1491I-40-3	GCTGCTCATGAGCTTGTAAATGGTCTTCCTATTGTATCG
P1-G1491V-3	ATGCTGCTCATGAGCTTGTACACGGT
P1-G1491A-3	ATGCTGCTCATGAGCTTGTACGCGGT
P4-I1468V-40-3	TTGGCCGATCATCTTGTACACCATGCCTTTGCTGAAGCG
P4-I1468A-40-3	TTGGCCGATCATCTTGTACGCCATGCCTTTGCTGAAGCG
P4-I1468G-40-3	TTGGCCGATCATCTTGTAGCCCATGCCTTTGCTGAAGCG
P1-H1345A-3	AGGGATCCCATCGACAAGCGCCCCTCGAGCGAGGCCGTT
P6-H1330A-3	AGGGGTGCCATTGCACAACGCGCCCTCGACCAAAGGACG
T1-S1937A-3	AGCGAATGCACCACCGGACGCCCAGCCGCCCAGGTGATA
T6-S2009A-3	GGCAATGACACCGCCCGCCGCCCAACCCGCCAAGGAGTA

E.1.4. Protein Expression Constructs

Appendix Table E.6│Details of chimeric NR-PKS expression constructs used in this study.

Protein	Encodes		Plasmid	Vector	Tag	MW (kDa)
	SAT-KS-MAT	PT-ACP-TE				
M1P2	PksA (M1–P1304)	Pks4 (A1285–N2036)	pEM1P2-2	pET-24a(+)	C-His	226.6
M1P3	PksA (M1–P1304)	ACAS (S1318–S1771)	pEM1P3-2	pET-24a(+)	C-His	194.3
M1P4	PksA (M1–P1304)	CTB1 (N1294–S2196)	pEM1P4-1	pET-24a(+)	C-His	242.6
M1P5	PksA (M1–P1304)	wA (P1286–V2157)	pEM1*P5-1	pET-24a(+)	C-His	240.2
M1P6	PksA (M1–P1304)	Pks1 (P1290–E2183)	pEM1*P6-2	pET-24a(+)	C-His	241.8
M2P1	Pks4 (M1–R1284)	PksA (E1305–A2109)	pEM2P1-3	pET-24a(+)	C-His	228.2
M2P3	Pks4 (M1–R1284)	ACAS (S1318–S1771)	pEM2P3-2	pET-24a(+)	C-His	190.5
M2P4	Pks4 (M1–R1284)	CTB1 (N1294–S2196)	pEM2P4-4	pET-24a(+)	C-His	238.8
M2P5	Pks4 (M1–R1284)	wA (P1286–V2157)	pEM2P5-2	pET-24a(+)	C-His	236.4
M2P6	Pks4 (M1–R1284)	Pks1 (P1290–E2183)	pEM2P6-3	pET-24a(+)	C-His	238.0
M3P1	ACAS (M1–P1317)	PksA (E1305–A2109)	pEM3P1-4	pET-24a(+)	C-His	233.1
M3P2	ACAS (M1–P1317)	Pks4 (A1285–N2036)	pEM3P2-1	pET-24a(+)	C-His	227.7
M3P4	ACAS (M1–P1317)	CTB1 (N1294–S2196)	pEM3P4-2	pET-24a(+)	C-His	243.7
M3P5	ACAS (M1–P1317)	wA (P1286–V2157)	pEM3P5-1	pET-24a(+)	C-His	241.3
M3P6	ACAS (M1–P1317)	Pks1 (P1290–E2183)	pEM3P6-3	pET-24a(+)	C-His	242.9
M4P1	CTB1 (M1–S1293)	PksA (E1305–A2109)	pEM4P1-7	pET-24a(+)	C-His	227.3
M4P2	CTB1 (M1–S1293)	Pks4 (A1285–N2036)	pEM4P2-3	pET-24a(+)	C-His	221.9
M4P3	CTB1 (M1–S1293)	ACAS (S1318–S1771)	pEM4P3-8	pET-24a(+)	C-His	189.6
M4P5	CTB1 (M1–S1293)	wA (P1286–V2157)	pEM4P5-1	pET-24a(+)	C-His	235.6
M4P6	CTB1 (M1–S1293)	Pks1 (P1290–E2183)	pEM4P6-1	pET-24a(+)	C-His	237.1
M5P1	wA (M1–Q1285)	PksA (E1305–A2109)	pEM5P1-1	pET-24a(+)	C-His	228.6
M5P2	wA (M1–Q1285)	Pks4 (A1285–N2036)	pEM5P2-1	pET-24a(+)	C-His	223.2
M5P3	wA (M1–Q1285)	ACAS (S1318–S1771)	pEM5P3-1	pET-24a(+)	C-His	190.9
M5P4	wA (M1–Q1285)	CTB1 (N1294–S2196)	pEM5P4-1	pET-24a(+)	C-His	239.3
M5P6	wA (M1–Q1285)	Pks1 (P1290–E2183)	pEM5P6-1	pET-24a(+)	C-His	238.4
M6P1	Pks1 (M1–R1289)	PksA (E1305–A2109)	pEM6P1-1	pET-24a(+)	C-His	228.3
M6P2	Pks1 (M1–R1289)	Pks4 (A1285–N2036)	pEM6P2-2	pET-24a(+)	C-His	222.9
M6P3	Pks1 (M1–R1289)	ACAS (S1318–S1771)	pEM6P3-1	pET-24a(+)	C-His	190.6
M6P4	Pks1 (M1–R1289)	CTB1 (N1294–S2196)	pEM6P4-2	pET-24a(+)	C-His	238.9
M6P5	Pks1 (M1–R1289)	wA (P1286–V2157)	pEM6P5-1	pET-24a(+)	C-His	236.6

Appendix Table E.7│ Details of potential topopyrone synthase expression constructs used in this study. PT or TE null control constructs for Pks1 are also included. The numbering for point mutations are derived from the parent sequence's numbering scheme and do not reflect the true position of mutation in the final chimeric sequence.

Protein	SKM[a]	Encodes P[b]	A_nT[c]	Plasmid	Vector	Tag	MW (kDa)
M2P1	Pks4	PksA	PksA	pEM2P1-3	pET-24a(+)	C-His	228.2
M2P1-GL	Pks4	PksA G1491L	PksA	pEM2P1-GL	pET-24a(+)	C-His	228.2
M2P1-GI	Pks4	PksA G1491I	PksA	pEM2P1-GI	pET-24a(+)	C-His	228.2
M2P1-GV	Pks4	PksA G1491V	PksA	pEM2P1-GV	pET-24a(+)	C-His	228.2
M2P1-GA	Pks4	PksA G1491A	PksA	pEM2P1-GA	pET-24a(+)	C-His	228.2
M2P1[P-T-]-GL	Pks4	PksA H1345A G1491L S2007A	PksA	pEM2P1[P-T-]-GL	pET-24a(+)	C-His	228.1
M2P1[P-]-GL	Pks4	PksA H1345A G1491L	PksA	pEM2P1[P-]-GL	pET-24a(+)	C-His	228.1
M2P4A1	Pks4	CTB1	PksA	pEM2P4A1	pET-24a(+)	C-His	224.8
M2P4A1-IV	Pks4	CTB1 I1468V	PksA	pEM2P4A1-IV	pET-24a(+)	C-His	224.8
M2P4A1-IA	Pks4	CTB1 I1468A	PksA	pEM2P4A1-IA	pET-24a(+)	C-His	224.7
M2P4A1-IG	Pks4	CTB1 I1468G	PksA	pEM2P4A1-IG	pET-24a(+)	C-His	224.7
Pks1	Pks1	Pks1	Pks1	pEPks1alt	pET-24a(+)	C-His	238.0
Pks1[P-T-]	Pks1	Pks1 H1330A S2009A	Pks1	pEPks1[P-T-]	pET-24a(+)	C-His	238.0
Pks1[P-]	Pks1	Pks1 H1330A	Pks1	pEPks1[P-]	pET-24a(+)	C-His	238.0

[a] SAT-KS-MAT (SKM) fraction encodes native Pks4 M1–R1284.
[b] PT (P) fraction encodes native PksA E1305–A1704 or CTB1 N1294–K1666.
[c] ACP_n-TE (A_nT) fraction encodes native PksA A1705–A2109.

Appendix Table E.8│Details of deconstructed NR-PKS constructs used in this study. Revised sequencing number is used for constructs with asterisks. References for the revised sequences are given.

Protein	Encodes	Plasmid	Vector	Tag	MW (kDa)
Pks4 SAT-KS-MAT	M1-A1283*[2]	pEPks4-NKAn	pET-24a(+)	C-His	140.8
PksA PT-G1491L	E1305-I1677	pEPT-G1491L	pET-24a(+)	C-His	42.4
CTB1 PT-I1468G	S1293-I1654	pECTB1-PT-I1468G	pET-24a(+)	C-His	40.4
PksA ACP	I1677-S1839	pEACP41	pET-28a(+)	N-His	20.6
Pks4 ACP	P1631–K1764*[2]	pEPks4-ACP	pET-28a(+)	N-His	16.2
CTB1 ACP$_2$	S1637–K1909	PECTB1-ACP	pET-28a(+)	N-His	31.7
PksA TE	D1812-A2109	pETE2	pET-28a(+)	N-His	36.4
Pks4 TE	P1765-N2036*[2]	pEPks4-TE	pET-28a(+)	N-His	32.2
wA TE	L1859-V2157	pEwA-TE	pET-28a(+)	N-His	34.9

E.2. *Chemical Characterization of Polyketide Products*
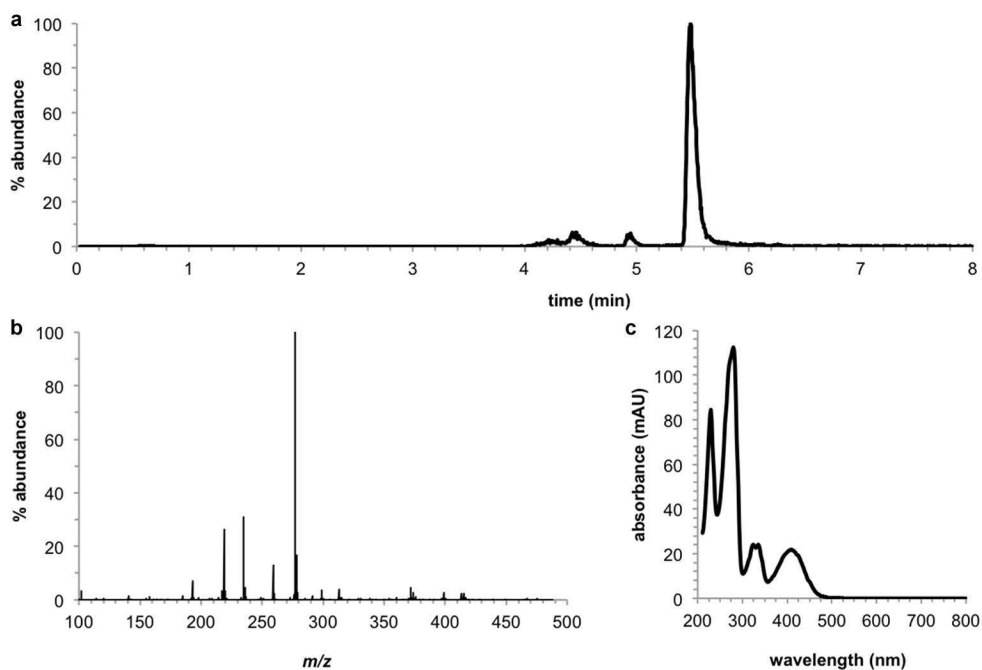
Appendix Figure E.2│Spectral characterization of YWA1. (a) Extracted ion chromatogram for *m/z* 277.0712. (b) MS for YWA1. (c) UV spectrum for YWA1. These data are consistent with literature values.[3]
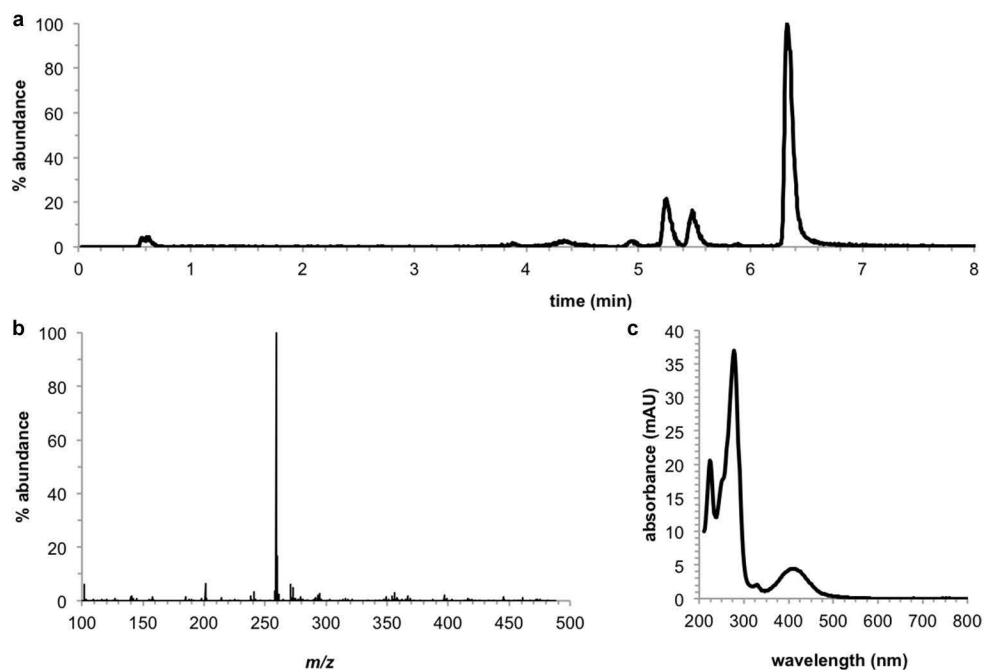
Appendix Figure E.3│Spectral characterization of *nor*-rubrofusarin. (a) Extracted ion chromatogram for *m/z* 259.0607. (b) MS for *nor*-rubrofusarin. (c) UV spectrum for *nor*-rubrofusarin. These data are consistent with literature values.[3d]
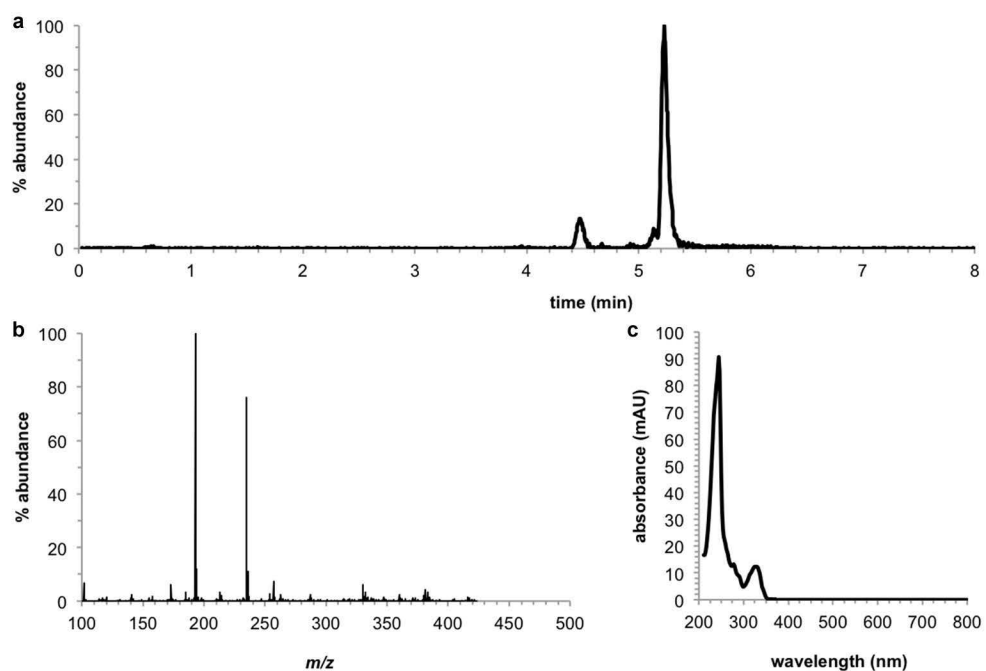
Appendix Figure E.4│Spectral characterization of isocoumarin 5. (a) Extracted ion chromatogram for *m/z* 235.0607. (b) MS for isocoumarin **5**. (c) UV spectrum for isocoumarin **5**. These data are consistent with literature values.[3d, 4]

231
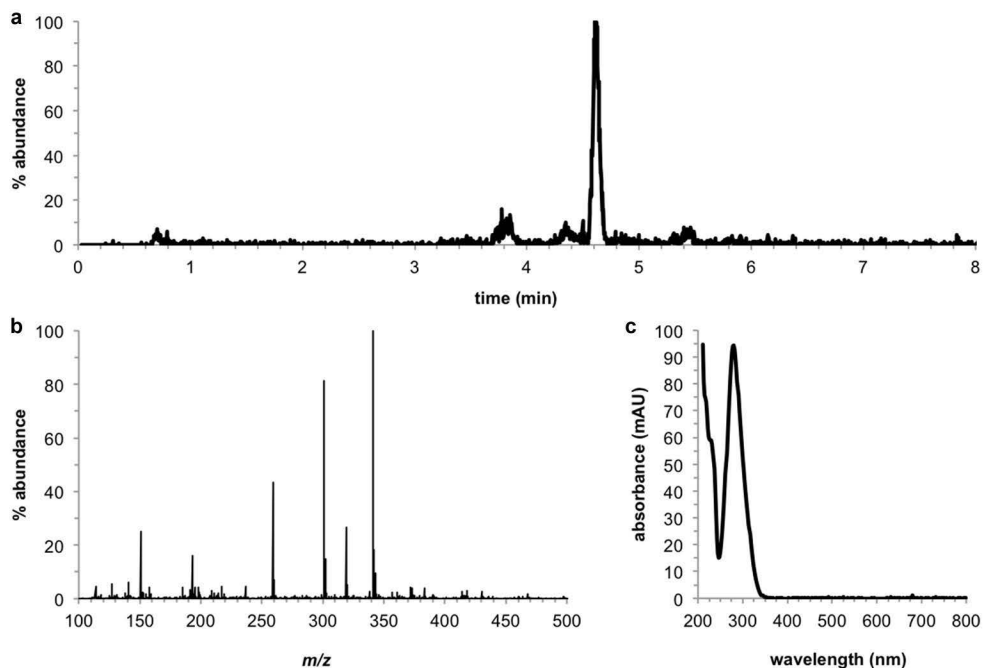
Appendix Figure E.5│Spectral characterization of derailment products SEK4 and SEK4b. (a) Extracted ion chromatogram for *m/z* 319.0818. (b) MS for products SEK4 and SEK4b. (c) UV spectrum for products SEK4 and SEK4b. These data are consistent with literature values.[3d]
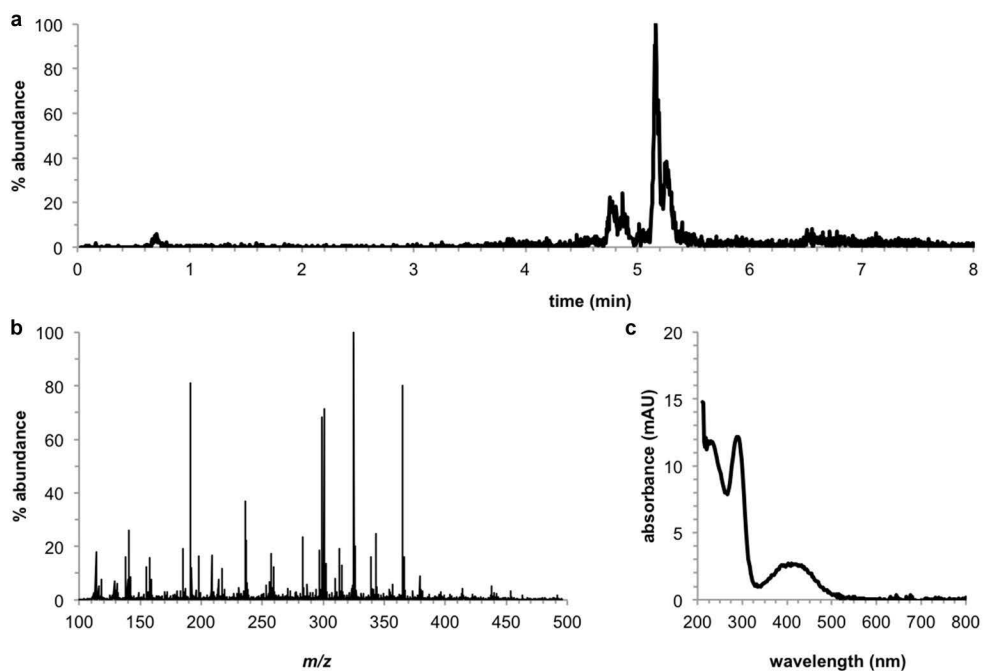
Appendix Figure E.6│Spectral characterization of topoketone. (a) Extracted ion chromatogram for *m/z* 325.0712. (b) MS for topoketone. (c) UV spectrum for topoketone.

E.3. Flaviolin Standard Curve

An authentic sample of flaviolin was prepared by Dr. Eric A. Hill (Johns Hopkins University, Baltimore, MD, 21218).[4] The area under the 280 nm chromatogram peak for flaviolin was used to generate a standard curve for determining flaviolin titers by producing strains (Figure E.7).
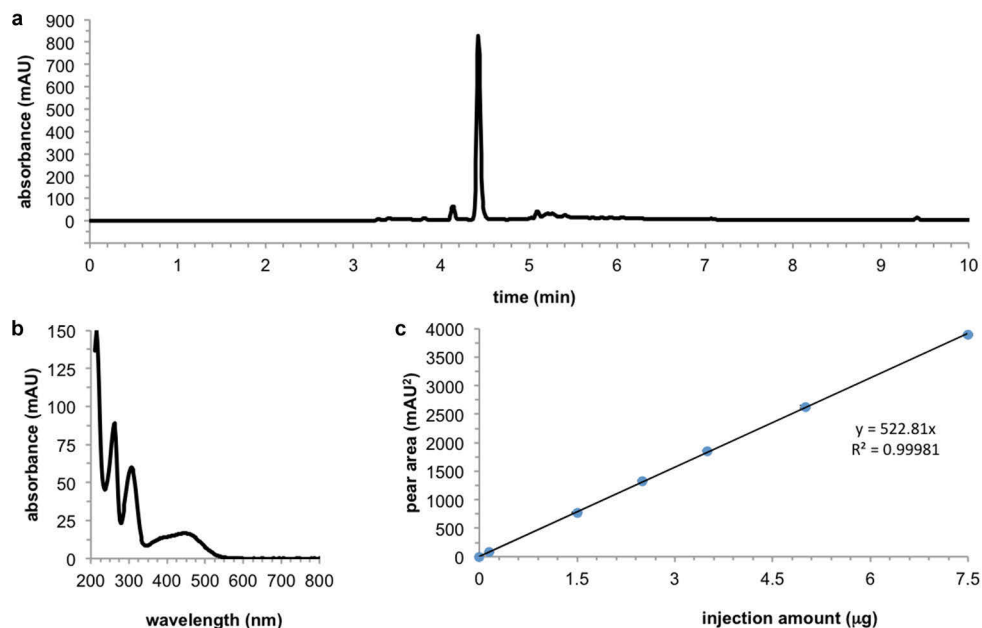
Appendix Figure E.7│Flaviolin standard curve. (a) HPLC 280 nm chromatogram of synthetic flaviolin. (b) UV spectrum of synthetic flaviolin. (c) Standard curve for flaviolin. Area under the flaviolin peak from 280 nm chromatograms are plotted against amount of flaviolin injected. Error bars are the standard deviation (n = 3).

E.4. References

[1]   X. Robert, P. Gouet, *Nucleic Acids Res* **2014**, *42*, W320-324.
[2]   J. M. Crawford, A. L. Vagstad, K. R. Whitworth, K. C. Ehrlich, C. A. Townsend, *Chembiochem* **2008**, *9*, 1019-1023.
[3]   (a) I. Fujii, A. Watanabe, U. Sankawa, Y. Ebizuka, *Chem Biol* **2001**, *8*, 189-197; (b) A. G. Newman, A. L. Vagstad, K. Belecki, J. R. Scheerer, C. A. Townsend, *Chem Commun* **2012**, *48*, 11772-11774; (c) A. L. Vagstad, A. G. Newman, P. A. Storm, K. Belecki, J. M. Crawford, C. A. Townsend, *Angew Chem Int Ed Engl* **2013**, *52*, 1718-1721; (d) A. G. Newman, A. L. Vagstad, P. A. Storm, C. A. Townsend, *J Am Chem Soc* **2014**, *136*, 7348-7362.

[4]   A. L. Vagstad, E. A. Hill, J. W. Labonte, C. A. Townsend, *Chem Biol* **2012**, *19*, 1525-1534.

Appendix F: Supplementary Material to Chapter 5

F.1.   Cloning CTB3 and CTB2

Expression constructs were prepared from CTB3 and CTB2. CTB3 was

deconstructed into monodomain fragments for ease of expression as detailed in Chapter

5. Specific sequence information relevant to cloning expression constructs for CTB3 and

CTB2 are detailed in this section.

F.1.1. Primer Sequences

Appendix Table F.1│Primers used to clone new constructs in this study. Underlined sequence aligns to adjacent exon. Bold sequences are introduced restriction enzyme cut sites. Italicized sequences are introduced stop codons.

Primer	Sequence
CTB3-5	ATGATGCAGTTCCAACGCGATCTTGAGGCG
CTB3-3	TCAAAACACAAACTCTGAACAATCCACCAG
CTB3-AseI-5	GCCG**ATTAAT**ATGATGCAGTTCCAACGCGAT
CTB3-NotI-3	ATTA**GCGGCCGC**AAACACAAACTCTGAACAATC
CTB3-Stop-3	ATTA**GCGGCCGC***TCA*AAACACAAACTCTGAACA
CTB3-ex1-3	ATGTTGATAGCCGTTTTGTAGATGCTCCAG
CTB3-ex2-5	CAAAACGGCTATCAACATTTAACCTGTTTT
CTB3-ex2-3	GTAGTTGACCATCCATCTAGCCCAACCTAA
CTB3-ex3-5	AGATGGATGGTCAACTACTCTGTGCCAGCT
CTB3-MT3-3	ATTA**GCGGCCGC**GGAGAGGGCAGTTTGGCCGGCTGA
CTB3-MT3-Stop-3	ATTA**GCGGCCGC***TTA*GGAGAGGGCAGTTTGGCCGGCTGA
CTB3-MO1-5	ATTA**CATATG**GATTCGAGGCCAGTCCTCATCGCA
CTB2-5	ATGGTTAAACGAATCGAAGCGGACAATCTC
CTB2-3	TTAATGACCATTGACACCATTGGTGCCATT
CTB2-NdeI-5	ATTA**CATATG**GTTAAACGAATCGAAGCG
CTB2-NotI-3	ATTA**GCGGCCGC**ATGACCATTGACACCATTGGT
CTB2-Stop-3	ATTA**GCGGCCGC***TT*AATGACCATTGACACCATT
CTB2-ex1-3	GAAGACGACCTGCATCGACGCCGTAGCGCA
CTB2-ex2-5	TCGATGCAGGTCGTCTTCAAATACAAGTTT

F.2.   Protein Expression Constructs

Proteins were expressed and purified as described in Chapter 5. Specific

characteristics of these proteins are presented in this section. We report here the corrected

CTB3 sequence. Five corrections were made in the CTB3 nucleotide sequence: A292G,

C2038A, G2075A, G2550A, and G2679A. These changes resulted in three in the CTB3

protein sequence: D635N, G793D, G836E. Large sections of the CTB2 nucleotide

sequence were corrected. These corrections led to a corrected CTB2 protein sequence

with a new numbering scheme. The corrected CTB2 sequence showed greater alignment

with other *O*-methyltransferases as determined by BLAST.

Edited CTB3 Nucleotide Sequence from *C. nicotianae*
Changes in sequence are in boldface. Changes resulting in changes to
the protein sequence are underlined.
ATGATGCAGTTCCAACGCGATCTTGAGGCGTCCTTGGAGGCCGTATCGGCCAACGCCCAGGAGCTGCTCAA
ATCTCTCAAGAGTCGCAAGGATGTTCAAGACCTCAACGCGTCGTTGCCGAAGGATCCTTTAGACAACTGCG
ATGCTCAAACTCAAGCCGCTCGTGCGCAGCTGGCAGAGGCAGCGACAAGAATCTTGCAGTTGTCGATCCGA
CCTCAAGAGTATCTGGAGCATCTACAAAACGGCTATCAACATTTAACCTGTTTTCGCTGGCTGGTGGAACT
CAACATATTGGACCACCTTCCACATAGCGGAACGATCAGCTACACAGATCTTGCGAGAAAAGCCAGCGTGC
CGCCTATGCAATTGAGAAGCATCTGTCGCATGGCCATATGCAATGGATTCCTGGAAGAGCCCGAGGCCAAC
CAAGTCCGCCACAGTCGCATTTCCGCCTTGTTCGCTCGCGATGAAAGCTATTTAGGTTGGGCTAGATGGAT
GGTCAACTACTCTGTGCCAGCTGCATACAAGCTTAGCGACGCCACGCGATCGTGGGGCGAGACTGTCGCCA
AAGATCAGACCGCGTTCAATCTGGGAATGGATGTGAAAGTCCCATTCTTTGACCATCTCCGCCAGACGCCC
GCAATGAAGGACGCCTTTGCAGCTTATATGCGTAATGTGACTTCGAACGCAACTTGGGGCCTCCAGCACGC
AGTCACCGGCTTCGACTGGGCTTCCCTTCCGCGGGGCGCAAAAGTCGTGGATGTCGGTGGCTCTCTTGGGC
ATGGTAGCATTGCCATTGCCAAGGAGCACACTCACCTTACCTTCGTCATTCAAGATCTGCCAGAGACGGTC
GCTGGTGCCAGGAAAGAAATGGCCCAAAATGACAAGATTGAAGCTTCTGTTAAATCTCGCATCACCTTTCA
GGAACATGACTTCTTTGGTCCTCAAACAGTGAAGGATGCCGATGTTTACTTTCTTCGCATGATCTGTCACG
ACTGGCCCGACAACGAAGCCAAGGTCATCCTCTCTCAGATTCGCGCTGCACTGAAACCTGGGGCGCAAATA
GTCATCATGGACACCATTCTTCCCCAGCCCGGCACAATTAGCGTTTTGCAAGAGCAACAACTACGCATTCG
GGATCTAACAATGATGGAAGTCTTCAATGCCAAGGAGCGTGAATTGGAGGACTGGAGCTCATTGATGCAAT
CTGCCGGTCTCGAGATTTCTCGCGTGAACCAGCCGCTCAACAGTGTGATGGGTCTGCTCACAGTCCGCTCA
GCCGGCCAAACTGCCCTCTCCGGAACGAATACACTGACGCCAGAGTTGGTGGCGGCAGTCTCCGCAAGCAC
TGGCTCTGCTGATTCGAGGCCAGTCCTCATCGCAGGCGCGGGTATTGCTGGGCTCTGCCTTGCACAGGCTT
TGAAGAAGGCCGGCATTGACTTTCGCGTCTTCGAAAGGGACTTCCATGTCGATGCTCGGCCACAAGGATAC
CGACTCAAATTCGAAGCAGACGCCGCACAGTCTCTCAAGAACATCCTGCCTGACGATGTTTATGAGGCTTT
CGAACTGTCAAATGCCGTCACCGCCGTAGGCGAGACGGACTTCAATCCCTTCAATGGTAACATAATCCACA
GCCGCACTGGTGGCGGCCTGTCTGGCAAGAAGGGACTGTATGCGACATTCACTGTTGACCGCAAAGCATTC
AGAACTCAGCTCATGACTGGCATTGAGGACAAGATTTCGTTCGGGAAGGAAATCGCGTACTACAAGACCGA
TGACGCTACATCTACGGTCAACGCAGAATTCAAGGACGGCACTCACGTCACCGGAAGTTTCCTGGCCGGCA
CTGATGGCTTACACTCTGTAGTTCGCAAGACATGTGTACCAAACCATCGTATCGTG**A**ATACTGGTGCTGCC
TGCATCTACGGCAAGACTGTAATGACACCGGAATTCCTCGCGCGATTCCCCGAGAAAGGCTTGAGATTCAT
GACAGTGGTCAGCGACATCGCACCTATGCTACAGTCTTGTCTCATCGGCGACAGCCCAGTCACCTTACTAC
TGGAGCCTATCCGATTCAGCGAAGCCTCGCGTGCCCGCTATCCAGAACTGCCTCCAGACTACGTCTACTGG
GCCCTCATCGGACCCAAGGAACGTTTCGGATCGCAAGAGGTGACTTCCATGAAGAACTTCGTCTCACTGGA
CCAAGCGGCAGAACAGGCTGCCAAGCTCAGTCTCGCAGTCACCGAGGAATGGCATCCGAGCCTTCGCGCGC
TGTTTGAGCTTCAAGACACGAAGCAAGCATCGCTCATTCGCGTTGCATCCACAATTCCCGATATCCCATCT
TGGGAGTCCCACTCCAATGTTACCGTTCTTGGCG**A**TAGCATTCATCCGATGAGCCCTTGTGGTGGAGTCGG
AGCGAACACCGCAATAGTCGATGCCGACGCCTTGG**C**TAAAGTGCTCGTTGAGCATGGTACGAAGCCACCGG
TGAACGCAATCGCCGAGTTCG**A**GGCCGCGATGAGAACGAGAGCGAAGAGGAACATTTGGAGGAGTGAGGTT
GGCAGCAAGAGGATGTTTGGACAGAAGAATCTGGTGGATTGTTCAGAGTTTGTGTTTTGA

Edited CTB3 Protein Sequence from *C. nicotianae*
Changes in protein sequence are in boldface and underlined.
MMQFQRDLEASLEAVSANAQELLKSLKSRKDVQDLNASLPKDPLDNCDAQTQAARAQLAEAATRILQLSIR
PQEYLEHLQNGYQHLTCFRWLVELNILDHLPHSGTISYTDLARKASVPPMQLRSICRMAICNGFLEEPEAN

236

```
QVRHSRISALFARDESYLGWARWMVNYSVPAAYKLSDATRSWGETVAKDQTAFNLGMDVKVPFFDHLRQTP
AMKDAFAAYMRNVTSNATWGLQHAVTGFDWASLPRGAKVVDVGGSLGHGSIAIAKEHTHLTFVIQDLPETV
AGARKEMAQNDKIEASVKSRITFQEHDFFGPQTVKDADVYFLRMICHDWPDNEAKVILSQIRAALKPGAQI
VIMDTILPQPGTISVLQEQQLRIRDLTMMEVFNAKERELEDWSSLMQSAGLEISRVNQPLNSVMGLLTVRS
AGQTALSGTNTLTPELVAAVSASTGSADSRPVLIAGAGIAGLCLAQALKKAGIDFRVFERDFHVDARPQGY
RLKFEADAAQSLKNILPDDVYEAFELSNAVTAVGETDFNPFNGNIIHSRTGGGLSGKKGLYATFTVDRKAF
RTQLMTGIEDKISFGKEIAYYKTDDATSTVNAEFKDGTHVTGSFLAGTDGLHSVVRKTCVPNHRIVNTGAA
CIYGKTVMTPEFLARFPEKGLRFMTVVSDIAPMLQSCLIGDSPVTLLLEPIRFSEASRARYPELPPDYVYW
ALIGPKERFGSQEVTSMKNFVSLDQAAEQAAKLSLAVTEEWHPSLRALFELQDTKQASLIRVASTIPDIPS
WESHSNVTVLGDSIHPMSPCGGVGANTAIVDADALAKVLVEHGTKPPVNAIAEFAAMRTRAKRNIWRSEV
GSKRMFGQKNLVDCSEFVF
```

Edited CTB2 Nucleotide Sequence from *C. nicotianae*
Changes in sequence are in boldface. Changes resulting in changes to
the protein sequence are underlined. Insertions are italicized.
```
ATGGTTAAACGAATCGAAGCGGACAATCTCTTTGAGCTCACGGCAGAGCTGGTCTCAGCCTCCTCAAAGCT
TCACAAATTTCTAGACCAGAAGAACCTTCCGCAGCCATCGTTTGATGCTCCAGCTCCATCGGTTGCTCTCA
ACTCAGCCAACAAGCCGTACTATGATGCGAGAAGCGCGATTGTAGAGGCTGCTGAGCAACTCATTCGCCTG
GTCCGTGGACCTCGCGACACACTCCTCGCTCTGTCTTTCGAGCACTGCGCTACGGCGTCGATGCAGGTCGT
CTTCAAATACAAGTTTGCGAATCACATCCCGCTACATGGCTCGACAACTTATAGCAAGATTGCCGAAGCAG
TCGGAGATGGTGTGACAACAGCTCTCGTTGAGCGCACGATACAACATTGCGCTTCCTTTGGCCTATTCGAG
ACGATCCCTGGCG*G*CTATGTTACTC_A_CAATGCTACCTCGTC_A_CTACTGGTCACCGATCCAGATCTCGAAGC
CTGGATGTATCTCTCGGCGGTGATAGCCTACCCAGCTGGCGCAGCTATCCCTAAGGCTGTAGAGCAGTATG
GCGTTTCCCATGAAGCTGACGAATCAGGGTACGGTGCCAGCATAGGAAGAAAGATTGCACAATTCCAGCGA
TTCCGTGAGCCCGATGGGAAGAAGGACCACGAGATGTTCGCACGAGCCATGCGTGGTATCGCGGCTGGTGG
TGCGTATGACTTCCGCCATGCAGTCGATGGCGGATACCCTTGGCACCTCCTCGCAGAGGGCGCAGGTCACC
TGGTCGTGGATGTTGGTGGAGGTCCCGGCCACGTCGCCATGGCACTCGCGGAAAAGTACCCAAGCTTGCGT
TTCCAAGTCCAAGATCTGCCCGAGACCGTCCAAGTGGGAGCGAAGAATTGCCCTGAGCACTTGAAGTCCCG
CGTGTCATTCCAGAGCCACGATTTCTTCACCTCGCAACCTGCTCATGAAGTGCAAGACGGCGAAGGAATCG
TCTACTTTGCTCGATTCATCTTGCACGACTGGAGTGACAAGTATGCCACCAAGATCGTGCAGCAACTTGCA
ACTGGCTTGAGGCCACAAGATCGCATCATTTTGAACGAAGTGGTCGTTCCCGAAGCCGGACAAGTTGGCAG
AGAGACCGAACGAAGAATGCACGATCGTGATCTGCTGATGTTGATGAACCTCAACGGCCGTGAGCGGACAC
AGAGTGCATTCGAGGCGATCTTCGCTTCAGTGACTCCCAAGCTGCGGCTGCAGAGGGTCATTCACCCAGAG
CAAGGCGAATTGTCGCTTATCGAGGTGACTCTTGATGGCGTTGAGCTTCCTGCCCAGGCTAATGGCGTCAA
CGGCCATGCGAATGGCACCAATGGTGTCAATGGTCA_G_TAA
```

Edited CTB2 Nucleotide Sequence from *C. nicotianae*
Changes in sequence are in boldface and underlined.
```
MVKRIEADNLFELTAELVSASSKLHKFLDQKNLPQPSFDAPAPSVALNSANKPYYDARSAIVEAAEQLIRL
VRGPRDTLLALSFEHCATASMQVVFKYKFANHIPLHGSTTYSKIAEAVGDGVTTALVERTIQHCASFGLFE
TIPGGYVTHNATSSLLVTDPDLEAWMYLSAVIAYPAGAAIPKAVEQYGVSHEADESGYGASIGRKIAQFQR
FREPDGKKDHEMFARAMRGIAAGGAYDFRHAVDGGYPWHLLAEGAGHLVVDVGGGPGHVAMALAEKYPSLR
FQVQDLPETVQVGAKNCPEHLKSRVSFQSHDFFTSQPAHEVQDGEGIVYFARFILHDWSDKYATKIVQQLA
TGLRPQDRIILNEVVVPEAGQVGRETERRMHDRDLLMLMNLNGRERTQSAFEAIFASVTPKLRLQRVIHPE
QGELSLIEVTLDGVELPAQANGVNGHANGTNGVNGQ
```

Appendix Table F.2 │ Details of expression constructs used in this study.

Protein	Encodes	Plasmid	Vector	Tag	MW (kDa)
CTB3	M1–F871[a]	p28CTB3-3	pET-28a(+)	N-His	98.2
CTB3-MT	M1–S433	p28CTB3-MT3-3	pET-28a(+)	N-His	50.7
CTB3-MO	D454–F871[a]	pECTB3-MO1-3	pET-24a(+)	C-His	47.0
CTB2	M1–Q462[b]	pECTB2-1	pET-24a(+)	C-His	51.8

[a] New CTB3 protein sequence as described above.
[b] New CTB2 protein sequence as described above.
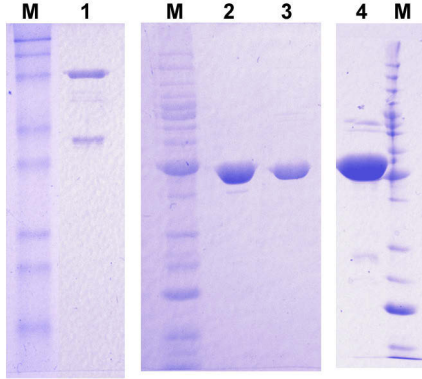
Appendix Figure F.1│12% SDS-PAGE analysis of proteins used in this study. Lane M: Ladder; 1: CTB3; 2: CTB3-MT; 3: CTB3-MO; 4: CTB2.
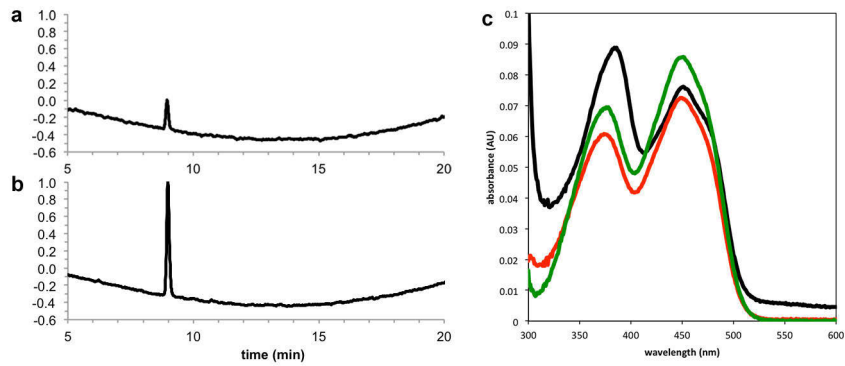
Appendix Figure F.2│Analysis of FAD content of CTB3-MO. (a) HPLC analysis of FAD released from heat denatured CTB3-MO. (b) HPLC analysis of FAD standard. (c) UV-vis spectra of CTB3-MO (black), FAD released from heat denatured CTB3-MO (red), and free FAD (green).

F.3. Cercosporin Complementation Assays
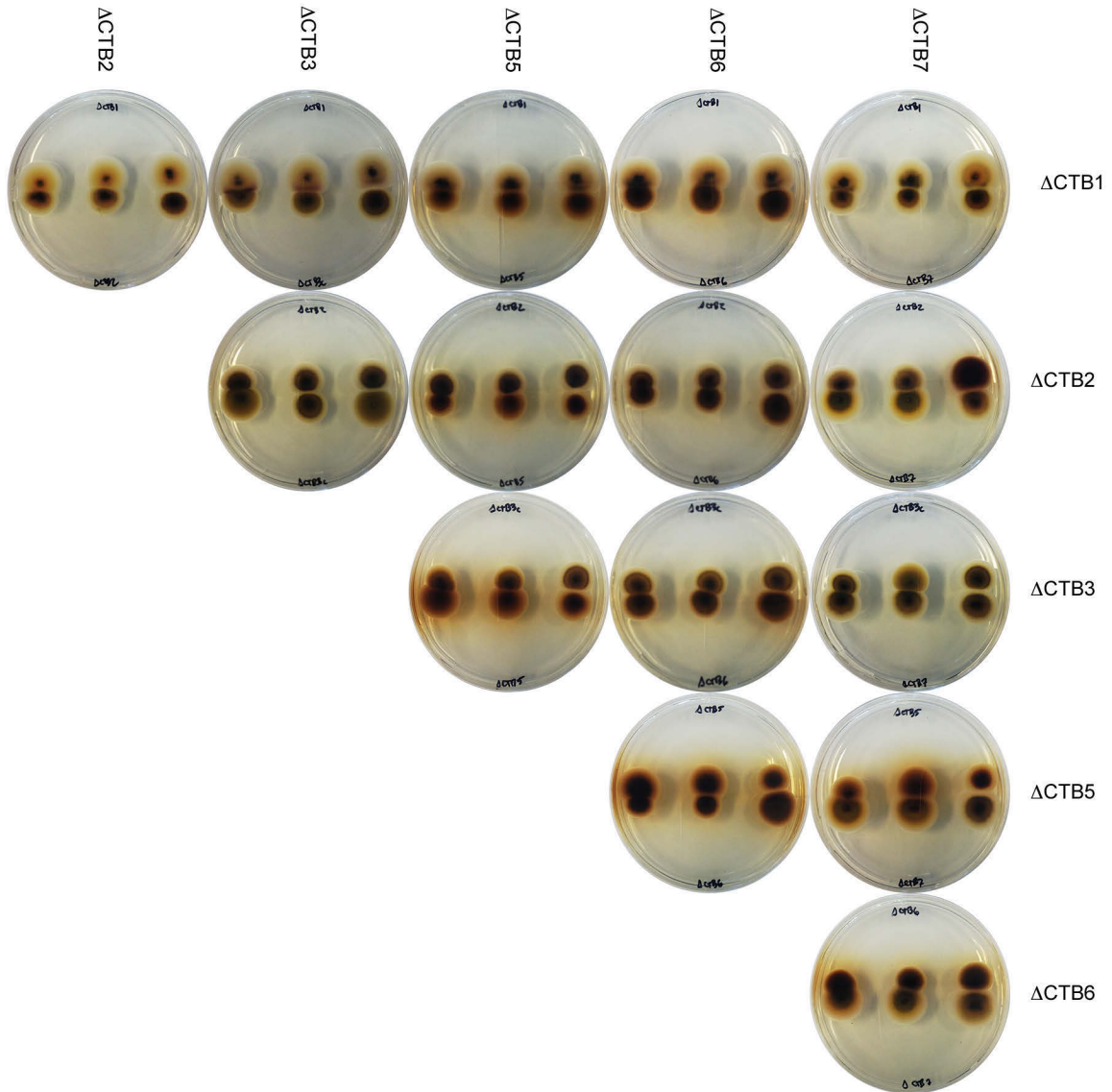
Appendix Figure F.3 | Pairwise complementation assay of *C. nicotianae* cercosporin biosynthetic gene knockout strains. Labels for each row indicate the identity of the top colony on each plate. Labels for each column indicate the identity of the bottom colony on each plate. Cercosporin complementation is indicated by the presence of a red pigment at the colony–colony interface. Colonies are spaced 0.5, 0.7, and 1.0 cm apart, from left to right.

F.4.   In Vitro Analysis of CTB2 Activity
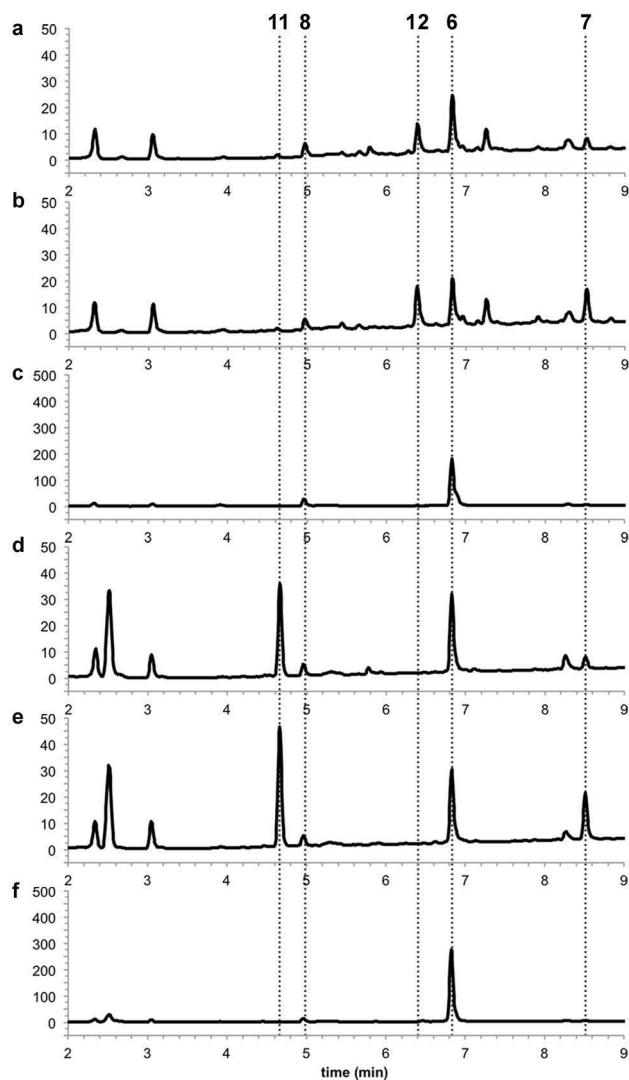
Appendix Figure F.4│Product profiles of *in vitro* reactions of CTB3 and CTB2. The 280 nm chromatograms of the following reactions are displayed: (a) CTB3-MT, CTB3-MO, CTB2 with *nor*-toralactone, (b) CTB3-MT, CTB3-MO with *nor*-toralactone, (c) CTB2 with *nor*-toralactone, (d) CTB3-MT, CTB3-MO, CTB2 with *nor*-toralactone under reductive conditions, (e) CTB3-MT, CTB3-MO with *nor*-toralactone under reductive conditions, (f) CTB2 with *nor*-toralactone under reductive conditions.

F.5. Chemical Characterization of Cercosporin Metabolites

F.5.1. ^1H NMR
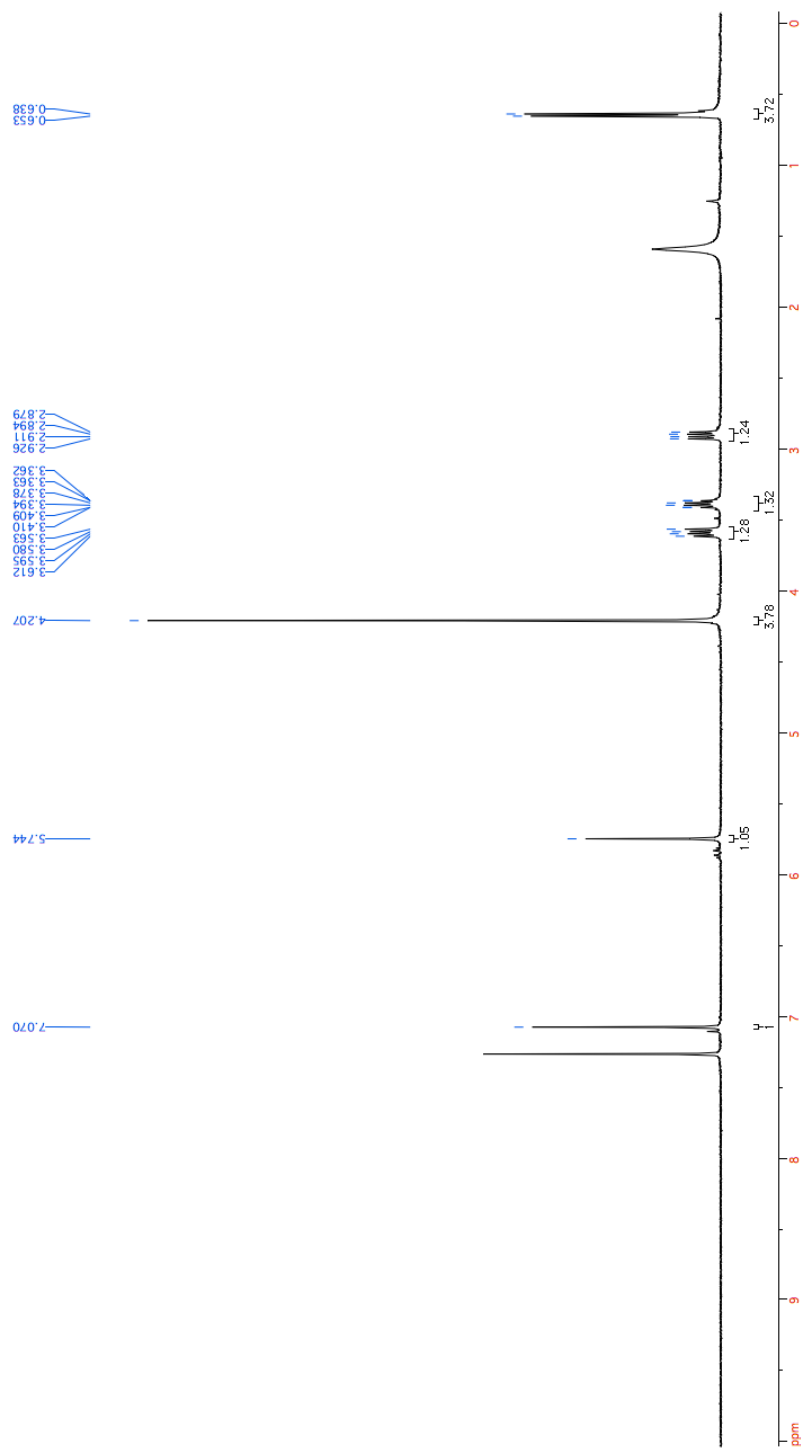
Appendix Figure F.5│^{1}H NMR (400 MHz, CCl$_3$D) of cercosporin (1) purified from *C. nicotianae* (ATCC$^©$ 18366$^{™}$).
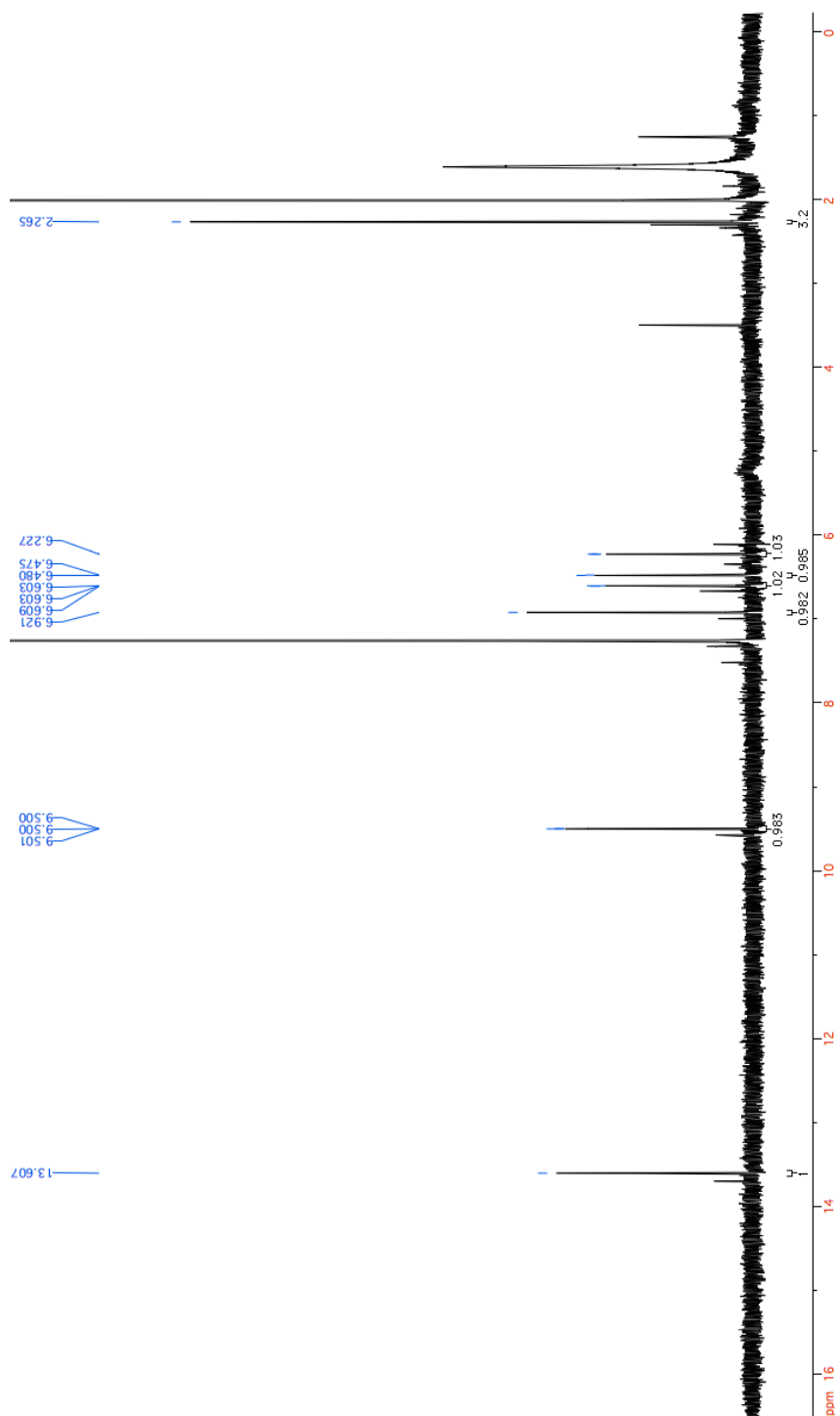
Appendix Figure F.6│¹H NMR (400 MHz, CCl₃D) of *nor*-toralactone (6) purified from *C. nicotianae* ΔCTB3c.
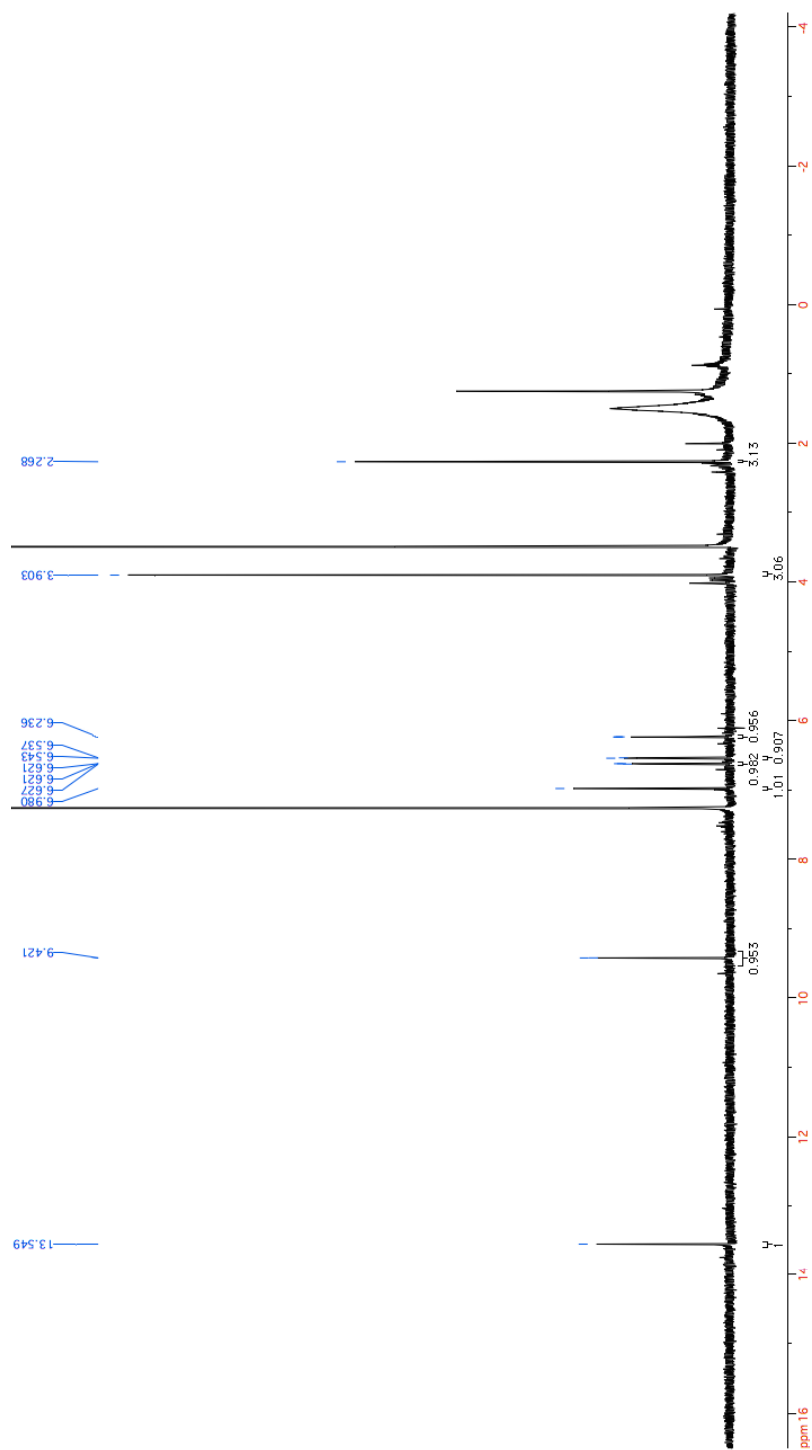
Appendix Figure F.7 │ ^{1}H NMR (400 MHz, CCl$_3$D) of toralactone (7) purified from *C. nicotianae* ΔCTB3c.
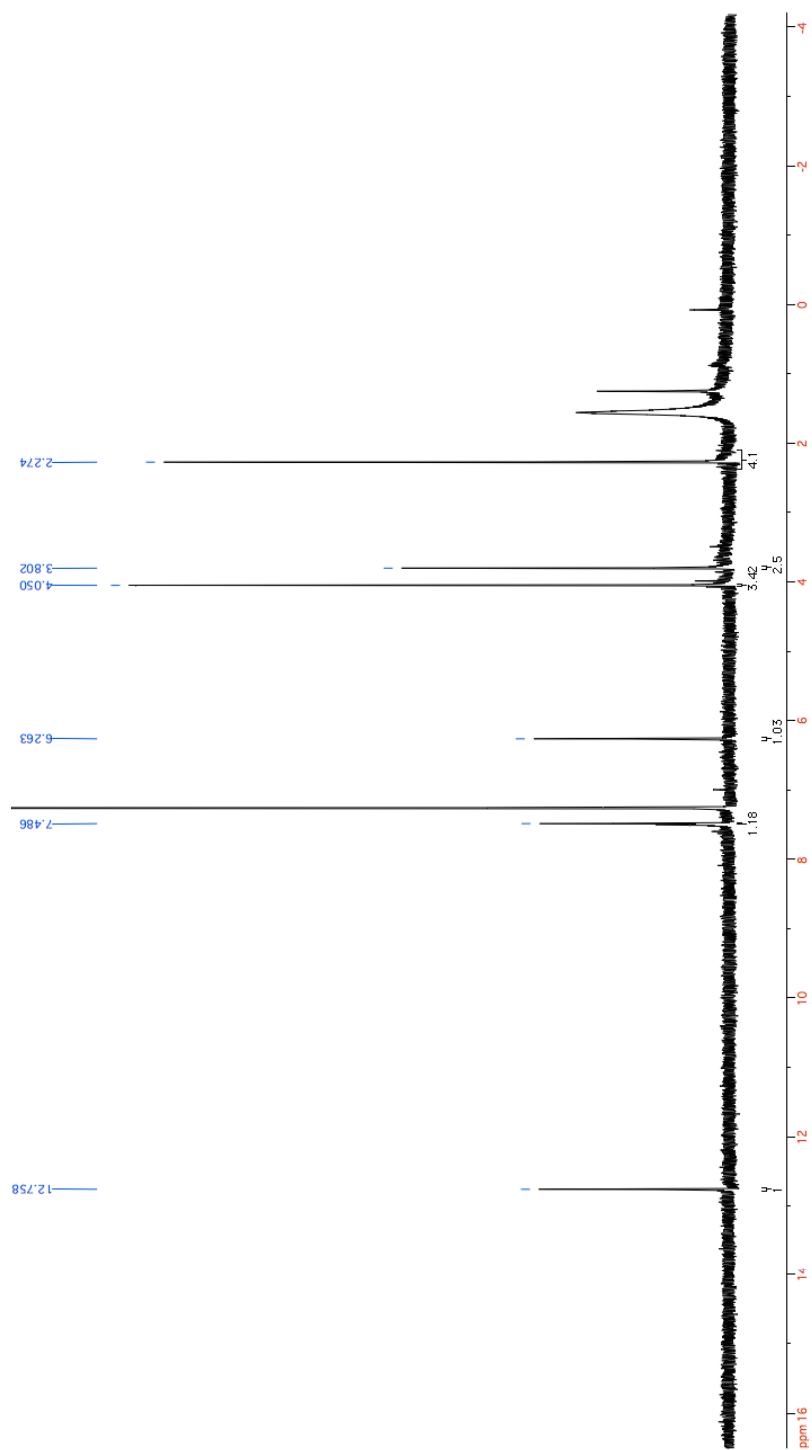
Appendix Figure F.8│¹H NMR (400 MHz, CCl₃D) of cercoquione A (9) purified from *C. nicotianae* ΔCTB6.
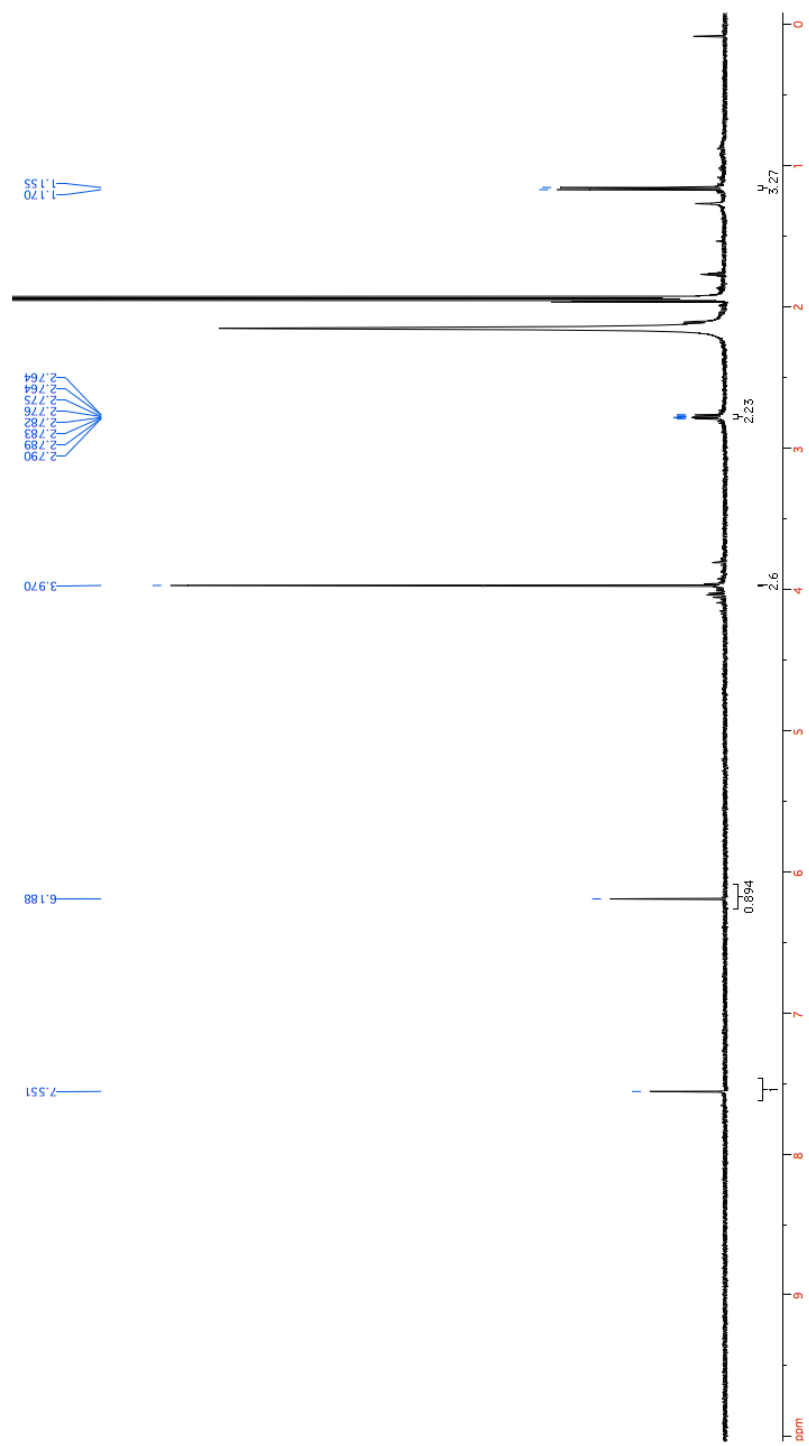
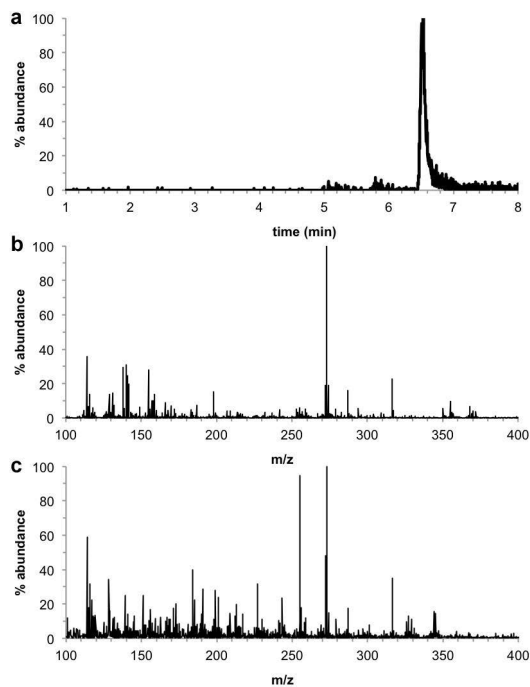**Appendix Figure F.9** | **^{1}H NMR (400 MHz, CD$_3$CN) of cercoquinone B (10) purified from *C. nicotianae* ΔCTB5.**
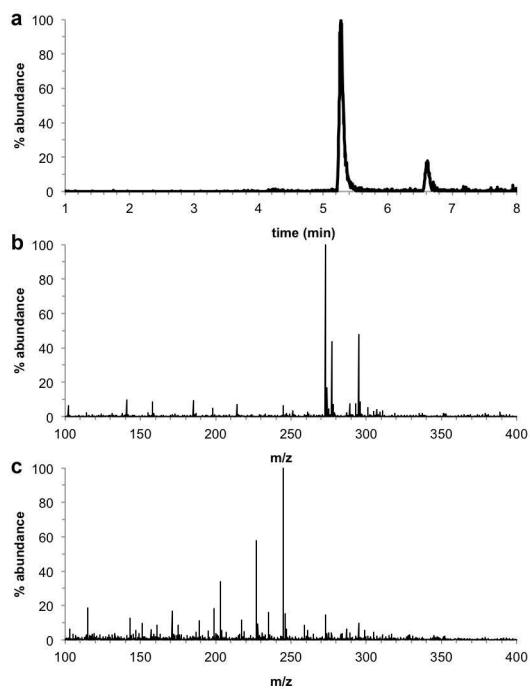
F.5.2. HRMS
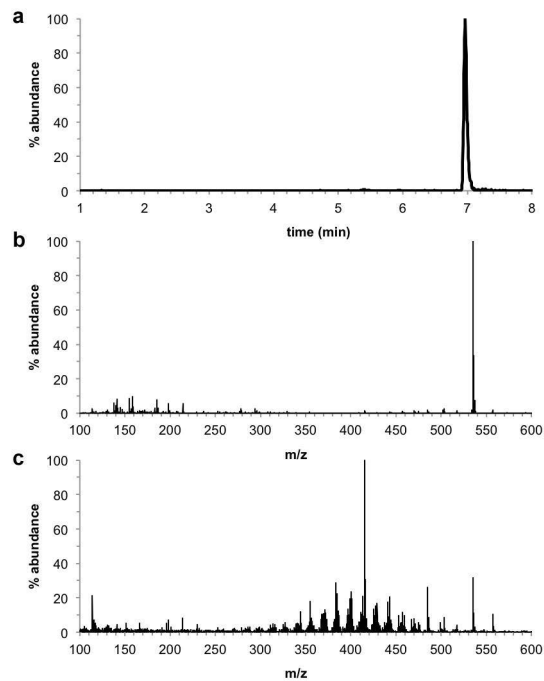
Appendix Figure F.10│HRMS spectra for cercosporin (1). (a) Extracted ion chromatogram for *m/z* 535.1604, and (b) MS, (c) MS/MS for cercosporin.
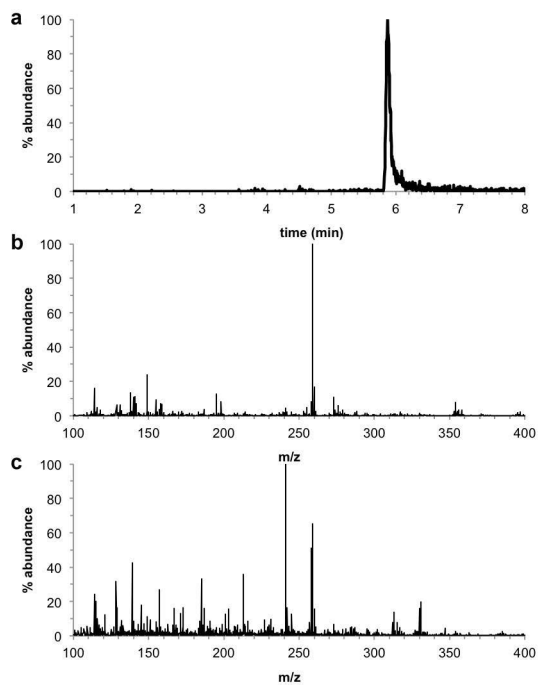
Appendix Figure F.11│HRMS spectra for *nor*-toralactone (6). (a) Extracted ion chromatogram for *m/z* 259.0607, and (b) MS, (c) MS/MS for *nor*-toralactone.
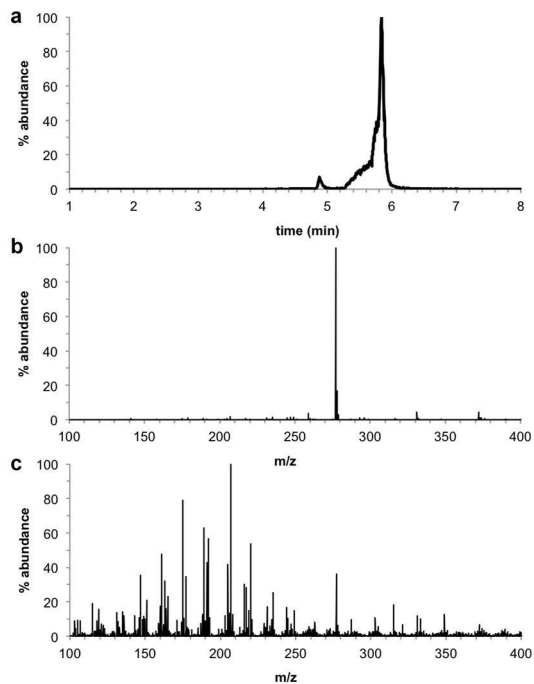
Appendix Figure F.12│HRMS spectra for toralactone (7). (a) Extracted ion chromatogram for *m/z* 273.0763, and (b) MS, (c) MS/MS for toralactone.
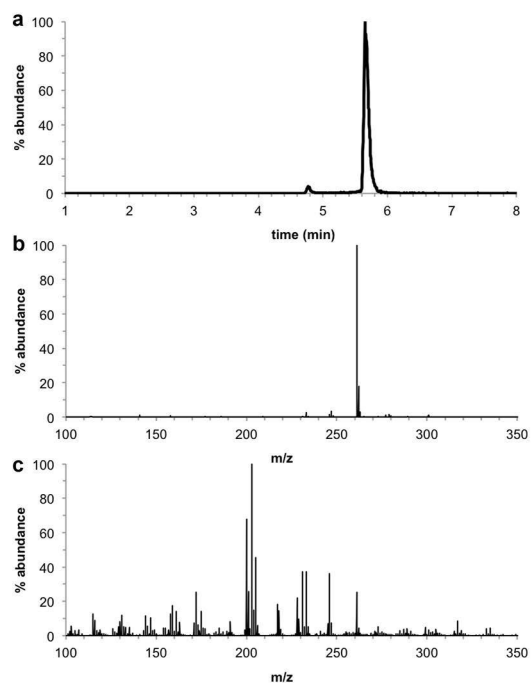

Appendix Figure F.13│HRMS spectra for compound 8. (a) Extracted ion chromatogram for *m/z* 273.0399, and (b) MS, (c) MS/MS for compound 8.
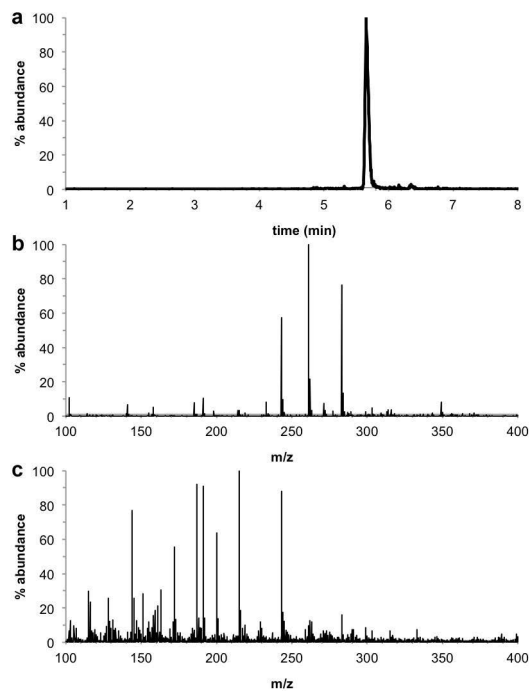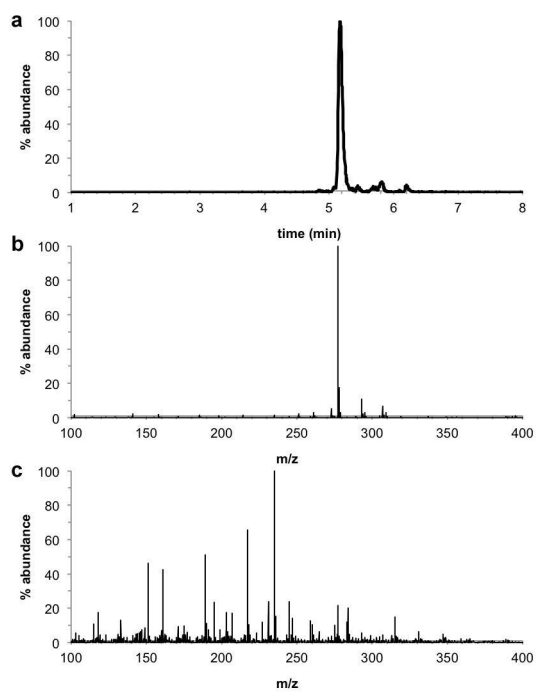
Appendix Figure F.14│HRMS spectra for cercoquinone A (9). (a) Extracted ion chromatogram for *m/z* 277.0712, and (b) MS, (c) MS/MS for cercoquinone A.

Appendix Figure F.15│HRMS spectra for cercoquinone B (10). (a) Extracted ion chromatogram for *m/z* 261.0763, and (b) MS, (c) MS/MS for cercoquinone B.
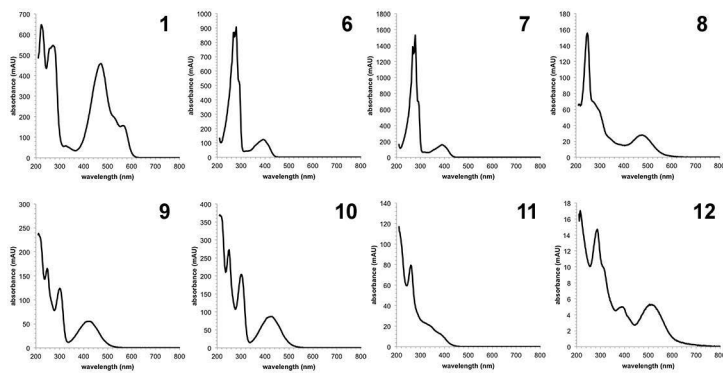
Appendix Figure F.16│HRMS spectra for cercoquinone C (12). (a) Extracted ion chromatogram for *m/z* 261.0763, and (b) MS, (c) MS/MS for cercoquinone C.

Appendix Figure F.17│HRMS spectra for cercoquinone D (11). (a) Extracted ion chromatogram for *m/z* 277.0712, and (b) MS, (c) MS/MS for cercoquinone D.

F.5.3. UV

Appendix Figure F.18│UV-visible spectra for cercosporin metabolites. The spectra for cercosporin (**1**),[1] *nor*-toralactone (**6**),[2] toralactone (**7**)[3] match those in the literature for these molecules.

F.6. References

[1]   S. Yamazaki, T. Ogawa, *Agric Biol Chem* **1972**, *36*, 1707-&.
[2]   A. G. Newman, A. L. Vagstad, K. Belecki, J. R. Scheerer, C. A. Townsend, *Chem Commun* **2012**, *48*, 11772-11774.
[3]   Z. J. Zhang, B. Yu, *J Org Chem* **2003**, *68*, 6309-6313.

Curriculum Vitae

ADAM G. NEWMAN

2834 N. Calvert Street
Apt. 1R
Baltimore, MD 21218

401-575-3389

agnewman85@gmail.com

adam.g.newman

linkedin.com/in/adamgnewman1

PROFILE

Natural product chemist specializing in synthetic biology, pathway characterization, and natural product biosynthesis now seeks role in the biotechnology sector. Over 7 years research experience characterizing and engineering fungal natural product pathways. Ph.D. in Chemical Biology Fall 2015.

EDUCATION

Johns Hopkins University — Baltimore, MD

Ph.D., Chemical Biology — *2015*
M.S., Chemical Biology — *2009*

Colby College — Waterville, ME

B.A., Chemistry, Conc. in Cell and Molec. Biol./Biochem. — *2007*
Honors and Distinction in Major, *summa cum laude*
Honors Thesis: *Quantitative PCR Reveals Nuclear DNA Alkylation by Epichlorohydrin in the Chicken Genome*
Departmental Biochemistry Award, ACS Biochemistry Award

RESEARCH

Johns Hopkins University — Baltimore, MD
Graduate Student with Dr. Craig A. Townsend — *2008-present*
Elucidation and engineering of fungal natural product pathways (polyketides).
In vitro biochemical characterization of polyketide synthases (PKSs).
Synthetic biology of PKSs for *in vitro* and heterologous production of non-native natural products.
In vitro and *in vivo* characterization of cercosporin biosynthetic pathway from fungal plant pathogen *Cercospora nicotianae*.

Colby College — Waterville, ME
Research Assistant with Dr. Julie T. Millard — *2006-2007*
Examination of *in vitro* and *in vivo* DNA interstrand crosslinking by epihalohydrins.
Determined crosslinking DNA sequence specificity.
Determined cytotoxicity of epihalohydrins using cell-based assays.
Measured *in vivo* differential chromatin crosslinking by epihalohydrins.

Ceimic Corporation — Narragansett, RI
Intern — *2005*
Isolation of organic components from environmental samples (soil, water, air).

SPECIAL SKILLS

Natural product chemistry	Synthetic biology
Biochemistry	Bioorganic chemistry
Biosynthetic chemistry	Molecular biology
Natural product preparation	Pathway characterization
Protein/Pathway engineering	Enzymology
Protein expression/purification	Metabolic analysis
Cell culture (bacteria, fungi)	Cloning (traditional, advanced)
PCR, RT-PCR, qPCR	Analytical chemistry
LC-MS, GC-MS, MS/MS	FTIR, UV, Fluorescence, CD
NMR	HPLC, UPLC (Agilent, Waters)
HPLC repair/maintenance	Chromatography (IC, AC, SEC)
Computer proficiency	Communication (oral, written)

LEADERSHIP

Johns Hopkins University — Baltimore, MD

Research Mentor, Townsend Laboratory — *2009-present*
Trained 1st year graduate students (4 total) in research practices during 8–12 week research rotations.
Guided students in completion of short-term research projects.

Co-organizer and Host, CBI Graduate Program 10th Annual Retreat — *2014*
Schedule faculty and student speakers.
Organized student and postdoctoral poster session.

Teaching Assistant, Chemical Biology — *2008-2009*

Colby College — Waterville, ME

Teaching Assistant, General and Organic Chemistry — *2004-2007*

Tutor, General Chemistry — *2006-2007*

PUBLICATIONS

A.G. Newman, C.A. Townsend (2015) Molecular characterization of the cercosporin biosynthetic pathway in the plant pathogen *Cercospora nicotianae*, *Manuscript in Preparation.*

J.F. Barajas, K. Finzel, T.R. Valentic, G. Shakya, N. Gamarra, D. Martinez, J.L. Meier, A.L. Vagstad, **A.G. Newman**, C.A Townsend, M.D. Burkart, S.C. Tsai (2015) Structural and biochemical analysis of protein-protein interactions between the acyl carrier protein and the product template domains. *Under Review.*

A.G. Newman, A.L. Vagstad, P.A. Storm, C.A. Townsend. (2014) Systematic domain swaps of iterative, non-reducing polyketide synthases provide a mechanistic understanding and rationale for catalytic reprogramming*, J. Am. Chem. Soc. 136,* 7348-7362.

A.L. Vagstad, **A.G. Newman**, P.A. Storm, K. Belecki, J.M. Crawford, and C.A. Townsend. (2013) Combinatorial domain swaps provide insights into the rules of fungal polyketide synthase programming and the rational synthesis of non-native aromatic products, *Angew. Chem. Int. Ed. Engl. 52*, 1718-1721.

A.G. Newman, A.L. Vagstad, K. Belecki, J.R. Scheerer, and C.A. Townsend. (2012) Analysis of the cercosporin polyketide synthase CTB1 reveals a new fungal thioesterase function, *Chem. Commun. 48*, 11772-11774.

T.P. Korman, J.M. Crawford, J.W. Labonte, **A.G. Newman**, J. Wong, C.A. Townsend, and S.C. Tsai. (2010) Structure and function of an iterative polyketide synthase thioesterase domain catalyzing Claisen cyclization in aflatoxin biosynthesis, *Proc. Natl. Acad. Sci. U. S. A. 107*, 6246-6251.

F.J. LaRiviere, **A.G. Newman**, M.L. Watts, S.Q. Bradley, J.E. Juskewitch, P.G. Greenwood, and J.T. Millard. (2009) Quantitative PCR analysis of diepoxybutane and epihalohydrin damage to nuclear versus mitochondrial DNA, *Mutat. Res., Fundam. Mol. Mech. Mutagen. 664*, 48-54.

K.P. Romano, **A.G. Newman**, R.W. Zahran, and J.T. Millard. (2007) DNA interstrand cross-linking by epichlorohydrin, *Chem. Res. Toxicol. 20*, 832-838.

ABSTRACTS

A.G. Newman, A.L. Vagstad, P.A. Storm, C.A. Townsend. (2015) Combinatorial Biosynthesis Accomplished Through Systematic Domain Swaps of Polyketide Synthases*, Abstract of Papers of the 8th Frontiers at the Chemistry-Biology Interface Symposium.*

A.G. Newman, A.L. Vagstad, P.A. Storm, C.A. Townsend. (2015) Combinatorial Biosynthesis Accomplished Through Systematic Domain Swaps of Polyketide Synthases*, Abstract of Papers of the Society for Industrial Microbiology and Biotechnology Natural Product Discovery and Development in the Post Genomic Era.*

A.G. Newman, F.J. LaRiviere, and J.T. Millard (2007) Quantitative PCR suggests preferential nuclear DNA alkylation by epichlorohydrin in the chicken genome, *Abstract of Papers of the American Chemical Society 233,* 219.